Title
Advancements in Electron Microscopy Enabling the Structural Characterization of the Amyloid Protein LECT2 and other Macromolecules

Author
Richards, Logan

Publication Date
2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Advancements in Electron Microscopy Enabling the Structural Characterization of the Amyloid Protein LECT2 and other Macromolecules

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

by

Logan Scott Richards

2023

ABSTRACT OF THE DISSERTATION

Advancements in Electron Microscopy Enabling the Structural Characterization of the Amyloid Protein LECT2 and other Macromolecules

by

Logan Scott Richards

Doctor of Philosophy in Biochemistry, Molecular and Structural Biology

University of California, Los Angeles, 2023

Professor Jose Alfonso Rodriguez, Chair

The last two decades have seen revolutionary progress in the field of cryo electron microscopy, granting views of previously unresolved biological structures utilizing both electron diffraction and electron imaging. Microcrystal electron diffraction (MicroED) has become applicable to protein, peptide, and small molecule structures while cryo-electron microscopy (cryo-EM) has allowed for high-resolution structure determination of proteins and macromolecular complexes outside the confines of crystallographic restraints.

The use of MicroED to interrogate a wider variety of structures has revealed important molecular insights in the worlds of biology and chemistry but has also revealed some of the shortcomings of the method that could be limiting to novel structure determination. Namely, there remains a limited selection of phasing methodologies for electron diffraction data: molecular replacement, requiring a highly similar structural model, or direct methods, requiring

an atomic resolution dataset. The experiments presented in chapters 2 and 3 of this thesis explore the implementation of fragment-based phasing methodologies to electron diffraction data to overcome these limitations and demonstrate their applicability to protein and amyloid peptide structures.

Electron microscopy has been particularly important in the study of amyloid proteins. The defining characteristic of amyloid proteins is the formation long, unbranched fibrillar aggregates which can be disease-associated, as is observed in many neurodegenerative diseases, or functional within the context of cell. Short peptide segments of these proteins which recapitulate the underlying cross-β structure of the fibrils have been lucrative targets for both X-ray crystallography and MicroED due to the small size of crystals they form. Meanwhile, improvements in cryo-EM resolution and data processing have permitted the first high-resolution views of full-length amyloid fibril structures. The fourth chapter of this thesis takes advantage of these advancements to elucidate the structure of the amyloid fibril form of human leukocyte cell-derived chemotaxin-2 (LECT2) by cryo-EM. ALECT2, the systemic amyloid disease associated with LECT2 aggregation, is clinically well characterized but poorly understood molecularly. While predictions of the aggregation-prone segments which contribute to amyloid formation have been made, a lack of structural information has limited the rational design of targeting agents and therapeutics. The work presented here offers a first step in uncovering the molecular mechanisms behind LECT2 aggregation and their role in ALECT2 disease progression.

The dissertation of Logan Scott Richards is approved.

David S. Eisenberg

Todd O. Yeates

Lin Jiang

Jose Alfonso Rodriguez, Committee Chair

University of California, Los Angeles

2023

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to thank the many people who have made this work possible. I would like to thank: My advisor, Dr. Jose Rodriguez, for his guidance, support, and intellectual contributions to this dissertation. The members of my committee, Dr. David Eisenberg, Dr. Todd Yeates, and Dr. Lin Jiang for their time, advice, important contributions to this work, and valued feedback at many group meetings. Dr. Duilio Cascio and Dr. Michael Sawaya for sharing their wealth of structural biology knowledge with me and teaching me so much about X-ray crystallography and electron microscopy. Instrumental to this work has been the assistance of my many collaborators, including Dr. Isabel Usón, Dr. Rafael Borges, and Dr. Claudia Millán.

This dissertation would not be possible without the guidance and support from numerous friends and past and present lab members including Maria, Ambarneil, Sam, Natalie, Becky, Lily, Niko, Shervin, Callie, Marcus and many more who each have uniquely and positively influenced my life and experience in the lab.

I especially want to thank my boyfriend, Ben, for his love and support. Last, and most importantly, I would like to thank my family: my brother, Brody, and my parents, Janet and Scott, without whom none of this would have been possible.

The work presented here has been supported in part by the USPHS National Research Service Award 5T32GM008496, DOE Grant DE-FC02-02ER63421, NSF Grant DMR-1548924, the NIH-NIGMS Grant R35 GM128867, and the UCLA Audree V. fowler graduate Fellowships in Protein Science.

Chapter 2 in this thesis is presented (Richards et al., 2020) with permission from the authors.

Chapter 3 in this thesis is presented (Richards et al., 2023) with permission from the authors.

(1) Richards, L. S.; Millán, C.; Miao, J.; Martynowycz, M. W.; Sawaya, M. R.; Gonen, T.; Borges, R. J.; Usón, I.; Rodriguez, J. A. Fragment-Based Determination of a Proteinase K Structure from MicroED Data Using ARCIMBOLDO_SHREDDER. Acta Cryst D 2020, 76 (8), 703–712. https://doi.org/10.1107/S2059798320008049.

(2) Richards, L. S.; Flores, M. D.; Millán, C.; Glynn, C.; Zee, C.-T.; Sawaya, M. R.; Gallagher-Jones, M.; Borges, R. J.; Usón, I.; Rodriguez, J. A. Fragment-Based Ab Initio Phasing of Peptidic Nanocrystals by MicroED. ACS Bio Med Chem Au 2023. https://doi.org/10.1021/acsbiomedchemau.2c00082

CURRICULUM VITAE

Logan Scott Richards

Education

2013-2017	University of California, Irvine Bachelor of Science in Molecular Biology and Biochemistry
2017-2019	University of California, Los Angeles Master of Science in Biochemistry, Molecular and Structural Biology
2017-2023	University of California, Los Angeles Doctor of Philosophy in Biochemistry, Molecular and Structural Biology, Expected

Honors and Awards

2018-2020	UCLA Chemistry-Biology Interface Training Grant
2021-2022	Audree V. Fowler Fellowship in Protein Science

Publications: Original Research

(1) Batabyal, D.; **Richards, L. S.**; Poulos, T. L. Effect of Redox Partner Binding on Cytochrome P450 Conformational Dynamics. J. Am. Chem. Soc. 2017, 139 (37), 13193–13199. https://doi.org/10.1021/jacs.7b07656.

(2) Warmack, R. A.; Boyer, D. R.; Zee, C.-T.; **Richards, L. S.**; Sawaya, M. R.; Cascio, D.; Gonen, T.; Eisenberg, D. S.; Clarke, S. G. Structure of Amyloid-β (20-34) with Alzheimer's-Associated Isomerization at Asp23 Reveals a Distinct Protofilament Interface. Nat Commun 2019, 10 (1), 3357. https://doi.org/10.1038/s41467-019-11183-z.

(3) Gallagher-Jones, M.; Bustillo, K. C.; Ophus, C.; **Richards, L. S.**; Ciston, J.; Lee, S.; Minor, A. M.; Rodriguez, J. A. Atomic Structures Determined from Digitally Defined Nanocrystalline Regions. IUCrJ 2020, 7 (Pt 3), 490–499. https://doi.org/10.1107/S2052252520004030.https://doi.org/10.1021/jacs.7b07656.

(4) **Richards, L. S.**; Millán, C.; Miao, J.; Martynowycz, M. W.; Sawaya, M. R.; Gonen, T.; Borges, R. J.; Usón, I.; Rodriguez, J. A. Fragment-Based Determination of a Proteinase K Structure from MicroED Data Using ARCIMBOLDO_SHREDDER. Acta Cryst D 2020, 76 (8), 703–712. https://doi.org/10.1107/S2059798320008049.

(5) **Richards, L. S.**; Flores, M. D.; Millán, C.; Glynn, C.; Zee, C.-T.; Sawaya, M. R.; Gallagher-Jones, M.; Borges, R. J.; Usón, I.; Rodriguez, J. A. Fragment-Based Ab Initio Phasing of Peptidic Nanocrystals by MicroED. ACS Bio Med Chem Au 2023. https://doi.org/10.1021/acsbiomedchemau.2c00082.

(6) **Richards, L. S.**; Flores, M. D.; Zink, S.; Schibrowsky, N. A.; Sawaya, M. R.; Rodriguez, J. A. Cryo-EM Structure of a Human LECT2 Amyloid Fibril Reveals a Network of Polar Ladders at Its Core. bioRxiv February 11, 2023, p 2023.02.08.527771. https://doi.org/10.1101/2023.02.08.527771.

Chapter 1

Introduction

1.1 Advancements in Electron Microscopy

The interrogation of matter using electrons began with Davisson and Germer's landmark diffraction experiments on crystalline nickel nearly a century ago[1,2]. Since these humble beginnings, the field of electron microscopy has grown to encompass an immense variety of techniques for structure elucidation based on imaging and diffraction methods. In particular, the last two decades have seen a boon of technical and methodological advancements which have, for the first time, enabled high-resolution structure determination of biological molecules by cryo-electron microscopy (cryo-EM) and microcrystal electron diffraction (MicroED). The so-called "resolution revolution"[3] in cryo-EM has been enabled largely by the development of ultra-fast direct electron detectors,[4] which allow the detection of low dose electron micrographs at speeds that allow for correction of specimen motion[5]. This, in combination with improved electron sources in the form of field emission guns, phase plate technology[6], and the streamlining of data collection and processing[7], have catalyzed nearly atomic resolution structures of biological molecules to be resolved from non-crystalline samples[8]. Electron diffraction techniques have seen a similarly rapid advancement in the last decade fueled by the previously listed technical developments in combination with modifications to data collection which helped minimize the inherent problem of multiple scattering[9–11]. MicroED has become a useful technique for the analysis of biological specimens[12], but challenges remain for novel structure determination.

1.2 The Phase Problem in MicroED

Electron diffraction draws much from the foundational knowledge established by X-ray crystallography and, as with all diffraction techniques, is in turn limited by the phase problem. When recording diffraction itensities, we lose information about the relative phases associated

with each structure factor. Recapitulating a structure from the measured intensities requires the retrieval of those phases and it is at this step where the fundamental differences between X-ray and electron diffraction begin to limit the application of previously established techniques. Well-known phasing techniques for X-ray crystallography include phase retrieval by isomorphous replacement[13], which relies on scattering differences derived from the addition of a heavy atom, anomalous dispersion[14], where anomalous scattering is measured using radiation in the vicinity of an absorption edge of one of the constituent elements, molecular replacement[15], requiring closely related structural model to derive an initial structural hypothesis, and direct methods[16], which relies on atomic resolution data for *ab initio* structure determination. However, isomorphous replacement methods have never been demonstrated for electron diffraction and may be an intractable approach due to the much smaller scattering differences observed relative to X-ray diffraction[17,18], and its strict need for isomorphism. Further, the absence of atomic absorption edges at the energies used for electron diffraction excludes the application of anomalous dispersion-based phasing[18,19]. This means that phasing for MicroED structures must generally be accomplished by either molecular replacement, assuming a suitable model is available, or by direct methods, assuming atomic resolution is achievable. Novel macromolecular structures may not have a suitably similar model to draw on and, except for rare exceptions, atomic resolution is a rarity in diffraction of protein crystals[20] so it is easy to imagine a scenario where neither of these criteria can be met. These limitations demand an expansion of phasing techniques for MicroED data to ensure that novel structures are not lost to the age-old phase problem.

1.3 Interrogating Amyloids with Electrons

Debilitating neurodegenerative diseases such as Alzheimer's and Parkinson's disease are well-known to involve amyloid fibril deposition in the brain[21,22], but many other examples of amyloid diseases involve the systemic deposition of persistent amyloid fibrils throughout various

organs[23]. The accumulation of amyloid aggregates is a common feature among these diseases, but the factors contributing to the misfolding, stability and toxicity of each disease-associated fibril are unique and likely dependent on their underlying fine structures[12]. This diversity underscores a drive to characterize the breadth of amyloid fibril structures. Amyloid proteins are a varied group, encompassing both functional and pathogenic proteins[24], which all share a common structural motif. Almost without exception, amyloid fibrils adopt beta-strand-rich two-dimensional conformations which stack in the third dimension to produce long, unbranching, twisting fibrils[25]. In the case of pathogenic amyloids these fibrils can be extremely stable, resistant to denaturation by chaotropic agents and digestion by proteolytic enzymes, while functional amyloids have their stability tempered to allow for, in some cases, reversibility of the fold[24]. Amyloids are often also polymorphic, by which a single protein sequence can adopt different two-dimensional folds within the context of a fibril[12]. The structures of amyloids are inherently tied to their energetic profile, pathogenicity, and function, but these structures are not trivial to interrogate. X-ray fiber diffraction originally defined the amyloid according to its cross-β diffraction pattern[26], by which it can be deduced that the stacked β-strands are spaced 4.8 Å apart and the β sheets packed across from each other are 8–12 Å apart. While the one-dimensional elongation of fibrils limited crystallographic structural studies of full-length amyloids, short peptide segments could be crystallized whose structures recapitulated the cross-β fold[27]. These crystals tend to be small, limited by the strain of stacked β strands that cannot twist due to crystallographic constraints[28], and required advancements first in X-ray microfocus beamlines[27] and then later MicroED[28] to facilitate structure determination. As advancements in MicroED have facilitated structure solution for peptide segments, the previously described advancements in cryo-EM have also permitted the first structures of full-length amyloid fibrils to be resolved[29]. These structures relied on the development of helical reconstruction algorithms applied to single-particle cryo-EM data [30] and the rapid resolution improvement of collected data

to produce the first comprehensive and detailed structural information for pathogenic and functional amyloids alike.

1.4 LECT2 and its Role in Renal Amyloidosis

1.4.1 ALECT2

ALECT2 is a systemic amyloid disease characterized by the accumulation of amyloid fibrils of the protein leukocyte cell-derived chemotaxin-2 (LECT2) in the kidneys, liver, spleen, adrenal glands, and colon[31,32]. ALECT2 causes impaired renal function, renal failure and nephrotic syndrome,[31] is age associated, and primarily affects patients of Mexican, Punjabi, South Asian, Sudanese, Native American, and Egyptian descent[33]. The bulk extracellular amyloid aggregates in the interstitial space and glomeruli stain strongly with the amyloid marker Congo Red and the protein identity was initially confirmed by staining with an anti-LECT2 antibody[34]. Since then, LECT2 aggregates have been observed surrounding the kidney vasculature as well[35]. In the absence of specific and non-invasive identification methods, ALECT2 can often be mistaken for more well-known kidney amyloid diseases such as immunoglobulin light chain amyloidosis (AL) or Amyloid A amyloidosis (AA), resulting in patients being administered inappropriate and often strenuous treatments such as aggressive chemotherapy for AL or targeted treatment of the underlying inflammatory process for AA[36]. Little is known about what predisposes individuals to acquire ALECT2, but there is a common nucleotide polymorphism among most patients at nucleotide 172 which results in an isoleucine to valine sequence polymorphism difference at position 40 of the full-length protein[37]. There are currently no available treatments for ALECT2 and no molecules that effectively target the amyloid form of LECT2, underscoring the need to establish fundamental knowledge about the genesis and structure of LECT2 amyloid fibrils.

1.4.2 LECT2 Structure and Disfunction

Nascent LECT2 is expressed in the liver as a 151 amino acid polypeptide with an 18-residue N-terminal signal peptide that is cleaved before its secretion into the blood as a globular protein[38]. LECT2 has widespread functions as a chemotactic factor to promote the migration of neutrophils to sites of infection[38,39], a regulator of cartilage growth[40], and a hepatokine important for metabolic homeostasis[41]. The structure of globular LECT2 has been determined previously and is a 133 amino acid long, 16kDa Zinc binding protein with no catalytic activity despite structural similarity to the M23 family of metalloendopeptidases[42]. Loss of bound Zn(II) is hypothesized play an important role in the conversion of properly folded LECT2 into an amyloid form by destabilizing the central beta-barrel motif within the protein[43]. The presence of the I40V mutation appears to be important, but not sufficient for inducing amyloid formation. It does not seem to destabilize the metal binding properties of the functional protein[43]. Amyloidogenic segments of LECT2 have been identified, and predicted amyloid-forming LECT2 peptides assemble into amyloid-like fibrils in isolation[44]. However, the design of diagnostic agents and effective treatments is limited by a lack of structural knowledge identifying protein segments that may be important for amyloid formation and stability in the context of a full-length fibril.

1.5 Overview

Chapters 2 and 3 of this thesis explore the application of fragment-based phasing techniques to globular protein and amyloid peptide structures determined by MicroED to expand the available phasing solutions for previously recalcitrant datasets. Chapter 4 focuses on the use of cryo-EM to interrogate a stable amyloid fibril structure formed by LECT2, yielding the first high-resolution structural detail of LECT2 in an amyloid form and identifying key residues that may play a role in ALECT2 disease progression.

References

(1) Davisson, C.; Germer, L. H. Diffraction of Electrons by a Crystal of Nickel. *Phys. Rev.* **1927**, *30* (6), 705–740. https://doi.org/10.1103/PhysRev.30.705.

(2) Davisson, C.; Germer, L. H. The Scattering of Electrons by a Single Crystal of Nickel. *Nature* **1927**, *119* (2998), 558–560. https://doi.org/10.1038/119558a0.

(3) Kühlbrandt, W. The Resolution Revolution. *Science* **2014**, *343* (6178), 1443–1444. https://doi.org/10.1126/science.1251652.

(4) McMullan, G.; Clark, A. T.; Turchetta, R.; Faruqi, A. R. Enhanced Imaging in Low Dose Electron Microscopy Using Electron Counting. *Ultramicroscopy* **2009**, *109* (12), 1411–1416. https://doi.org/10.1016/j.ultramic.2009.07.004.

(5) Li, X.; Mooney, P.; Zheng, S.; Booth, C. R.; Braunfeld, M. B.; Gubbens, S.; Agard, D. A.; Cheng, Y. Electron Counting and Beam-Induced Motion Correction Enable near-Atomic-Resolution Single-Particle Cryo-EM. *Nat Methods* **2013**, *10* (6), 584–590. https://doi.org/10.1038/nmeth.2472.

(6) Danev, R.; Buijsse, B.; Khoshouei, M.; Plitzko, J. M.; Baumeister, W. Volta Potential Phase Plate for In-Focus Phase Contrast Transmission Electron Microscopy. *Proceedings of the National Academy of Sciences* **2014**, *111* (44), 15635–15640. https://doi.org/10.1073/pnas.1418377111.

(7) Suloway, C.; Pulokas, J.; Fellmann, D.; Cheng, A.; Guerra, F.; Quispe, J.; Stagg, S.; Potter, C. S.; Carragher, B. Automated Molecular Microscopy: The New Leginon System. *J Struct Biol* **2005**, *151* (1), 41–60. https://doi.org/10.1016/j.jsb.2005.03.010.

(8) Nakane, T.; Kotecha, A.; Sente, A.; McMullan, G.; Masiulis, S.; Brown, P. M. G. E.; Grigoras, I. T.; Malinauskaite, L.; Malinauskas, T.; Miehling, J.; Uchański, T.; Yu, L.; Karia, D.; Pechnikova, E. V.; de Jong, E.; Keizer, J.; Bischoff, M.; McCormack, J.; Tiemeijer, P.; Hardwick, S. W.; Chirgadze, D. Y.; Murshudov, G.; Aricescu, A. R.; Scheres, S. H. W. Single-Particle Cryo-EM at Atomic Resolution. *Nature* **2020**, *587* (7832), 152–156. https://doi.org/10.1038/s41586-020-2829-0.

(9) Kolb, U.; Gorelik, T.; Kübel, C.; Otten, M. T.; Hubert, D. Towards Automated Diffraction Tomography: Part I—Data Acquisition. *Ultramicroscopy* **2007**, *107* (6), 507–513. https://doi.org/10.1016/j.ultramic.2006.10.007.

(10) Kolb, U.; Gorelik, T.; Otten, M. T. Towards Automated Diffraction Tomography. Part II—Cell Parameter Determination. *Ultramicroscopy* **2008**, *108* (8), 763–772. https://doi.org/10.1016/j.ultramic.2007.12.002.

(11) Mugnaioli, E.; Gorelik, T.; Kolb, U. “Ab Initio” Structure Solution from Electron Diffraction Data Obtained by a Combination of Automated Diffraction Tomography and Precession Technique. *Ultramicroscopy* **2009**, *109* (6), 758–765. https://doi.org/10.1016/j.ultramic.2009.01.011.

(12) Shi, Y.; Zhang, W.; Yang, Y.; Murzin, A. G.; Falcon, B.; Kotecha, A.; van Beers, M.; Tarutani, A.; Kametani, F.; Garringer, H. J.; Vidal, R.; Hallinan, G. I.; Lashley, T.; Saito, Y.;

Murayama, S.; Yoshida, M.; Tanaka, H.; Kakita, A.; Ikeuchi, T.; Robinson, A. C.; Mann, D. M. A.; Kovacs, G. G.; Revesz, T.; Ghetti, B.; Hasegawa, M.; Goedert, M.; Scheres, S. H. W. Structure-Based Classification of Tauopathies. *Nature* **2021**, *598* (7880), 359–363. https://doi.org/10.1038/s41586-021-03911-7.

(13) Green, D. W.; Ingram, V. M.; Perutz, M. F.; Bragg, W. L. The Structure of Haemoglobin - IV. Sign Determination by the Isomorphous Replacement Method. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **1997**, *225* (1162), 287–307. https://doi.org/10.1098/rspa.1954.0203.

(14) Matthews, B. W. The Determination of the Position of Anomalously Scattering Heavy Atom Groups in Protein Crystals. *Acta Cryst* **1966**, *20* (2), 230–239. https://doi.org/10.1107/S0365110X6600046X.

(15) Rossmann, M. G.; Blow, D. M.; IUCr. *The detection of sub-units within the crystallographic asymmetric unit*. Acta Crystallographica. https://doi.org/10.1107/S0365110X62000067.

(16) Sheldrick, G. M.; Gilmore, C. J.; Hauptman, H. A.; Weeks, C. M.; Miller, R.; Usón, I. Ab Initio Phasing. In *International Tables for Crystallography*; John Wiley & Sons, Ltd, 2012; pp 413–432. https://doi.org/10.1107/97809553602060000850.

(17) Ceska, T. A.; Henderson, R. Analysis of High-Resolution Electron Diffraction Patterns from Purple Membrane Labelled with Heavy-Atoms. *J Mol Biol* **1990**, *213* (3), 539–560. https://doi.org/10.1016/S0022-2836(05)80214-1.

(18) Burmester, C.; Schröder, R. SOLVING THE PHASE PROBLEM IN PROTEIN ELECTRON CRYSTALLOGRAPHY: MULTIPLE ISOMORPHOUS REPLACEMENT AND ANOMALOUS DISPERSION AS ALTERNATIVES TO IMAGING; 1997.

(19) Doyle, P. A.; Turner, P. S. Relativistic Hartree–Fock X-Ray and Electron Scattering Factors. *Acta Cryst A* **1968**, *24* (3), 390–397. https://doi.org/10.1107/S0567739468000756.

(20) Martynowycz, M. W.; Clabbers, M. T. B.; Hattne, J.; Gonen, T. Ab Initio Phasing Macromolecular Structures Using Electron-Counted MicroED Data. *Nat Methods* **2022**, *19* (6), 724–729. https://doi.org/10.1038/s41592-022-01485-4.

(21) Selkoe, D. J. The Molecular Pathology of Alzheimer's Disease. *Neuron* **1991**, *6* (4), 487–498. https://doi.org/10.1016/0896-6273(91)90052-2.

(22) Spillantini, M. G.; Crowther, R. A.; Jakes, R.; Hasegawa, M.; Goedert, M. α-Synuclein in Filamentous Inclusions of Lewy Bodies from Parkinson's Disease and Dementia with Lewy Bodies. *Proceedings of the National Academy of Sciences* **1998**, *95* (11), 6469–6473. https://doi.org/10.1073/pnas.95.11.6469.

(23) Muchtar, E.; Dispenzieri, A.; Magen, H.; Grogan, M.; Mauermann, M.; McPhail, E. D.; Kurtin, P. J.; Leung, N.; Buadi, F. K.; Dingli, D.; Kumar, S. K.; Gertz, M. A. Systemic Amyloidosis from A (AA) to T (ATTR): A Review. *Journal of Internal Medicine* **2021**, *289* (3), 268–292. https://doi.org/10.1111/joim.13169.

(24) Sawaya, M. R.; Hughes, M. P.; Rodriguez, J. A.; Riek, R.; Eisenberg, D. S. The Expanding Amyloid Family: Structure, Stability, Function, and Pathogenesis. *Cell* **2021**, *184* (19), 4857–4873. https://doi.org/10.1016/j.cell.2021.08.013.

(25) Eisenberg, D.; Jucker, M. The Amyloid State of Proteins in Human Diseases. *Cell* **2012**, *148* (6), 1188–1203. https://doi.org/10.1016/j.cell.2012.02.022.

(26) Astbury, W. T.; Dickinson, S.; Bailey, K. The X-Ray Interpretation of Denaturation and the Structure of the Seed Globulins. *Biochem J* **1935**, *29* (10), 2351-2360.1.

(27) Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A. Ø.; Riekel, C.; Grothe, R.; Eisenberg, D. Structure of the Cross-β Spine of Amyloid-like Fibrils. *Nature* **2005**, *435* (7043), 773–778. https://doi.org/10.1038/nature03680.

(28) Rodriguez, J. A.; Ivanova, M. I.; Sawaya, M. R.; Cascio, D.; Reyes, F. E.; Shi, D.; Sangwan, S.; Guenther, E. L.; Johnson, L. M.; Zhang, M.; Jiang, L.; Arbing, M. A.; Nannenga, B. L.; Hattne, J.; Whitelegge, J.; Brewster, A. S.; Messerschmidt, M.; Boutet, S.; Sauter, N. K.; Gonen, T.; Eisenberg, D. S. Structure of the Toxic Core of α-Synuclein from Invisible Crystals. *Nature* **2015**, *525* (7570), 486–490. https://doi.org/10.1038/nature15368.

(29) 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. Cryo-EM Structures of Tau Filaments from Alzheimer's Disease. *Nature* **2017**, *547* (7662), 185–190. https://doi.org/10.1038/nature23002.

(30) He, S.; Scheres, S. H. W. Helical Reconstruction in RELION. *Journal of Structural Biology* **2017**, *198* (3), 163–176. https://doi.org/10.1016/j.jsb.2017.02.003.

(31) Benson, M. D.; James, S.; Scott, K.; Liepnieks, J. J.; Kluve-Beckerman, B. Leukocyte Chemotactic Factor 2: A Novel Renal Amyloid Protein. *Kidney International* **2008**, *74* (2), 218–222. https://doi.org/10.1038/ki.2008.152.

(32) Dogan, A. Amyloidosis: Insights from Proteomics. *Annual Review of Pathology: Mechanisms of Disease* **2017**, *12* (1), 277–304. https://doi.org/10.1146/annurev-pathol-052016-100200.

(33) Larsen, C. P.; Ismail, W.; Kurtin, P. J.; Vrana, J. A.; Dasari, S.; Nasr, S. H. Leukocyte Chemotactic Factor 2 Amyloidosis (ALECT2) Is a Common Form of Renal Amyloidosis among Egyptians. *Mod Pathol* **2016**, *29* (4), 416–420. https://doi.org/10.1038/modpathol.2016.29.

(34) Benson, M. D. LECT2 Amyloidosis. *Kidney International* **2010**, *77* (9), 757–759. https://doi.org/10.1038/ki.2010.18.

(35) Nasr, S. H.; Dogan, A.; Larsen, C. P. Leukocyte Cell–Derived Chemotaxin 2–Associated Amyloidosis: A Recently Recognized Disease with Distinct Clinicopathologic Characteristics. *Clin J Am Soc Nephrol* **2015**, *10* (11), 2084–2093. https://doi.org/10.2215/CJN.12551214.

(36) Picken, M. M. Alect2 Amyloidosis: Primum Non Nocere (First, Do No Harm). *Kidney Int* **2014**, *86* (2), 229–232. https://doi.org/10.1038/ki.2014.45.

(37) Larsen, C. P.; Kossmann, R. J.; Beggs, M. L.; Solomon, A.; Walker, P. D. Clinical, Morphologic, and Genetic Features of Renal Leukocyte Chemotactic Factor 2 Amyloidosis. *Kidney International* **2014**, *86* (2), 378–382. https://doi.org/10.1038/ki.2014.11.

(38) Yamagoe, S.; Yamakawa, Y.; Matsuo, Y.; Minowada, J.; Mizuno, S.; Suzuki, K. Purification and Primary Amino Acid Sequence of a Novel Neutrophil Chemotactic Factor LECT2. *Immunology Letters* **1996**, *52* (1), 9–13. https://doi.org/10.1016/0165-2478(96)02572-2.

(39) Yamagoe, S.; Kameoka, Y.; Hashimoto, K.; Mizuno, S.; Suzuki, K. Molecular Cloning, Structural Characterization, and Chromosomal Mapping of the Human LECT2 Gene. *Genomics* **1998**, *48* (3), 324–329. https://doi.org/10.1006/geno.1997.5198.

(40) Hiraki, Y.; Inoue, H.; Kondo, J.; Kamizono, A.; Yoshitake, Y.; Shukunami, C.; Suzuki, F. A Novel Growth-Promoting Factor Derived from Fetal Bovine Cartilage, Chondromodulin II: PURIFICATION AND AMINO ACID SEQUENCE*. *Journal of Biological Chemistry* **1996**, *271* (37), 22657–22662. https://doi.org/10.1074/jbc.271.37.22657.

(41) Willis, S. A.; Sargeant, J. A.; Yates, T.; Takamura, T.; Takayama, H.; Gupta, V.; Brittain, E.; Crawford, J.; Parry, S. A.; Thackray, A. E.; Varela-Mato, V.; Stensel, D. J.; Woods, R. M.; Hulston, C. J.; Aithal, G. P.; King, J. A. Acute Hyperenergetic, High-Fat Feeding Increases Circulating FGF21, LECT2, and Fetuin-A in Healthy Men. *The Journal of Nutrition* **2020**, *150* (5), 1076–1085. https://doi.org/10.1093/jn/nxz333.

(42) Zheng, H.; Miyakawa, T.; Sawano, Y.; Asano, A.; Okumura, A.; Yamagoe, S.; Tanokura, M. Crystal Structure of Human Leukocyte Cell-Derived Chemotaxin 2 (LECT2) Reveals a Mechanistic Basis of Functional Evolution in a Mammalian Protein with an M23 Metalloendopeptidase Fold*. *Journal of Biological Chemistry* **2016**, *291* (33), 17133–17142. https://doi.org/10.1074/jbc.M116.720375.

(43) Ha, J.-H.; Tu, H.-C.; Wilkens, S.; Loh, S. N. Loss of Bound Zinc Facilitates Amyloid Fibril Formation of Leukocyte-Cell-Derived Chemotaxin 2 (LECT2). *Journal of Biological Chemistry* **2021**, *296*, 100446. https://doi.org/10.1016/j.jbc.2021.100446.

(44) Tsiolaki, P. L.; Nasi, G. I.; Baltoumas, F. A.; Fishman, S.; Tu, H.-C.; Iconomidou, V. A. Delving into the Amyloidogenic Core of Human Leukocyte Chemotactic Factor 2. *Journal of Structural Biology* **2019**, *207* (3), 260–269. https://doi.org/10.1016/j.jsb.2019.06.001.

Chapter 2

This chapter is based on the published article “Fragment-based determination of a proteinase K structure from MicroED data using ARCIMBOLDO_SHREDDER.” [1]

Summary

This article explores the application of fragment-based phasing techniques, using the program ARCIMBOLDO_SHREDDER, to the determination of MicroED protein structures. While the phase problem is a common obstacle in all crystallographic methods, it is particularly limiting in the case of MicroED data as isomorphous replacement methods have not yet been demonstrated and anomalous dispersion-based phasing is limited by the lack of atomic absorption edges at the energies used for electron diffraction. Without these experimental phasing methodologies phasing of MicroED data is limited to molecular replacement, requiring an accurate model from a previously determined structure, or direct methods, requiring atomic resolution data which is often unattainable in MicroED datasets of large proteins. This article expands the available phasing options in MicroED data processing by demonstrating that fragment-based phasing techniques can be advantageous when applied to MicroED data at resolutions too poor for direct methods. In this work, MicroED data of proteinase K is phased using a library of secondary-structure and tertiary-structure fragments derived from a distantly related homologous protein. These fragments are initially used as molecular-replacement search models so an ensemble of polyalanine fragments may be placed and further refined with density modification and autotracing in SHELXE, thus overcoming the need for a single closely related model to phase MicroED data.

(1) Richards, L. S.; Millán, C.; Miao, J.; Martynowycz, M. W.; Sawaya, M. R.; Gonen, T.; Borges, R. J.; Usón, I.; Rodriguez, J. A. Fragment-Based Determination of a Proteinase K Structure from MicroED Data Using ARCIMBOLDO_SHREDDER. Acta Cryst D 2020, 76 (8), 703–712. https://doi.org/10.1107/S2059798320008049.

STRUCTURAL BIOLOGY

ISSN 2059-7983

Fragment-based determination of a proteinase K structure from MicroED data using *ARCIMBOLDO_SHREDDER*

Logan S. Richards,[a]‡ Claudia Millán,[b]‡ Jennifer Miao,[a]‡ Michael W. Martynowycz,[c,d] Michael R. Sawaya,[c,d] Tamir Gonen,[c,d,e] Rafael J. Borges,[b] Isabel Usón[b,f]* and Jose A. Rodriguez[a]*

[a]Department of Chemistry and Biochemistry; UCLA–DOE Institute for Genomics and Proteomics; STROBE, NSF Science and Technology Center, University of California Los Angeles (UCLA), Los Angeles, CA 90095, USA, [b]Crystallographic Methods, Institute of Molecular Biology of Barcelona (IBMB–CSIC), Barcelona Science Park, Helix Building, Baldiri Reixac 15, 08028 Barcelona, Spain, [c]Howard Hughes Medical Institute, University of California Los Angeles (UCLA), Los Angeles, California, USA, [d]Department of Biological Chemistry, University of California Los Angeles (UCLA), Los Angeles, CA 90095, USA, [e]Department of Physiology, University of California Los Angeles (UCLA), Los Angeles, California, USA, and [f]ICREA, Institució Catalana de Recerca i Estudis Avançats, Passeig Lluís Companys 23, 08003 Barcelona, Spain. *Correspondence e-mail: uson@ibmb.csic.es, jrodriguez@mbi.ucla.edu

Received 22 January 2020
Accepted 16 June 2020

Edited by R. J. Read, University of Cambridge, United Kingdom

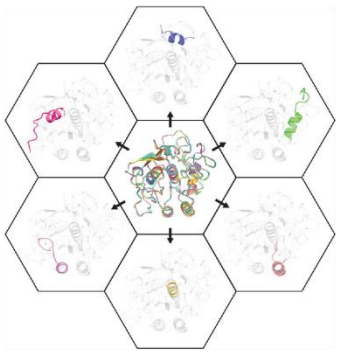
‡ These authors made equivalent contributions.

Keywords: phasing; crystal; fragment; electron diffraction; cryoEM; MicroED; proteinase K; *ARCIMBOLDO_SHREDDER*.

PDB reference: proteinase K, 6v8r

Supporting information: this article has supporting information at journals.iucr.org/d

Structure determination of novel biological macromolecules by X-ray crystallography can be facilitated by the use of small structural fragments, some of only a few residues in length, as effective search models for molecular replacement to overcome the phase problem. Independence from the need for a complete pre-existing model with sequence similarity to the crystallized molecule is the primary appeal of *ARCIMBOLDO*, a suite of programs which employs this *ab initio* algorithm for phase determination. Here, the use of *ARCIMBOLDO* is investigated to overcome the phase problem with the electron cryomicroscopy (cryoEM) method known as microcrystal electron diffraction (MicroED). The results support the use of the *ARCIMBOLDO_SHREDDER* pipeline to provide phasing solutions for a structure of proteinase K from 1.6 Å resolution data using model fragments derived from the structures of proteins sharing a sequence identity of as low as 20%. *ARCIMBOLDO_SHREDDER* identified the most accurate polyalanine fragments from a set of distantly related sequence homologues. Alternatively, such templates were extracted in spherical volumes and given internal degrees of freedom to refine towards the target structure. Both modes relied on the rotation function in *Phaser* to identify or refine fragment models and its translation function to place them. Model completion from the placed fragments proceeded through phase combination of partial solutions and/or density modification and main-chain autotracing using *SHELXE*. The combined set of fragments was sufficient to arrive at a solution that resembled that determined by conventional molecular replacement using the known target structure as a search model. This approach obviates the need for a single, complete and highly accurate search model when phasing MicroED data, and permits the evaluation of large fragment libraries for this purpose.



1. Introduction

Crystallography has remained an indispensable method for structure determination since its initial demonstration over a century ago (Bragg & Bragg, 1913). Beyond X-ray diffraction, neutron and electron diffraction have contributed important advances to the crystallographic determination of macromolecular structures (Glaeser, 1999; Shi *et al.*, 2013; Gemmi *et al.*, 2019). Recently, an electron crystallography method called microcrystal electron diffraction (MicroED) has been developed to obtain high-resolution structures from frozen-hydrated three-dimensional macromolecular crystals

(Supplementary Fig. S1; Nannenga, Shi, Hattne *et al.*, 2014). In MicroED, crystals of a few hundred nanometres in thickness are continuously rotated in an electron beam while diffraction is measured from a region of the specimen defined by the selected area aperture; the latter is positioned at the conjugate image plane of the objective lens. The recorded diffraction is reduced using conventional X-ray crystallography software to yield data that are suitable for structure determination. Phasing of MicroED data for biomolecules has been achieved by three approaches: by molecular replacement (Shi *et al.*, 2013), by direct methods (Sawaya *et al.*, 2016) or by using radiation damage (Martynowycz *et al.*, 2020). Refinement proceeds through programs such as *REFMAC* (Kovalevskiy *et al.*, 2018), *phenix.refine* (Afonine *et al.*, 2012) or *SHELXL* (Sheldrick, 2015*b*) using electron scattering factors.

Important differences between X-ray and electron diffraction can impact the phase problem. Differences in wavelength (λ) impact Ewald sphere curvature such that λ is $\sim$1 Å in a typical X-ray experiment but $\sim$0.025 Å in a MicroED experiment performed using 200 keV electrons (Rodriguez & Gonen, 2016). While the relatively large scattering cross section of electrons in the 200–300 keV energy range is beneficial for extracting signal from very thin nanocrystals, the strong interaction between electrons and matter gives rise to a greater fraction of multiple scattering and absorption for thicker samples (Jansen *et al.*, 1998). These effects can be mitigated by using higher energy electrons, since penetration depth and kinematic scattering increase with electron beam energy, allowing thicker crystals to be interrogated at higher energies (300 keV). However, high beam energies increase the likelihood of knock-on radiation damage (Subramanian *et al.*, 2015). These challenges, combined with limited rotation of the electron microscope stage ($\sim$140° maximally) and the possibility of low symmetry and some crystallites oriented preferentially on the grid surface, can lead to reduced completeness in MicroED movies from single crystals (Rodriguez & Gonen, 2016; Wennmacher *et al.*, 2019). Near-total completeness data sets are achieved through the merging of data from several crystals, but the merged data quality can be hindered by non-isomorphism as well as variations in crystal size and thickness at the nanoscale, all of which introduce difficulties in scaling (de la Cruz *et al.*, 2017). These, in addition to differences in X-ray and electron scattering factors (Colliex *et al.*, 2006), and limitations in the existing electron scattering factor libraries, make experimental phasing more challenging for electron diffraction applications, even without accounting for the impact of charged atoms (Yonekura *et al.*, 2015).

The phase problem is a common obstacle in all crystallographic methods, including MicroED (Hattne *et al.*, 2015; Shi *et al.*, 2016). Determination of the first protein structure by MicroED, a 2.9 Å resolution structure of hen egg-white lysozyme, was achieved by molecular replacement using a known lysozyme polyalanine model (Shi *et al.*, 2013), akin to previous efforts in electron diffraction (Gonen *et al.*, 2005). No globular protein structure has been determined by direct methods from MicroED data; the approach has thus far only succeeded for MicroED of peptides and small molecules (Sawaya *et al.*, 2016; Genderen *et al.*, 2016). Isomorphous replacement methods have not yet been demonstrated for MicroED and this is considered to be a potentially intractable approach (Ceska & Henderson, 1990; Burmester & Schroeder, 1997). The lack of atomic absorption edges at the energies used for electron diffraction leaves little opportunity for anomalous dispersion-based phasing (Doyle & Turner, 1968; Burmester & Schroeder, 1997; Colliex *et al.*, 2006). Furthermore, initial efforts in structure determination by MicroED were overcast by concerns that dynamical scattering would scramble the intensities recorded from 3D protein crystals. The application of continuous rotation, yielding more accurate intensities (Nannenga, Shi, Leslie *et al.*, 2014), and the determination of novel biostructures has helped to dispel some of these concerns (see, for example, Rodriguez *et al.*, 2015).

A growing number of MicroED structures have been determined at resolutions outside the high-resolution regime by molecular replacement (Nannenga & Gonen, 2019). These include the structures of a fragment of α-synuclein at 1.4 Å resolution (Rodriguez *et al.*, 2015), of bovine liver catalase at 3.2 Å resolution (Nannenga, Shi, Hattne *et al.*, 2014) and of a Ca^{2+}-ATPase at 3.2 Å resolution (Yonekura *et al.*, 2015). In each case, the use of near-ideal models also overcame potential issues with data quality that may pose barriers to phasing, including low completeness or high integration errors (Hattne *et al.*, 2015). With continued improvements to data collection and processing, novel structures continue to be determined by MicroED (Hughes *et al.*, 2018; Jones *et al.*, 2018; Purdy *et al.*, 2018; de la Cruz *et al.*, 2017; Zhou *et al.*, 2019; Xu *et al.*, 2018, 2019). Despite these successes, caution is prudent when evaluating the influence of model bias on the final structures, particularly where the model-to-structure r.m.s.d. is low; this has been the norm for many MicroED structures to date. Phasing in MicroED without atomic resolution data ($\sim$1 Å) is a challenge and, given the complications regarding the experimental data, the phasing of protein structures by *ab initio* methods has immediate advantages: it does not require stereochemical knowledge, experimental modification of crystals or the collection of data at specific wavelengths (Hauptman, 1986; Sheldrick *et al.*, 2012; Usón & Sheldrick, 1999). Ultimately, atom placements must be computed whose transforms best correlate with the measured data and allow the generation of density maps that yield a refined structure (Sheldrick, 2015*a*).

ARCIMBOLDO is a suite of software distributed within *CCP*4 (Winn *et al.*, 2011) that uses libraries of secondary-structure and tertiary-structure elements as initial search fragments for molecular replacement executed by *Phaser*, in which each fragment is oriented and positioned in the unit cell (McCoy *et al.*, 2007). Initial maps are then computed and improved by density modification using *SHELXE* (Thorn & Sheldrick, 2013). Finally, main-chain autotracing (Sheldrick, 2010) is performed to provide a reliable figure of merit at a given resolution in the form of a correlation coefficient (CC; Fujinaga & Read, 1987). In this way, *ARCIMBOLDO* substitutes the atomicity requirement in direct methods with the enforcement of secondary structure in order to accomplish

fragment-based molecular replacement at resolutions near 2.0 Å (Rodríguez *et al.*, 2009).

ARCIMBOLDO can generate libraries of secondary-structure or tertiary-structure fragment search models in multiple ways (Rodríguez *et al.*, 2012; Medina *et al.*, 2020). The most effective search model in *ARCIMBOLDO* is an α-helix owing to its ubiquitous presence in protein structures, its constant geometry and its generally low *B* factors given its structural rigidity (Millán, Sammito & Usón, 2015). Libraries of idealized polyalanine helices can be generated for use in *ARCIMBOLDO_LITE* (Sammito *et al.*, 2015), while both secondary-structure and tertiary-structure elements can be made by extraction from the wide variety of existing structures deposited in the PDB using *ARCIMBOLDO_BORGES* (Sammito *et al.*, 2013). Another variant of *ARCIMBOLDO*, *ARCIMBOLDO_SHREDDER*, uses distantly related homologous structures to generate these polyalanine fragments for use as initial search models (Sammito *et al.*, 2014). This approach to molecular replacement eliminates the need for a single model and instead generates many possible models from low sequence-similarity homologues, low-resolution NMR structures or DNA-binding motifs (Pröpper *et al.*, 2014).

We now expand the available methods for the phasing of MicroED data at resolutions outside the atomic regime. We rely on fragments of homologue structures with low identity to the target for phasing using *ARCIMBOLDO_SHREDDER*. The success of this approach is demonstrated for proteinase K using a library of secondary-structure and tertiary-structure fragments from a distantly related homologous protein as molecular-replacement search models. An ensemble of polyalanine fragments from this library are placed and refined with density modification and autotracing in *SHELXE* (Usón & Sheldrick, 2018), obviating the need for a single closely related model to phase MicroED data.

2. Materials and methods

2.1. Collection and processing of MicroED data for proteinase K

The MicroED data sets originally used to produce the structures with PDB codes 5k7s (de la Cruz *et al.*, 2017) and 6cl7 (Hattne *et al.*, 2018) were integrated using *MOSFLM*. However, this procedure initially generated swayed intensities, as shown by analyzing normalized structure factors. This in turn suggested that the data might suffer from twinning when examined using an *L*-test. To address these issues, the data were reprocessed: they were indexed, integrated and scaled in *DIALS* and *XSCALE* to take advantage of 3D profile fitting (Clabbers *et al.*, 2018; Kabsch, 2010). Up to six high-damage frames were omitted from the end of each data set until no further increase in $CC_{1/2}$ (Karplus & Diederichs, 2012) could be achieved to 1.6 Å resolution. Exhaustive merging was attempted for 12 data sets. The merging results were evaluated based on the resultant completeness and $\langle I/\sigma(I)\rangle$. The selection criterion was such that the merged data had >90% completeness using the fewest crystals and the highest $\langle I/\sigma(I)\rangle$. The final merged data set was from six crystals integrated to 1.6 Å resolution with an overall completeness of 91.5% and an $\langle I/\sigma(I)\rangle$ of 3.3 (Table 1; Supplementary Table S1).

2.2. *ARCIMBOLDO_SHREDDER* in sequential mode

To accomplish fragment-based phasing from a homologous structure using *ARCIMBOLDO_SHREDDER*, a user-chosen homologue is first run through the *Phaser* rotation function. Positions from the peaks in the rotation search are clustered geometrically within a tolerance of 15° and each cluster is then used to systematically omit or extract segments from the template model. These segments are contiguous polyalanine fragments of variable length (Sammito *et al.*, 2014). The log-likelihood gain (LLG) scores of these fragments are then optimized by rotational analysis in *Phaser* (Storoni *et al.*, 2004). By comparing the LLG distribution for each sequentially ordered group of models of equal size, a single descriptor function, called the Shred-LLG function, is generated. Each point on the Shred-LLG function corresponds to a single residue and describes its individual contribution to the LLG score (Sammito *et al.*, 2014). Using this descriptor function, *ARCIMBOLDO_SHREDDER* then generates up to four models per rotation cluster. These are chosen by selecting peaks, plateaus, residues characterized by values above 75% of the maximum and values above the minimum peak height of the Shred-LLG function. These models are then run through independent *ARCIMBOLDO_LITE* searches, comprising both the *Phaser* location and refinement steps (McCoy *et al.*, 2005), and further trimming based on CC scores and successive rounds of density modification and main-chain autotracing with *SHELXE* until a final solution is reached (Supplementary Fig. S2). The parameters used for these *ARCIMBOLDO_SHREDDER* runs can be found in the supporting information (Section S1). These parameters, including the *SHELXE* inputs and fragment-generation options, can be varied for individual data sets, where data extrapolation can address a lack of completeness (Usón *et al.*, 2007). To inspire the best choice of these parameters, a tutorial describing our procedure for proteinase K structure determination by *ARCIMBOLDO* is available at http://chango.ibmb.csic.es/tutorial_microed.

2.3. *ARCIMBOLDO_SHREDDER* in spherical mode

In spherical mode (Millán *et al.*, 2018), *ARCIMBOLDO_SHREDDER* produces a set of compact, overlapping models starting from a distant homologue template that are run as a library in *ARCIMBOLDO_BORGES*. To increase the radius of convergence of this approach, additional degrees of freedom are given to the models, which are decomposed in rigid-body groups and subjected to refinement against the intensity-based likelihood rotation-function target (Read & McCoy, 2016) and again after they have been placed in the unit cell. This refinement is accomplished in *Phaser* with the *gyre* and *gimble* modes (McCoy *et al.*, 2018), although other modifications of the model relying on the experimental data can be performed, such as normal-mode deformation (McCoy

et al., 2013) or pruning to optimize the CC (Sheldrick & Gould, 1995) or LLG (Oeffner *et al.*, 2018). Combination of partial solutions representing parts of a general hypothesis for the target fold can be performed in reciprocal space with *ALIXE* (Millán, Sammito, Garcia-Ferrer *et al.*, 2015). A functional set of parameters used for *ARCIMBOLDO_SHREDDER* in a spherical mode run can be found in the supporting information (Section S2).

2.4. Homologues used as models for *ARCIMBOLDO_SHREDDER* and for molecular replacement

Using the sequence corresponding to the proteinase K structure deposited as PDB entry 5k7s, we searched for homologous structures using the *HHpred* server (Söding *et al.*, 2005). Model libraries were generated from structures with PDB codes 4dzt (B. L. Barnett, P. R. Green, L. C. Strickland, J. D. Oliver, T. Rydel & J. F. Sullivan, unpublished work), 5yl7 (Park *et al.*, 2018), 5jxg (Dahms *et al.*, 2016) and 1ga6 (Wlodawer *et al.*, 2001). Using *GESAMT* (Krissinel, 2012), the r.m.s.d. values for the models with PDB codes 4dzt, 5yl7, 5jxg and 1ga6 were 1.01, 1.43, 1.87 and 2.10 Å based on the alignment of 268, 245, 247 and 226 residues, respectively (Supplementary Table S2). Libraries of between 100 and 200 models were generated from these structures by *ARCIMBOLDO_SHREDDER* in spherical or sequential mode and were evaluated using *ARCIMBOLDO_BORGES*.

Table 1
MicroED structure of proteinase K determined by fragment-based phasing.

Phasing method	*ARCIMBOLDO_SHREDDER* (fragment library from PDB entry 4dzt)	*Phaser* (PDB entry 4dzt)
Data collection and processing		
No. of crystals	6	6
Total electron exposure (e Å^{-2})	0.86	0.86
Molecular weight (kDa)	28.9	28.9
Resolution (Å)	55.79–1.60 (1.657–1.600)	55.79–1.60 (1.657–1.600)
Space group	$P4_32_12$	$P4_32_12$
a, *b*, *c* (Å)	67.25, 67.25, 99.92	67.25, 67.25, 99.92
α, β, γ (°)	90, 90, 90	90, 90, 90
Total No. of reflections	194052	194052
No. of unique reflections	29058 (2506)	29058 (2506)
$CC_{1/2}$	0.912 (0.051)	0.912 (0.051)
$\langle I/\sigma(I)\rangle$	3.31	3.31
Completeness (%)	91.49 (66.19)	91.49 (66.19)
Multiplicity	6.68	6.68
Phasing		
Residues placed	175	
Fragments placed	13	
LLG	287.3	179
TFZ	20.4	19.6
CC (%)	23.31	
Refinement		
Resolution (Å)	55.79–1.60 (1.657–1.600)	55.79–1.60 (1.657–1.600)
R_{work} (%)	19.6	19.8
R_{free} (%)	23.3	23.7
RSCC	0.92	0.92
No. of residues	279	279
No. of protein atoms	2056	2038
No. of water molecules	122	138
No. of ligand atoms	2	2
Average *B* factor (Å^2)		
Overall	17.37	16.84
Protein	16.96	16.64
Water	18.89	19.72
Ligand	25.41	25.76
R.m.s.d., bonds (Å)	0.007	0.007
R.m.s.d., angles (°)	0.9	0.9
Ramachandran statistics		
Outliers (%)	0.36	0.36
Favored (%)	97.11	97.11
Clashscore	6.75	5.28

2.5. Refinement and analysis of the proteinase K structure determined by *ARCIMBOLDO_SHREDDER* in sequential mode with PDB entry 4dzt

An initial *ARCIMBOLDO_SHREDDER* solution determined from a set of 13 traced segments derived from the model with PDB code 4dzt and totaling 175 residues was refined in *Phenix* using *phenix.refine* (Liebschner *et al.*, 2019; Afonine *et al.*, 2012). Subsequent visualization and model building were performed in *Coot* (Emsley *et al.*, 2010). Fragment chains were first connected by building along the full length of the protein backbone. Side chains were then assigned unambiguously, and finally ordered waters were placed. The reported r.m.s.d. values were computed by secondary-structure matching superposition (Krissinel & Henrick, 2004) using *Super* in *PyMOL* (version 1.8; Schrödinger) or *GESAMT* (Krissinel, 2012), considering only core C^{α} atoms. Placed fragments were evaluated against our final, fully refined model by calculating the LLG with *Phaser* and the initial CC and weighted mean phase error (wMPE) with *SHELXE*.

3. Results

3.1. Using *ARCIMBOLDO_SHREDDER* in sequential mode to phase proteinase K

ARCIMBOLDO_SHREDDER was implemented to phase MicroED data using a 1.6 Å resolution data set for proteinase K that was collected as described previously (Hattne *et al.*, 2018) and used in part in the determination of the structures deposited as PDB entries 5k7s and 6cl7. This data set (Table 1), which is 91.49% complete to 1.6 Å resolution with an overall $\langle I/\sigma(I)\rangle$ of 3.3, was suitable for molecular replacement using a known crystal structure of proteinase K (Table 1). To evaluate fragment-based phasing, we chose a homologue of proteinase K with a sequence identity of 40% (PDB entry 4dzt). This structure has a 268-atom C^{α} r.m.s.d. of 1.01 Å to the proteinase K structure determined from these data, as calculated by *GESAMT*. A library of models was generated from this

starting template based on rotational analysis by *Phaser* using electron scattering factors, with top-scoring clusters of rotation solutions being used to perform an analysis of the effect of omitting continuous spans of the structure. Such omit fragments were generated by extracting 10–20-residue contiguous segments every four residues for the length of the protein, resulting in a total of 759 polyalanine fragments (Fig. 1*a*). The global evaluation of such fragments is performed in terms of a Shred-LLG function, which through joint scoring of the results obtained using all of these fragments assesses the local accuracy of the initial template (described in Section 2). The located models were input to *SHELXE* for autotracing expansion as outlined above. This implementation resulted in a correct output model composed of 175 residues encompassing 13 chains obtained from a solution characterized by a *Phaser* rotation LLG score of 287.30, a *Phaser* translation *Z*-score (TFZ) of 20.40 and a *SHELXE* final CC of 23.31% (Fig. 1*b*).

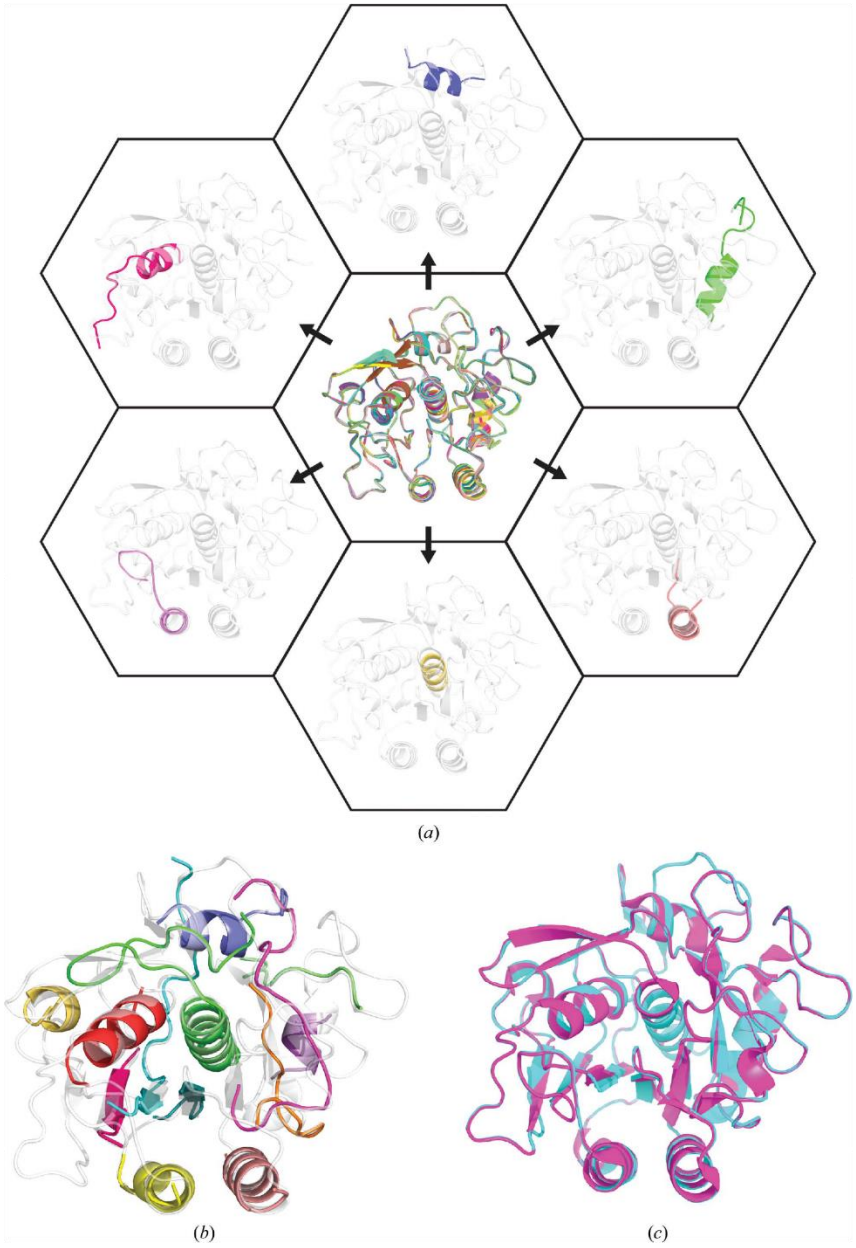


Figure 1
Schematic of fragment generation and structure determination of proteinase K using *ARCIMBOLDO_SHREDDER*. (*a*) At the center, an overlay of all 759 fragments sequentially generated from the template model (PDB entry 4dzt) is shown. Examples of individual fragments derived from the model template are shown extracted out of the center model in the context of the final structure of proteinase K. (*b*) The output solution from *ARCIMBOLDO_SHREDDER*, composed of the 13 placed individual fragments (colored chains), is shown overlaid with the final structure of proteinase K. (*c*) The final structure of proteinase K determined with *ARCIMBOLDO_SHREDDER* (PDB entry 6v8r; cyan) overlaid with the previously determined MicroED structure (PDB entry 5k7s; pink) gives a C^{α} r.m.s.d. of 0.12 Å.

The output model traced by *SHELXE* was composed of fragments from seven α-helices and three β-strands as well as a few loop regions that are conserved between proteinase K and the homologue. Missing structural elements appeared clearly as positive difference-map peaks in initial refinements (Fig. 2*a*), and subsequent rounds of manual model building and refinement revealed missing loops, side chains and ordered waters (Figs. 2*b* and 2*c*). The refined structure solution contained 279 unambiguously assigned residues and 122 ordered waters, and had a final R_{work} of 19.6% and R_{free} of 23.3%. Omit maps computed from the refined *ARCIMBOLDO_SHREDDER* solution or the solution determined by *Phaser* using PDB entry 4dzt as a search model, having deleted from each the sixth helix corresponding to residues 223–237 in proteinase K, resulted in positive difference density that outlined not only the location of the helix, but also revealed a continuous map at 3.0σ matching the appropriate side

chains for all but four residues in the helix (Fig. 3*a*). Additionally, omit maps of the two coordinated calcium ions give positive difference-map peaks at 17.69σ and 11.73σ (Fig. 3*b*), and the omit map for the removal of an ordered water molecule gives rise to a 6.95σ positive-density peak (Fig. 3*c*). Placement of bound ions and waters satisfied the difference map density and resulted in a decrease in the *R* factors.

3.2. Comparison of the solution from *ARCIMBOLDO_SHREDDER* with the known proteinase K structure

The 13 homologue fragments placed by *Phaser* overlay well with the final structure of proteinase K (Fig. 1*b*). The structure determined using model fragments from this *ARCIMBOLDO_SHREDDER* run is nearly identical to the previously determined MicroED structure of proteinase K (Hattne *et al.*, 2018), with a C^{α} r.m.s.d. of 0.12 Å (Fig. 1*c*). The input model aligned with the known structure of proteinase K gives a C^{α} r.m.s.d. of 0.65 Å when aligning 232 atoms and yields a correct solution when used for molecular replacement. Notably, the *Phaser* LLG and TFZ scores are lower for this solution (179 and 19.6, respectively) compared with the initial scores for the *ARCIMBOLDO_SHREDDER* solution, showing that the fragments placed reflect accurate structural components that are present in the final structure.

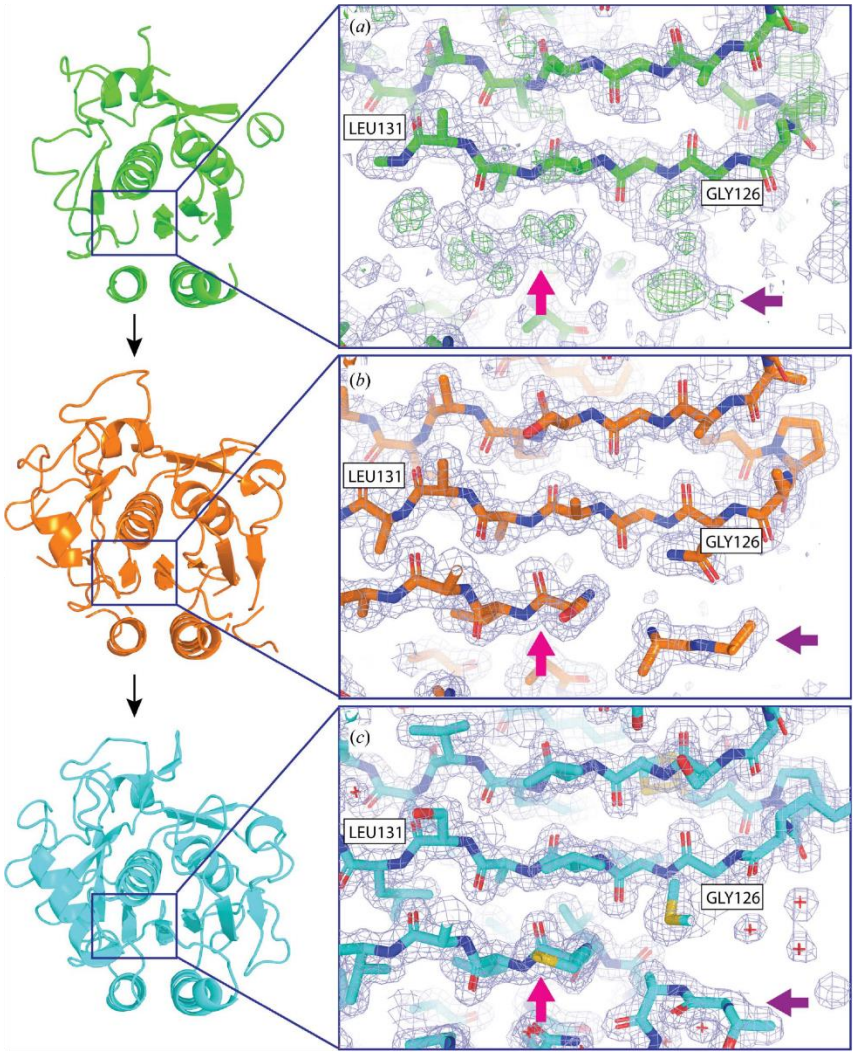


Figure 2
Building of missing structural elements. Starting from an *ARCIMBOLDO*-generated solution, advancement in refinement is shown in stages including (*a*) the initial *ARCIMBOLDO* output, (*b*) an intermediate stage of building and (*c*) the final structure (PDB entry 6v8r). Pink and purple arrows indicate positions in the map where structural elements, a β-sheet (pink) and a loop region (purple), were built into the positive difference-map peak density seen in the initial map.

3.3. Use of spherical fragment generation for structure determination

To further evaluate the potential of model improvement against the experimental MicroED data, we also attempted phasing using the recently developed spherical mode in *ARCIMBOLDO_SHREDDER* (Millán *et al.*, 2018). This mode is particularly appropriate for more structurally distant homologs that have an overall conserved fold and where deviations from the final model are distributed isotropically in Cartesian space. In such a case, simply removing the regions of largest deviation or extruding contiguous fragments, as is performed in sequential mode, may not be sufficient to obtain a phasing solution. Instead, in spherical mode, small compact fragments of pre-defined size are extracted from the distant homologue, given degrees of freedom and searched for independently, and subsequently combined in reciprocal space (Millán *et al.*, 2020). The spherical mode in *ARCIMBOLDO_SHREDDER* selects the size of its models based on the eLLG score. Given the data resolution, the expected r.m.s.d. of the models and a target eLLG (by default 30), the appropriate size for the models is derived. All of the models produced in the run were within a range of ten residues of such a value. The models ranged in size between 44 and 48 residues. Three homologues with

various degrees of sequence identity and structural similarity (PDB entries 5yl7, 5jxg and 1ga6), which did not produce viable solutions in sequential mode, were evaluated using *ARCIMBOLDO_SHREDDER* in spherical mode to attempt phasing of the proteinase K MicroED data. The results from this attempt are summarized in Table 2 and demonstrate the determination of correct partial solutions using this method. However, while solutions are identified during the search, the extension of these partial solutions in *SHELXE* can be notably more difficult for MicroED data than for X-ray data. This may be owing in part to the high initial mean phase errors (68–76°) associated with the placement of these fragments. This in turn will require improved algorithms, implemented in *SHELXE*, that take into account the unique aspects of electron scattering.

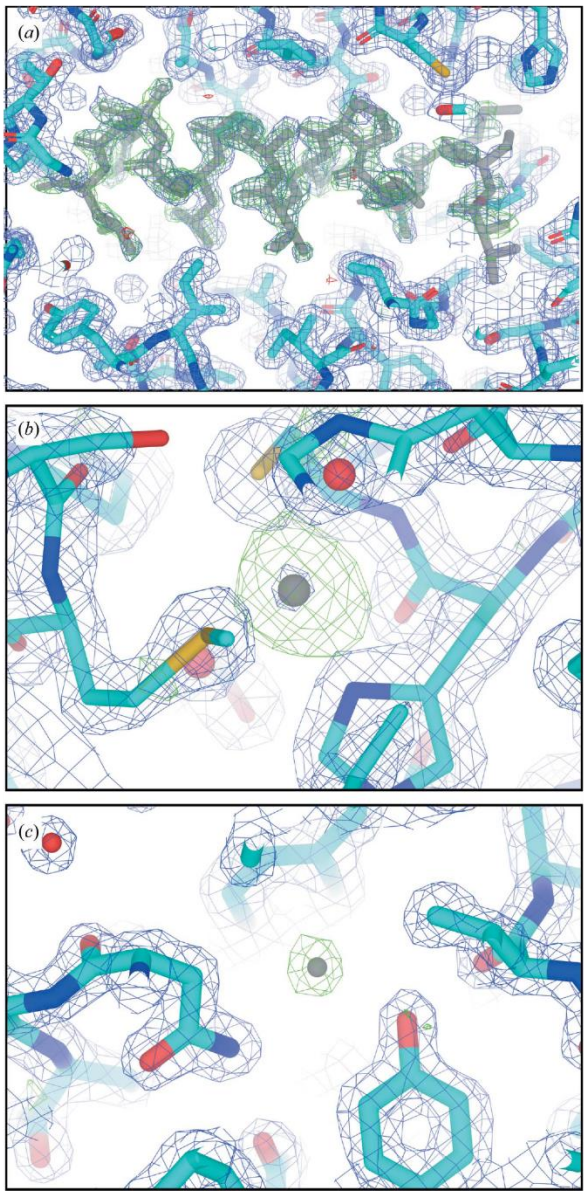


Figure 3
Representative omit maps. (*a*) Omit map for the *ARCIMBOLDO_SHREDDER* solution generated after removal of the sixth α-helix comprised of residues 223–237 (shown in gray). (*b*) Omit map generated after the removal of one of the calcium ions coordinated by the structure (shown in gray). (*c*) Omit map generated after the removal of one representative water molecule (shown in gray). The blue mesh is the $2mF_o - F_c$ map contoured at 1.5σ and the green mesh is the $F_o - F_c$ map contoured at 3.0σ.

Table 2
Results for fragment-based phasing of the MicroED data set.

Target: 5k7s	Identity†	R.m.s.d. (*GESAMT*) (Å)/No. of residues	No. of correct solutions/ total solutions	Best wMPE (°)
5yl7	0.310	1.43/245	39/394	68.8
5jxg	0.193	1.87/247	6/375	72.4
1ga6	0.208	2.10/226	2/632	76.0

† Identity is denoted as a fraction, where 1 represents perfect identity.

Table 3
Results for fragment-based phasing of the X-ray data set associated with PDB entry 4woc.

Target: 5k7s	Identity†	R.m.s.d. (*GESAMT*) (Å)/No. of residues	No. of correct solutions/ total solutions	Best wMPE (°)
5yl7	0.310	1.43/245	25/579	64.5
5jxg	0.193	1.87/247	10/486	71.0
1ga6	0.208	2.10/226	4/595	76.5

† Identity is denoted as a fraction, where 1 represents perfect identity.

3.4. Comparing the performance of *ARCIMBOLDO_SHREDDER* in spherical mode against both MicroED and X-ray diffraction data using more distant homologues of proteinase K

The same homologues used for phasing the MicroED data in the experiments described in Section 3.4 and Table 2 were used to phase an X-ray data set from an isostructural form of proteinase K: PDB entry 4woc (Guo *et al.*, 2015). Attempts at phasing using the X-ray data set are summarized in Table 3. With these data, fragment placement succeeds in generating correct placements with all models tested against both data sets. As expected, these trials yield better minimum wMPEs with X-ray data than with MicroED data. The overall trend in both cases favored the placement of fragments from structures with higher similarity to the known solution. For example, from the library of models generated by *ARCIMBOLDO_SHREDDER* from the proteinase K structure deposited as PDB entry 5yl7, 39 fragments were placed and yielded correct solutions in the MicroED data set. The best showed a weighted mean phase error (wMPE) of 68.8°. In the most extreme of cases, fragments generated from a pepstatin-insensitive carboxyl proteinase from *Pseudomonas* sp. 101 (PSCP) deposited as PDB entry 1ga6 (with only 21% sequence identity to the target) facilitated the placement of two correct fragments as solutions, with the best having a wMPE of 76°. These tests collectively demonstrate the

promise of some distant homologues for the accurate placement of fragments using MicroED data.

3.5. Phasing with idealized helices as search models in *ARCIMBOLDO_LITE*

Searches using idealized helix models ranging in size from three to 18 alanine residues were attempted on the MicroED data set (PDB entry 6v8r) and an example X-ray proteinase K data set (PDB entry 4woc). The parameters for these runs were set to the defaults for *ARCIMBOLDO_LITE* except for the implementation of electron scattering factors in *Phaser* for the 6v8r data set (supporting information, Section S3). None of these runs with either data set produced a solution with a wMPE of lower than 85°, indicating that no viable solution was identified by this method. This result is not surprising given that the helix fragments represent a very small scattering fraction of the full structure.

4. Discussion and conclusions

As the field of MicroED continues to expand, a growing number of novel structures may present phasing hurdles. Given that experimental phasing remains a challenge in MicroED, it is important to explore other ways to overcome the phase problem beyond direct methods and molecular replacement. To date, more than a dozen *ab initio* structures determined by direct methods from MicroED data have been deposited in the PDB, in comparison to several dozen structures determined by conventional molecular replacement with resolutions between 1.2 and 3 Å (Rodriguez & Gonen, 2016). Of the set determined by molecular replacement, approximately 13 are in some way novel, although many of these rely on highly similar search models determined by X-ray diffraction. The relatively low number of completely novel structures is due in part to the challenges associated with the experimental phasing of MicroED data. Given the smaller difference in scattering between heavy and light atoms in electron diffraction compared with X-ray diffraction, experimental phasing by isomorphous replacement remains undemonstrated and, at least for 2D crystals, might be intractable (Ceska & Henderson, 1990).

Many of the structures determined by MicroED to date have resolutions (1.2–2 Å) appropriate for attempts at phasing by *ARCIMBOLDO* or other fragment-based and *ab initio* phasing approaches. Fragment-based approaches are typically less restrictive than conventional molecular-replacement methods for phasing and have been demonstrated in electron crystallography of 2D and 3D crystals using image data combined with electron diffraction data (Wisedchaisri & Gonen, 2011). Requirements that are important for the success of structure determination by *ARCIMBOLDO* from both MicroED and X-ray diffraction data include (i) high completeness, (ii) data quality and perhaps resolution and (iii) models similar to the target structure from which fragments are derived. When these criteria are met, conventional molecular replacement is often also successful. For instance, phasing of polymeric amyloid peptide assemblies has been achieved using idealized β-strands that closely match the final geometry of the polypeptide structure (Rodriguez *et al.*, 2015).

Overall, while fragment placement succeeds with a variety of libraries, even those with distant homology to the known target structure for MicroED data, the extension of partial solutions remains a challenge. This may result in part from the nature of the maps, which represent a screened Coulomb potential rather than electron density, or from inherent features of the data. Additional limitations are likely to be present in MicroED maps. For example, some crystals may suffer from orientation bias on an EM grid, and this in turn may lead to a missing cone of information which can persist despite attempts at merging multiple data sets (Nannenga, Shi, Hattne *et al.*, 2014). Problems also arise from inaccuracies in the estimation of standard errors of the experimental data. The strong effects of anisotropy (Strong *et al.*, 2006) and the partial effects of directional lack of completeness, along with potential absorption and dynamic scattering (Cowley & Moodie, 1957; Dorset *et al.*, 1992; Glaeser & Downing, 1993), can add to a uniquely deleterious effect on maps and thus may influence density modification and autotracing. Despite these, density modification has been demonstrated for electron diffraction (Wisedchaisri & Gonen, 2011). The use of electron scattering form factors, data filtering by information content (Read *et al.*, 2020) and anisotropy correction are expected to be beneficial for these approaches, both during direct-method protocols and with fragment-based approaches. Future corrections implemented during data reduction may ameliorate these effects. Our present observations suggest that *ARCIMBOLDO* may be successful in identifying phasing solutions for MicroED data from structures of distantly related homologues. Various modes of search-model definition, be it linear fragments, structures with omitted segments or spherical regions of structures, could yield solutions with varying success.

After years of successful application to X-ray crystallographic data, this study demonstrates the utility of fragment-based phasing methods and *ARCIMBOLDO* with MicroED data. Our ability to determine a known structure using small structural fragments derived from a distantly related homologue opens the possibility of the *de novo* determination of structures by MicroED. This demonstration follows several reports of fragment-based phasing or phase extension for electron diffraction data (Wisedchaisri & Gonen, 2011). Phasing methods that employ the use of fragments are gaining in popularity for the determination of X-ray structures. An example of these is *AMPLE* (Bibby *et al.*, 2012; Rigden *et al.*, 2018), which in turn uses *ROSETTA* (Qian *et al.*, 2007), *QUARK* (Keegan *et al.*, 2015) or *CONCOORD* (de Groot *et al.*, 1997) to generate models. Some of these programs offer the possibility of generating *ab initio* fragments derived from the target sequence, for example *FRAGON* (Jenkins, 2018) and *FRAP* (Shrestha & Zhang, 2015). While the limited substrate scope of our study precludes conclusions on the general application of fragment-based phasing to MicroED data, our results demonstrate that fragment-based phasing is

advantageous when applied to MicroED data with a resolution that is too poor for direct methods. In such cases, *ARCIMBOLDO_SHREDDER* and perhaps other fragment-based phasing programs offer a potential solution to a problem that may otherwise remain unsolved.

5. Related literature

The following references are cited in the supporting information for this article: Arndt & Wonacott (1977), Nannenga & Gonen (2016).

Acknowledgements

We thank Drs Johan Hattne (HHMI) and Duilio Cascio (UCLA) for discussions and helpful analysis of this problem. Author contributions are as follows. JAR and IU directed the research. LSR, CM and JM generated fragment libraries and performed phasing with *ARCIMBOLDO*, refined initial fragment models and performed post-phasing analyses. CM, IU and RJB developed and evaluated MicroED-specific features in the *ARCIMBOLDO* framework. MWM collected and processed the data. MRS assisted with model refinement and phasing. All authors helped to write and provided critical feedback on the article.

Funding information

This work was performed as part of STROBE, an NSF Science and Technology Center, through grant DMR-1548924. This work was also supported by DOE grant DE-FC02-02ER63421 and NIH–NIGMS grants R35 GM128867 and P41GM136508. LSR is supported by USPHS National Research Service Award 5T32GM008496. RJB received a fellowship from FAPESP (16/24191-8 and 17/13485-3). CM is grateful to MICINN for her BES-2015-71397 scholarship associated with the Structural Biology Maria de Maeztu Unit of Excellence. This work was supported by grants BIO2015-64216-P, PGC2018-101370-B-100 and MDM2014-0435-01 (the Spanish Ministry of Economy and Competitiveness) and Generalitat de Catalunya (2017SGR-1192). JAR is supported as a Searle Scholar, a Pew Scholar and a Beckman Young Investigator. The Gonen laboratory is supported by the Howard Hughes Medical Institute; this work was also partially supported by the Janelia Research Campus Visitor Exchange Program.

References

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* D**68**, 352–367.

Arndt, U. W. & Wonacott, A. J. (1977). *The Rotation Method in Crystallography*. Amsterdam: North-Holland.

Bibby, J., Keegan, R. M., Mayans, O., Winn, M. D. & Rigden, D. J. (2012). *Acta Cryst.* D**68**, 1622–1631.

Bragg, W. H. & Bragg, W. L. (1913). *Nature*, **91**, 557.

Burmester, C. & Schroeder, R. R. (1997). *Scanning Microsc.* **11**, 332–334.

Ceska, T. A. & Henderson, R. (1990). *J. Mol. Biol.* **213**, 539–560.

Clabbers, M. T. B., Gruene, T., Parkhurst, J. M., Abrahams, J. P. & Waterman, D. G. (2018). *Acta Cryst.* D**74**, 506–518.

Colliex, C., Cowley, J. M., Dudarev, S. L., Fink, M., Gjønnes, J., Hilderbrandt, R., Howie, A., Lynch, D. F., Peng, L. M., Ren, G., Ross, A. W., Smith, V. H., Spence, J. C. H., Steeds, J. W., Wang, J., Whelan, M. J. & Zvyagin, B. B. (2006). *International Tables for Crystallography*, Vol. *C*, edited by E. Prince, pp. 259–429. Chester: International Union of Crystallography.

Cowley, J. M. & Moodie, A. F. (1957). *Acta Cryst.* **10**, 609–619.

Cruz, M. J. de la, Hattne, J., Shi, D., Seidler, P., Rodriguez, J., Reyes, F. E., Sawaya, M. R., Cascio, D., Weiss, S. C., Kim, S. K., Hinck, C. S., Hinck, A. P., Calero, G., Eisenberg, D. & Gonen, T. (2017). *Nat. Methods*, **14**, 399–402.

Dahms, S. O., Arciniega, M., Steinmetzer, T., Huber, R. & Than, M. E. (2016). *Proc. Natl Acad. Sci. USA*, **113**, 11196–11201.

Dorset, D. L., Tivol, W. F. & Turner, J. N. (1992). *Acta Cryst.* A**48**, 562–568.

Doyle, P. A. & Turner, P. S. (1968). *Acta Cryst.* A**24**, 390–397.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Fujinaga, M. & Read, R. J. (1987). *J. Appl. Cryst.* **20**, 517–521.

Gemmi, M., Mugnaioli, E., Gorelik, T. E., Kolb, U., Palatinus, L., Boullay, P., Hovmöller, S. & Abrahams, J. P. (2019). *ACS Cent. Sci.* **5**, 1315–1329.

Genderen, E. van, Clabbers, M. T. B., Das, P. P., Stewart, A., Nederlof, I., Barentsen, K. C., Portillo, Q., Pannu, N. S., Nicolopoulos, S., Gruene, T. & Abrahams, J. P. (2016). *Acta Cryst.* A**72**, 236–242.

Glaeser, R. M. (1999). *J. Struct. Biol.* **128**, 3–14.

Glaeser, R. M. & Downing, K. H. (1993). *Ultramicroscopy*, **52**, 478–486.

Gonen, T., Cheng, Y., Sliz, P., Hiroaki, Y., Fujiyoshi, Y., Harrison, S. C. & Walz, T. (2005). *Nature*, **438**, 633–638.

Groot, B. L. de, van Aalten, D. M. F., Scheek, R. M., Amadei, A., Vriend, G. & Berendsen, H. J. C. (1997). *Proteins*, **29**, 240–251.

Guo, F., Zhou, W., Li, P., Mao, Z., Yennawar, N. H., French, J. B. & Huang, T. J. (2015). *Small*, **11**, 2733–2737.

Hattne, J., Reyes, F. E., Nannenga, B. L., Shi, D., de la Cruz, M. J., Leslie, A. G. W. & Gonen, T. (2015). *Acta Cryst.* A**71**, 353–360.

Hattne, J., Shi, D., Glynn, C., Zee, C.-T., Gallagher-Jones, M., Martynowycz, M. W., Rodriguez, J. A. & Gonen, T. (2018). *Structure*, **26**, 759–766.

Hauptman, H. (1986). *Science*, **233**, 178–183.

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.

Jansen, J., Tang, D., Zandbergen, H. W. & Schenk, H. (1998). *Acta Cryst.* A**54**, 91–101.

Jenkins, H. T. (2018). *Acta Cryst.* D**74**, 205–214.

Jones, C. G., Martynowycz, M. W., Hattne, J., Fulton, T. J., Stoltz, B. M., Rodriguez, J. A., Nelson, H. M. & Gonen, T. (2018). *ACS Cent. Sci.* **4**, 1587–1592.

Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.

Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.

Keegan, R. M., Bibby, J., Thomas, J., Xu, D., Zhang, Y., Mayans, O., Winn, M. D. & Rigden, D. J. (2015). *Acta Cryst.* D**71**, 338–343.

Kovalevskiy, O., Nicholls, R. A., Long, F., Carlon, A. & Murshudov, G. N. (2018). *Acta Cryst.* D**74**, 215–227.

Krissinel, E. (2012). *J. Mol. Biochem.* **1**, 76–85.

Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). *Acta Cryst.* D**75**, 861–877.

Martynowycz, M. W., Hattne, J. & Gonen, T. (2020). *Structure*, **28**, 458–464.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). *Acta Cryst.* D**61**, 458–464.
McCoy, A. J., Nicholls, R. A. & Schneider, T. R. (2013). *Acta Cryst.* D**69**, 2216–2225.
McCoy, A. J., Oeffner, R. D., Millán, C., Sammito, M., Usón, I. & Read, R. J. (2018). *Acta Cryst.* D**74**, 279–289.
Medina, A., Triviño, J., Borges, R. J., Millán, C., Usón, I. & Sammito, M. D. (2020). *Acta Cryst.* D**76**, 193–208.
Millán, C., Jiménez, E., Schuster, A., Diederichs, K. & Usón, I. (2020). *Acta Cryst.* D**76**, 209–220.
Millán, C., Sammito, M., Garcia-Ferrer, I., Goulas, T., Sheldrick, G. M. & Usón, I. (2015). *Acta Cryst.* D**71**, 1931–1945.
Millán, C., Sammito, M. & Usón, I. (2015). *IUCrJ*, **2**, 95–105.
Millán, C., Sammito, M. D., McCoy, A. J., Nascimento, A. F. Z., Petrillo, G., Oeffner, R. D., Domínguez-Gil, T., Hermoso, J. A., Read, R. J. & Usón, I. (2018). *Acta Cryst.* D**74**, 290–304.
Nannenga, B. L. & Gonen, T. (2016). *Curr. Opin. Struct. Biol.* **40**, 128–135.
Nannenga, B. L. & Gonen, T. (2019). *Nat. Methods*, **16**, 369–379.
Nannenga, B. L., Shi, D., Hattne, J., Reyes, F. E. & Gonen, T. (2014). *eLife*, **3**, e03600.
Nannenga, B. L., Shi, D., Leslie, A. G. W. & Gonen, T. (2014). *Nat. Methods*, **11**, 927–930.
Oeffner, R. D., Afonine, P. V., Millán, C., Sammito, M., Usón, I., Read, R. J. & McCoy, A. J. (2018). *Acta Cryst.* D**74**, 245–255.
Park, H. J., Lee, C. W., Kim, D., Do, H., Han, S. J., Kim, J. E., Koo, B. H., Lee, J. H. & Yim, J. H. (2018). *PLoS One*, **13**, e0191740.
Pröpper, K., Meindl, K., Sammito, M., Dittrich, B., Sheldrick, G. M., Pohl, E. & Usón, I. (2014). *Acta Cryst.* D**70**, 1743–1757.
Purdy, M. D., Shi, D., Chrustowicz, J., Hattne, J., Gonen, T. & Yeager, M. (2018). *Proc. Natl Acad. Sci. USA*, **115**, 13258–13263.
Qian, B., Raman, S., Das, R., Bradley, P., McCoy, A. J., Read, R. J. & Baker, D. (2007). *Nature*, **450**, 259–264.
Read, R. J. & McCoy, A. J. (2016). *Acta Cryst.* D**72**, 375–387.
Read, R. J., Oeffner, R. D. & McCoy, A. J. (2020). *Acta Cryst.* D**76**, 238–247.
Rigden, D. J., Thomas, J. M. H., Simkovic, F., Simpkin, A., Winn, M. D., Mayans, O. & Keegan, R. M. (2018). *Acta Cryst.* D**74**, 183–193.
Rodríguez, D., Sammito, M., Meindl, K., de Ilarduya, I. M., Potratz, M., Sheldrick, G. M. & Usón, I. (2012). *Acta Cryst.* D**68**, 336–343.
Rodríguez, D. D., Grosse, C., Himmel, S., González, C., de Ilarduya, I. M., Becker, S., Sheldrick, G. M. & Usón, I. (2009). *Nat. Methods*, **6**, 651–653.
Rodriguez, J. A. & Gonen, T. (2016). *Methods Enzymol.* **579**, 369–392.
Rodriguez, J. A., Ivanova, M. I., Sawaya, M. R., Cascio, D., Reyes, F., 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. (2015). *Nature*, **525**, 486–490.
Sammito, M., Meindl, K., de Ilarduya, I. M., Millán, C., Artola-Recolons, C., Hermoso, J. A. & Usón, I. (2014). *FEBS J.* **281**, 4029–4045.
Sammito, M., Millán, C., Frieske, D., Rodríguez-Freire, E., Borges, R. J. & Usón, I. (2015). *Acta Cryst.* D**71**, 1921–1930.
Sammito, M., Millán, C., Rodríguez, D. D., de Ilarduya, I. M., Meindl, K., De Marino, I., Petrillo, G., Buey, R. M., de Pereda, J. M., Zeth, K., Sheldrick, G. M. & Usón, I. (2013). *Nat. Methods*, **10**, 1099–1101.
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.
Sheldrick, G. M. (2010). *Acta Cryst.* D**66**, 479–485.
Sheldrick, G. M. (2015*a*). *Acta Cryst.* C**71**, 3–8.
Sheldrick, G. M. (2015*b*). *Acta Cryst.* A**71**, 3–8.
Sheldrick, G. M., Gilmore, C. J., Hauptman, H. A., Weeks, C. M., Miller, M. & Usón, I. (2012). *International Tables for Crystallography*, Vol. *F*, edited by E. Arnold, D. M. Himmel & M. G. Rossmann, pp. 413–429. Chester: International Union of Crystallography.
Sheldrick, G. M. & Gould, R. O. (1995). *Acta Cryst.* B**51**, 423–431.
Shi, D., Nannenga, B. L., Iadanza, M. G. & Gonen, T. (2013). *eLife*, **2**, e01345.
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.
Shrestha, R. & Zhang, K. Y. J. (2015). *Acta Cryst.* D**71**, 304–312.
Söding, J., Biegert, A. & Lupas, A. N. (2005). *Nucleic Acids Res.* **33**, W244–W248.
Storoni, L. C., McCoy, A. J. & Read, R. J. (2004). *Acta Cryst.* D**60**, 432–438.
Strong, M., Sawaya, M. R., Wang, S., Phillips, M., Cascio, D. & Eisenberg, D. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 8060–8065.
Subramanian, G., Basu, S., Liu, H., Zuo, J.-M. & Spence, J. C. H. (2015). *Ultramicroscopy*, **148**, 87–93.
Thorn, A. & Sheldrick, G. M. (2013). *Acta Cryst.* D**69**, 2251–2256.
Usón, I. & Sheldrick, G. M. (1999). *Curr. Opin. Struct. Biol.* **9**, 643–648.
Usón, I. & Sheldrick, G. M. (2018). *Acta Cryst.* D**74**, 106–116.
Usón, I., Stevenson, C. E. M., Lawson, D. M. & Sheldrick, G. M. (2007). *Acta Cryst.* D**63**, 1069–1074.
Wennmacher, J. T. C., Zaubitzer, C., Li, T., Bahk, Y. K., Wang, J., van Bokhoven, J. A. & Gruene, T. (2019). *Nat. Commun.* **10**, 3316.
Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Cryst.* D**67**, 235–242.
Wisedchaisri, G. & Gonen, T. (2011). *Structure*, **19**, 976–987.
Wlodawer, A., Li, M., Dauter, Z., Gustchina, A., Uchida, K., Oyama, H., Dunn, B. M. & Oda, K. (2001). *Nat. Struct. Biol.* **8**, 442–446.
Xu, H., Lebrette, H., Clabbers, M. T. B., Zhao, J., Griese, J. J., Zou, X. & Högbom, M. (2019). *Sci. Adv.* **5**, eaax4621.
Xu, H., Lebrette, H., Yang, T., Srinivas, V., Hovmöller, S., Högbom, M. & Zou, X. (2018). *Structure*, **26**, 667–675.
Yonekura, K., Kato, K., Ogasawara, M., Tomita, M. & Toyoshima, C. (2015). *Proc. Natl Acad. Sci. USA*, **112**, 3368–3373.
Zhou, H., Luo, Z. & Li, X. (2019). *J. Struct. Biol.* **205**, 59–64.

Chapter 3

This chapter is based on the published article “Fragment-based ab initio phasing of peptidic nanocrystals by MicroED.” [1]

Summary

This article describes fragment-based phasing techniques for *ab initio* structure determination of novel peptide structures from MicroED. As previously described, the phase problem is particularly limiting in the case of MicroED data. Without access to isomorphous replacement or anomalous dispersion methods becomes limited to molecular replacement, which demands a highly similar model from a previously determined structure, or direct methods, which requires atomic resolution data. In this work, a difficult set of amyloid peptide structural targets were interrogated by MicroED and phased using the program ARCIMBOLDO_BORGES. The input for ARCIMBOLDO was custom libraries of peptide fragments, some composed of previously determined structures and others computed *de novo* using PyRosetta when no other models proved sufficient. These fragment libraries were designed to sample structural variability while precluding model bias by leveraging ARCIMBOLDO’s ability to assess a large collection of structural hypotheses in parallel. Several novel peptide structures were determined using this method including segments of the functional mammalian prion, CPEB3 (QIGLAQTQ), the human amyloid protein LECT2 (GSTVYAPFT), and the human zinc finger protein (ZFP) 292 (FRNWQAYMQ). Fragment-based phasing using custom libraries as demonstrated here lays the foundation for a more generalized phasing solution, with limited model bias, for a wider set of chemical structures that could employ large libraries of chemical fragments.

(1) Richards, L. S.; Flores, M. D.; Millán, C.; Glynn, C.; Zee, C.-T.; Sawaya, M. R.; Gallagher-Jones, M.; Borges, R. J.; Usón, I.; Rodriguez, J. A. Fragment-Based Ab Initio Phasing of Peptidic Nanocrystals by MicroED. ACS Bio Med Chem Au 2023. https://doi.org/10.1021/acsbiomedchemau.2c00082.

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

pubs.acs.org/biomedchemau

Article

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

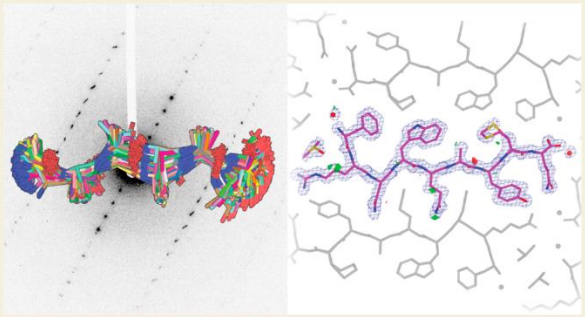
Logan S. Richards,[#] Maria D. Flores,[#] Claudia Millán,[#] Calina Glynn, Chih-Te Zee, Michael R. Sawaya, Marcus Gallagher-Jones, Rafael J. Borges, Isabel Usón,* and Jose A. Rodriguez*

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

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

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



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

INTRODUCTION

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

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

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

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

Table 1

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

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

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

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

RESULTS

High-Resolution MicroED Data from Peptide Nanocrystals

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

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

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

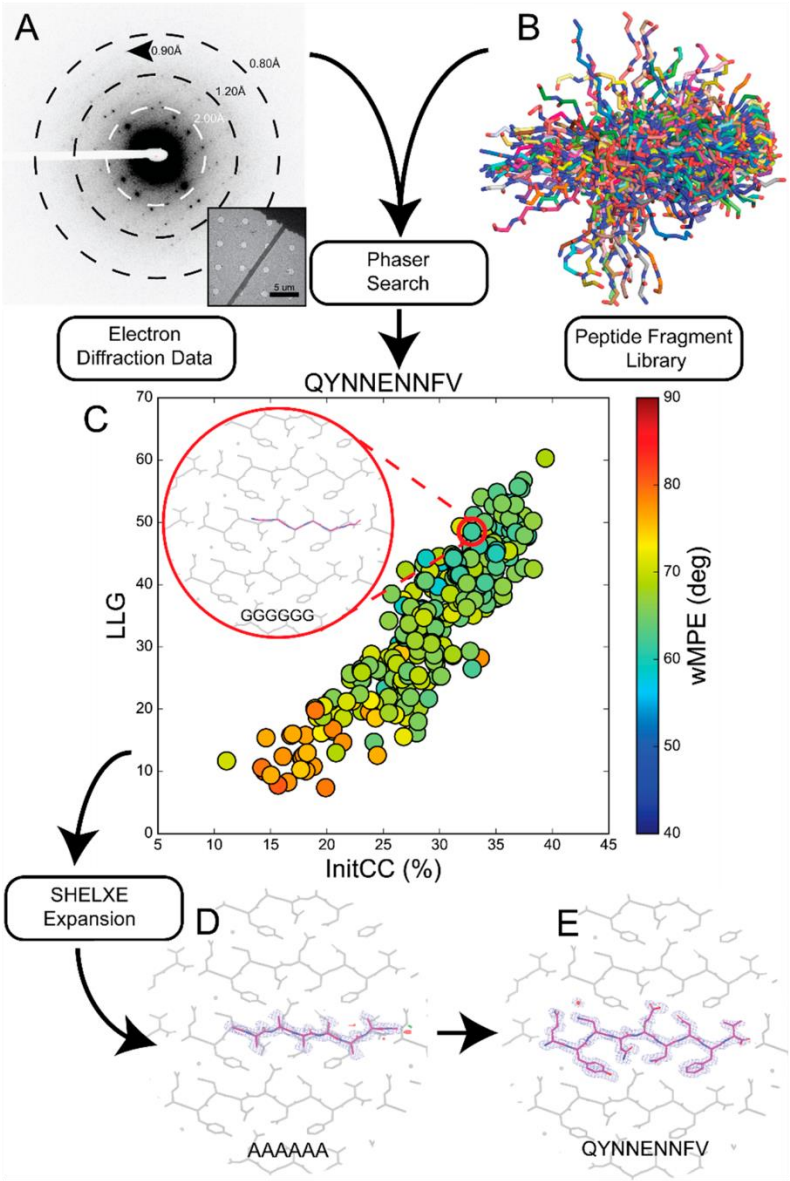


Figure 1. Workflow for using peptide fragments for phasing MicroED data. A) Electron diffraction pattern reaching atomic resolution for QYNNENNFV. Rings designate resolution ranges while arrowhead designates highest resolution spot. B) All fragments comprising the polyglycine hexapeptide library aligned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

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

Phasing with ARCIMBOLDO Using General Local Fold Libraries

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

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

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

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

Broad, Knowledge-Based Libraries to Phase Nanocrystals

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

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

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

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

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

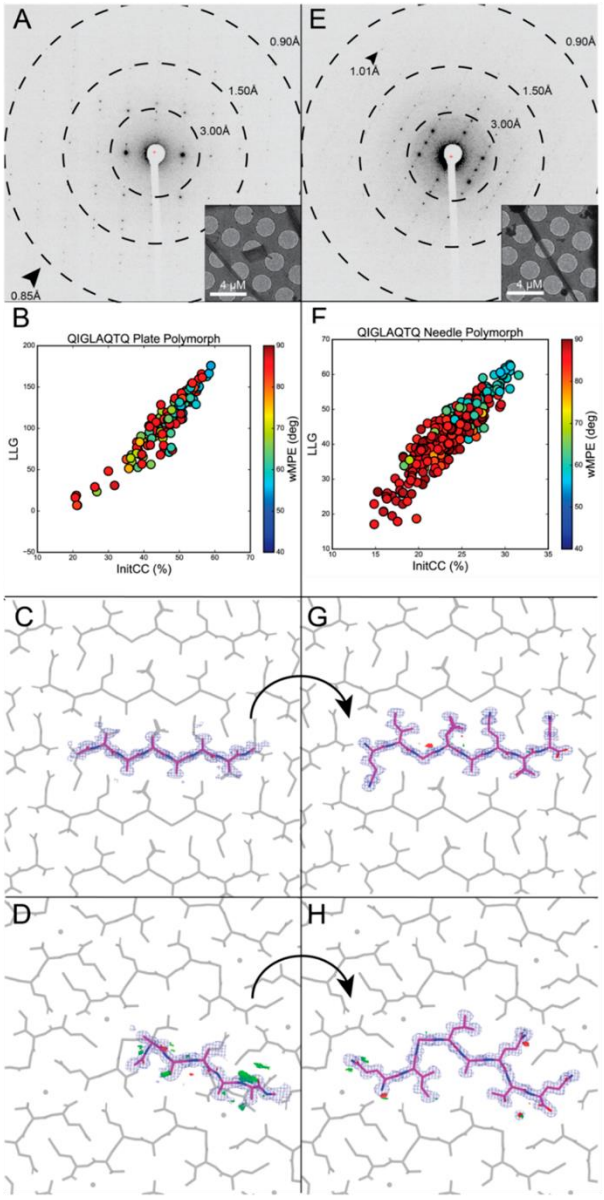


Figure 2. FBP of MicroED data from CPEB3 peptide QIGLAQTQ reveals two polymorphs. A) Electron diffraction pattern reaching atomic resolution for QIGLAQTQ plate morphology. Rings designate resolution ranges while arrowheads designate the highest resolution spot. Overfocused diffraction image of plate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

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

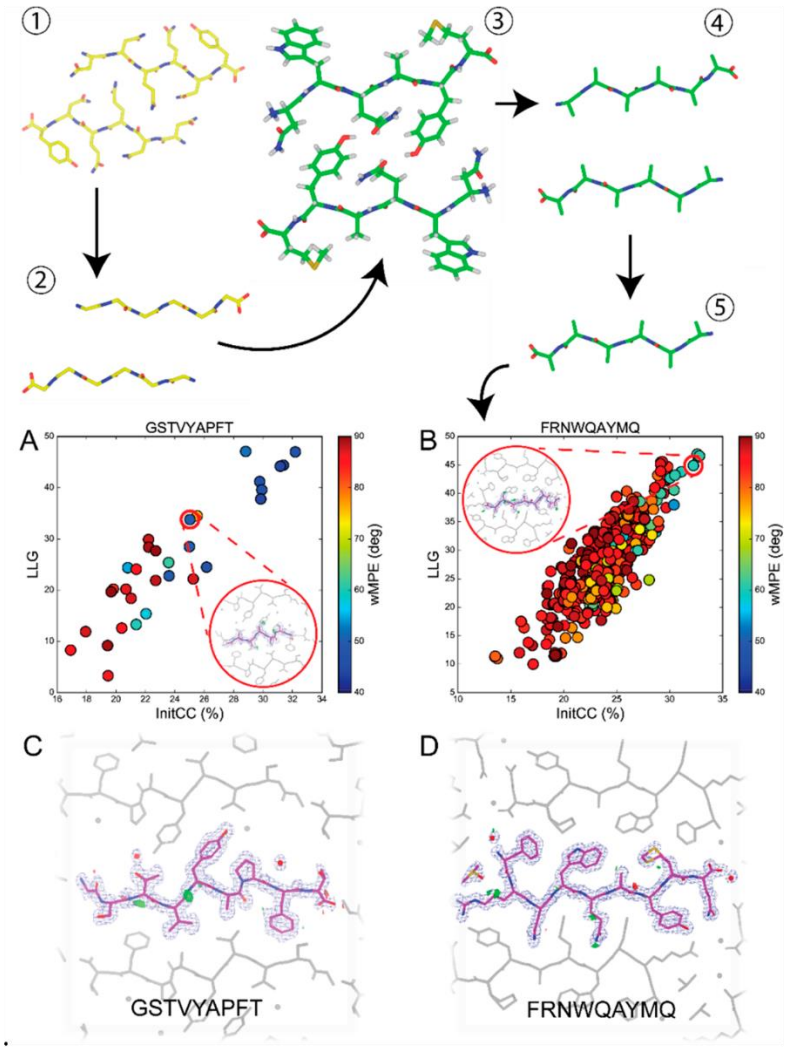


Figure 3. Rosetta library generation and wMPE analysis for GSTVYAPFT and FRNWQAYMQ peptide structures. 1) NNQQNY peptide structure used as a template steric zipper for Rosetta modeling. 2) NNQQNY peptide stripped to glycine residues in preparation for threading. 3) One example of a Rosetta-generated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

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

Structures Determined by Modeled Fragment Libraries

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

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

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

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

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

■ OUTLOOK AND CONCLUSIONS

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

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

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

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

Summary

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

■ EXPERIMENTAL CONSIDERATIONS

Collection of Microfocus X-ray Data and Structure Determination

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

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

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

Preparation of Peptide Nanocrystals

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

Collection of MicroED Data

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

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

MicroED Data Processing

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

Structure Determination by Direct Methods and Refinement

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

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

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

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

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

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

Generation of Custom and Rosetta Fragment Libraries

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

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

General Strand Libraries Generated with ALEPH

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

Structure Refinement

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

■ ASSOCIATED CONTENT

Data Availability Statement

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

Supporting Information

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

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

■ AUTHOR INFORMATION

Corresponding Authors

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

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

Authors

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

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

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

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

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

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

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

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

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

Author Contributions

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

Notes

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

■ ACKNOWLEDGMENTS

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

■ REFERENCES

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

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

(3) Rodriguez, J. A.; Ivanova, M. I.; Sawaya, M. R.; Cascio, D.; Reyes, F. E.; Shi, D.; Sangwan, S.; Guenther, E. L.; Johnson, L. M.; Zhang, M.; Jiang, L.; Arbing, M. A.; Nannenga, B. L.; Hattne, J.; Whitelegge, J.; Brewster, A. S.; Messerschmidt, M.; Boutet, S.; Sauter, N. K.; Gonen, T.; Eisenberg, D. S. Structure of the Toxic Core of α-Synuclein from Invisible Crystals. *Nature* **2015**, *525* (7570), 486−490.

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

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

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

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

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

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

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

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

Chapter 4

This chapter is based on an article in preparation titled “Cryo-EM Structure of a Human LECT2 Amyloid Fibril Reveals a Network of Polar Ladders at its Core.”

Summary

This article presents a cryo-EM structure of recombinant human leukocyte cell-derived chemotaxin-2 (LECT2) amyloid fibrils and provides the first molecular insights into interactions which might contribute to the amyloid aggregation of LECT2 observed in the progression of its associated disease, ALECT2 amyloidosis. ALECT2 is a common and likely underdiagnosed form of systemic amyloidosis which primarily affects the kidneys, leading to impaired renal function, renal failure and nephrotic syndrome. Previous work has predicted regions of the protein which may contribute to amyloid formation but have yet to identify the sequences and structures that may be important to the disease. There remains a need for diagnostic tools to distinguish ALECT2 from other renal amyloid diseases and are currently no available treatments for ALECT2 amyloidosis. In this work, we biochemically and structurally characterize one amyloid fibril conformation adopted by the full-length LECT2 human protein. The recombinant fibrils exhibited canonical amyloid thioflavin T fluorescence and were robust, showing stability in solution and resistance to denaturation agents and protease digestion. The cryo-EM structure identified a 21 residue stretch of amino acids, residues 55-75 of the 151-residue long LECT2 protein, which comprise an energetically favorable and stable amyloid fibril core. This core structure is dry and stabilized by alternating networks of hydrophobic contacts and hydrogen-bonded uncharged polar residues not previously predicted to be amyloidogenic. This structure provides the first step toward a more comprehensive molecular understanding of ALECT2 amyloidosis and a starting point for the future design of targeted therapeutics or diagnostic agents.

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

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

Affiliations:

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

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

Introduction

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

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

Nascent LECT2 is a 151 amino acid polypeptide with an 18-residue N-terminal signal peptide that is cleaved before its secretion into the blood as a globular protein[13]. In circulation, LECT2 performs multiple biological functions. It can act as a chemotactic factor to promote the migration of neutrophils to sites of infection[13,14]. It can also act as a regulator of cartilage growth[15], and it can act as a hepatokine important for metabolic homeostasis[16]. Structurally, LECT2 displays an M23 metalloendopeptidase fold that coordinates Zn(II), but lacks a critical histidine residue that would enable its catalysis[17]. While the clinical details of ALECT2 amyloidosis have been well characterized, the molecular details of ALECT2 amyloid aggregation remain unclear. Loss of this bound Zn(II) may play an important role in the conversion of properly folded LECT2 into an amyloid form *in vitro* as it is thought to destabilize a central beta-barrel motif within the protein[18]. Many patients who suffer from ALECT2 amyloidosis are also homozygous for the common I40V sequence polymorph of the protein[19]. This mutation appears to be important, but not sufficient for inducing amyloid formation, and does not destabilize the metal binding properties of the functional

protein; accordingly, its role in amyloid conversion remains unclear[18]. Amyloidogenic segments of LECT2 have been identified, and predicted amyloid-forming LECT2 peptides assemble into amyloid-like fibrils in isolation[20]. However, the identity of the ALECT2 amyloid core remains unknown.

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

Results

Assembly of recombinant human ALECT2 amyloid fibrils

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

ALECT2 amyloid fibrils are urea and protease-resistant

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

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

Structural determination of a LECT2 amyloid fibril core

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

resembling the teeth of a zipper (Figure S3E and S3F). These observations indicated that the two protofilaments might be related by a symmetry that approximated a 2_1-screw axis. X-ray fibril diffraction from the ALECT2 fibrils showed a strong 4.69Å reflection; corresponding to the true spacing between ALECT2 molecules within a protofilament (Figure S2A and S2B). This spacing was confirmed by analysis of Fourier Transforms from high-resolution images of the fibrils (Figure S3B). The 2D classes were sufficient to generate a *de novo* 3D initial model in RELION and this map was used as an initial model for 3D classification (Figure S3G and S3H).

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

Stability-promoting molecular features in the ALECT2 fibril core

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

Discussion

In search of amyloid folds adopted by the LECT2 protein, we find that recombinant LECT2 polypeptides encoding the I40V sequence polymorphism readily form amyloid fibrils that display a canonical beta-sheet rich architecture with stacked layers of the LECT2 protein separated by a rise of 4.69 Å per layer and obeying a regular, left-handed, helical twist. The amyloid core resolved within these fibrils is composed of two identical protofilaments that each contain 21 of the 133 residues in full-length LECT2: from Methionine 55 to Isoleucine 75. Interestingly, this sequence has not been predicted by previous work to form amyloid fibrils. However, other studies identified the regions immediately preceding and following it as possible amyloidogenic segments[20], and the newly identified core segment is predicted to form steric zipper structures by the amyloid-prediction algorithm in ZipperDB[25](Figure S1A).

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

A comparison of native and fibrillar ALECT2 structures reveals that the beta-strand conformation of N- and C-terminal segments of the fibrillar fold are retained from beta-strands 3 and 4 of the globular fold. However, the central segment of the fibril core represents a substantial reording of

secondary structure, namely rearrangement of loop 2 into 3 beta-strand segments, the largest of which stretches from N65 to N71 (Figure 2B). The outer surface of the ALECT2 fibril is decorated with larger amino acids including basic residues K61, K66, and R74 as well as one acidic residue, E60, and a pi-stacked Y63 (Figure 2B). These residues are also in solvent-exposed positions in the globular fold of LECT2. Residues I56, V73, and I75, which form the outer steric zipper segments of the fibril, are all buried within the central beta-barrel of globular LECT2, and I69, the central hydrophobic residue, is natively buried as part of loop 2 abutted to alpha-helix 1 (Figure S8B). In contrast, the polar-ladder-forming residues Q59, N65, N67, and N71 undergo a large conformational and environmental change from the globular fold. All four of these residues are solvent-facing in loop 2 of the globular structure of LECT2, and all are packed into a relatively dry interface in the fibril structure (Figure S8B). Polar ladder conformations like those revealed in this structure have long been hypothesized to stabilize amyloid assemblies[26,27] and were also favored during energy minimization of the ALECT2 fibril fold in Rosetta (Figure S7B). The structural transition to a polar ladder within a dry environment confers increased stability by joining chains along the fibril axis while fully satisfying hydrogen bonds within the fibril core. This stabilizing effect is further amplified when the two polar ladders in separate protofilaments interdigitate into a zipper conformation, as is observed between residues N67 and N71 in the ALECT2 fibril.

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

arrangement of loop 2 spans the gap between monomers of LECT2 well. A model of a domain-swapped LECT2 dimer through the loop 2 region using the amyloid core structure resolved here shows that this region could facilitate the domain-swapped conformation with minimal disruption to the rest of the globular LECT2 fold (Figure S10A). Further, this conformation maintains the general amyloid core arrangement and could adopt the fibrillar fold following destabilization of the rest of the dimeric interface (Figure S10B).

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

Methods

Protein Expression and Purification

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

Thioflavin T Fluorescence

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

Cryo-EM Sample Preparation and Data Collection

To grow fibrils for cryo-EM analysis, stock LECT2 protein was diluted to 2 mg/ml (120 µM) into a fibrillization buffer containing 50 mM MOPS 6.5, 30 mM TCEP, 10 mM EDTA, and 150 mM NaCl. A volume of 100 µL of protein solution per well was then agitated in a 96 well plate on an acoustic

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

Cryo-EM Data Processing

Collected images were input into MotionCor2 for drift correction and CTFFIND4 was used to calculate Contrast Transfer Functions (CTF). Fibrils were automatically picked using filament mode in crYOLO[30,31] after being trained on a manually selected pool of fibril segments picked from 100 movies. This yielded 1,622,942 segments using an overlap of 90% between neighboring segments and a box size of 384 pixels. The segments were then transferred to RELION 3.1[32–34] for all subsequence 2D and 3D classification. The segments were split into six equally sized groups and subjected to iterative rounds of reference-free two-dimensional (2D) classification using T = 8 and K = 100 to remove poorly aligned classes. After narrowing each split of segments, they were combined again (T = 8 and K = 100) resulting in 308,578 segments contributing to well-defined 2D classes to be used in 3D classification. These classes were used to generate a *de novo* 3D initial model which was used as a reference for further 3D classification. Initial

classification (K = 4, T = 4) did not impose pseudo-2_1-screw symmetry and used an initial helical twist of -1.06° and a helical rise of 4.8 Å. The best class from this classification was selected out and further subclassified by removing any segments with a CTF estimate poorer than 4 Å resolution. The resulting 51,543 segments were used for further 3D classification (K = 3, T = 4) allowing local optimization of helical twist and rise. The best resulting class showed clear strand separation along the fibril axis and the pseudo-2_1-screw symmetry became apparent without it having been imposed. The model and data from these 24,770 segments were used for high-resolution gold-standard 3D auto-refinement with the pseudo-2_1-screw symmetry now enforced. The resulting model underwent iterative Bayesian polishing[34] and CTF refinement[35] to further improve the map (Figure S6A). A final refined helical twist of 179.49° was derived from the auto-refine map and a helical rise of 2.345 Å was imposed on the post-processed map based on X-ray fibril diffraction data showing 4.69 Å strand separation in the fibril samples. The final resolution was ultimately calculated to be 2.715 Å from gold-standard Fourier shell correlations at 0.143 between two independently refined half-maps (Figure S6B).

Model Building

A poly-alanine model was initially built into the post-processed map in Coot[36]. Side chains were modified manually to test the fit of different registrations of the LECT2 sequence in the density. A close approach between tubes of density could be interpreted only as the G58-G72 interaction between symmetric strands, as the closeness of the approach would sterically exclude all other side chains. This feature helped to identify the likely segment of LECT2 and the entire 21 residue chain from Met55-Ile75 was built out from there. Final atomic refinements and statistical calculations for Supplementary Table 1 were performed in Phenix[37]. While the map seemed to agree clearly with this assignment, we decided to confirm its validity by testing the fit and favorability of all possible 21-residue segments from the LECT2 sequence within the map using Rosetta. Threading was performed with a custom python script utilizing the PyRosetta software

package[38]. In a sliding window approach, each sequence window was threaded onto a poly-alanine backbone, placed in the density and Fast Relaxed in PyRosetta. The pipeline used the REF2015 score function[39] with an increased electrostatics weight (fa_elec = 1.5) in combination with a density score term (elec_dens_fast = 25). Each sequence was energy minimized in triplicate and the resulting scores were averaged. Symmetry was applied to pose objects to increase speed of computation. The Fast Relaxed structure was then optimized and refined in Coot[36] and Phenix[37].

X-ray Fibril Diffraction

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

Fibril Stability Assays

ALECT2 amyloid fibrils were grown as described above for cryo-EM sample preparation and then pelleted at 14,300xg for 10 minutes. The fibril pellets were then resuspending to a protein concentration of 40 μM in water, fibrillization buffer, 3M urea with fibrillization buffer, and fibrillization buffer with 1:1000 Proteinase K (Sigma-Aldrich) to a final concentration of 40 nM. These solutions were immediately pipetted in triplicate into a 96 well tray alongside buffer control wells for each and placed on a NEPHELOstar Plus plate reader (BMG Labtech). The tray was stirred every ten minutes to homogenize the mixture and light scattering of particles in solution was measured immediately after each agitation. For analysis of protease resistant ALECT2 fragments, fibrils at 120 μM were incubated with 10 μM Proteinase K for various timepoints and

then evaluated by SDS-PAGE (Figure S9A). The protein gel band from the one-hour timepoint was extracted, digested with Trypsin, and analyzed by gel liquid chromatography tandem mass spectrometry.

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

Gel Liquid Chromatography tandem mass spectrometry spectra collected on Proteinase K digested ALECT2 fibrils were acquired on a ThermoFisher Q-Exactive Plus (UCLA Molecular Instrumentation Center, Los Angeles, CA, USA). LECT2 fibrils at 120 μM were incubated with 10 μM Proteinase K for one hour. These samples were removed at various timepoints, boiled for 10 minutes at 98 °C to halt digestion, and loaded on a 4–12% Bis-Tris SDS-PAGE gel. The gel band from the one-hour timepoint was excised and digested with 200 ng trypsin at 37°C overnight. The digested products were then extracted from the gel bands in 50% acetonitrile/49.9% H2O/ 0.1% trifluoroacetic acid (TFA) followed by desalting with C18 StageTips. Extracted peptides were then injected on an EASY-Spray HPLC column (25 cm × 75 μm ID, PepMap RSLC C18, 2 μm, ThermoScientific) and tandem mass spectra were acquired with a quadrupole orbitrap mass spectrometer (Q-Exactive Plus Thermo Fisher Scientific) interfaced to a nanoelectrospray ionization source. The raw MS/MS data were converted into MGF format by Thermo Proteome Discoverer (VER. 1.4, Thermo Scientific and analyzed by a MASCOT[40] sequence database search.

Acknowledgements

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

Data Statement

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

Figures:

Figure 1: Views of the recombinant ALECT2 fibril core revealed by cryo-EM helical reconstruction. A) Model fibril core structure of ALECT2 (teal) modelled within the refined 2.7Å resolution cryo-EM map (grey mesh). B) View of the ALECT2 fibril core with helical symmetry applied to show its full 850Å twist and 60Å width. C) Side view of the fibril model in the final cryo-EM map demonstrates its fit into density within the fibril layers.

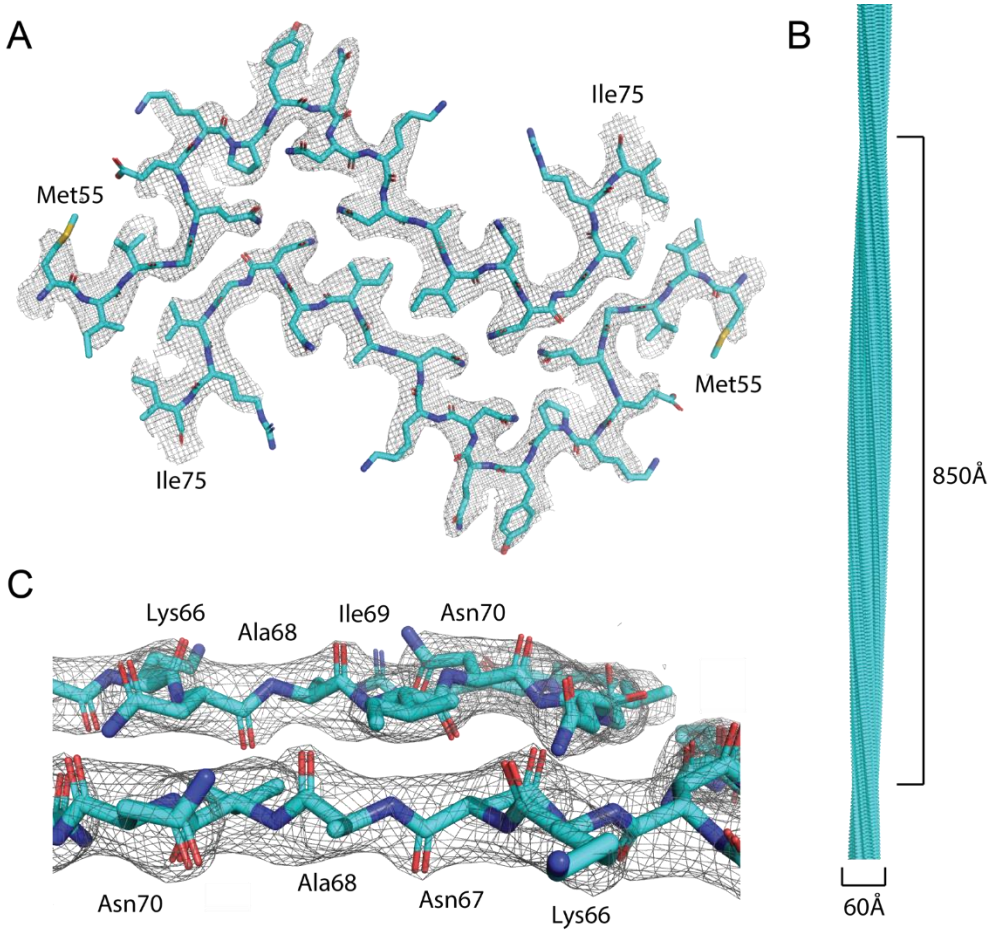


Figure 2: LECT2 fibril core sequence and structure overview. A) Sequence of ALECT2 showing native secondary structure elements as yellow arrows for beta-strands and green rectangles for alpha helices with disulfides shown as orange brackets and the I40V polymorphism noted by a red star. The sequence of the fibril core is highlighted in red. B) Depiction of the fibril core showing the distribution of polar and nonpolar residues. C) Depiction of the fibril core showing stabilizing residues and the calculated solvation energy of the fibril core structure. Depictions of the fibril core were generated using the Amyloid Illustrator web service (https://srv.mbi.ucla.edu/AmyloidAtlas/Illustrator/).

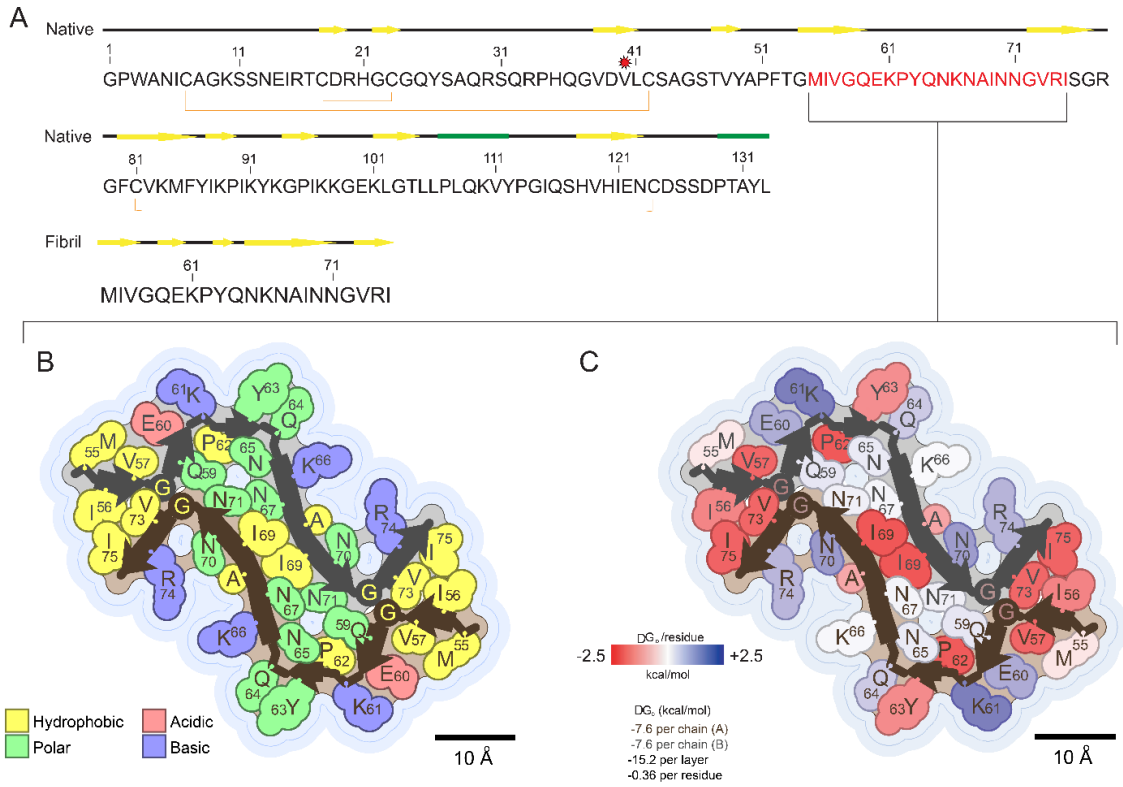
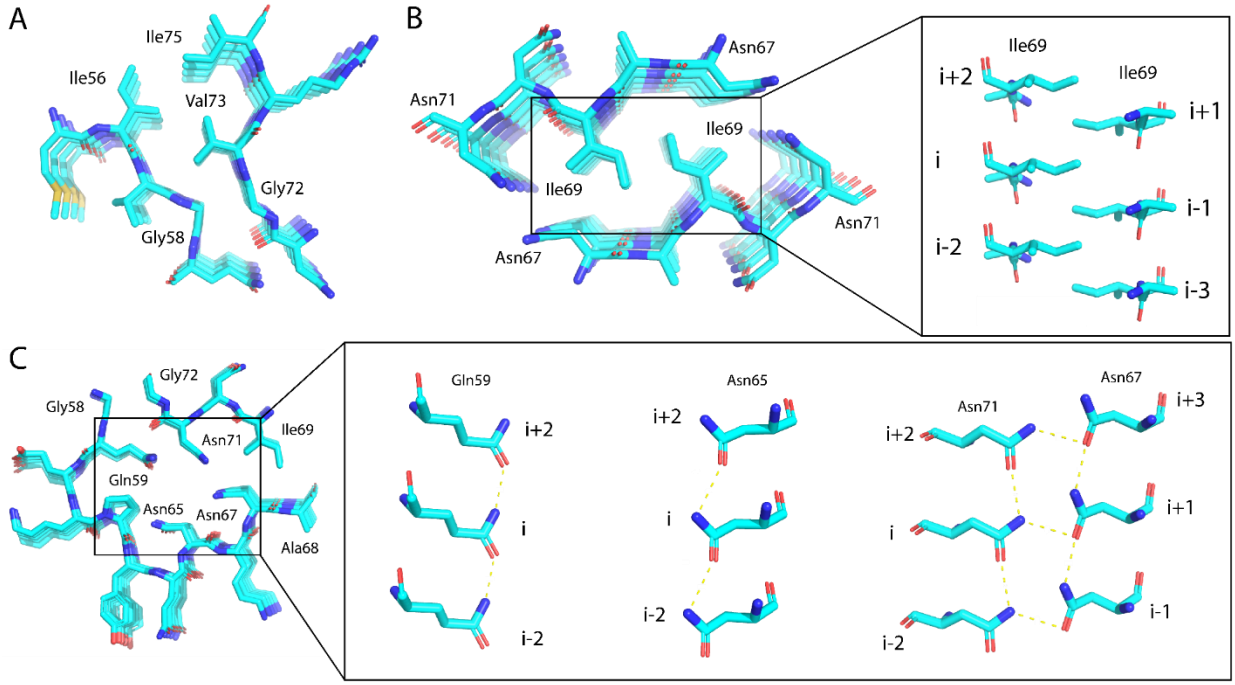


Figure 3: Stabilizing contacts at the core of the ALECT2 fibril. A) The outer steric zipper interface formed by residues Gly58, Gly72, Val73, Ile56, and Ile75. B) The central steric zipper interface formed by residue Ile69 is flanked by a paired polar ladder. Inset shows the pseudo-2_1 stacking of fibril layers. C) Hydrophilic pocket within the LECT2 fibril core contains polar ladders formed by uncharged polar residues Gln59, Asn65, and Asn67 with Asn71, spaced 4.69Å apart along the fibril axis (inset).



References

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

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

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

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

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

6. Spillantini, M.G., Crowther, R.A., Jakes, R., Hasegawa, M., and Goedert, M. (1998). α-Synuclein in filamentous inclusions of Lewy bodies from Parkinson's disease and dementia with Lewy bodies. Proceedings of the National Academy of Sciences *95*, 6469–6473. 10.1073/pnas.95.11.6469.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Chapter 5

Concluding Remarks

Rapid advancements in electron microscopy technology techniques have enabled us to interrogate the high-resolution structures of biological molecules previously out of reach of other structural biology methodologies. To fully take advantage of these advancements it is imperative that we also address the fundamental limiting factors of the techniques. The unavoidable question of the phase problem in electron diffraction, for example, requires new methodologies to help address novel structural data for which traditional phasing techniques have failed. The work on structure determination by electron diffraction presented in this thesis proposes one such technique, fragment-based-phasing, as an additional tool for overcoming this problem and outline the application of the technique to both protein and peptide structures. The techniques described here leave much room to be expanded upon. For example, fragment-based phasing can be integrated with and help rigorously test the increasingly accurate computationally generated structural hypothesis as potential phasing solutions for novel structures or could be used to apply a generalized, unbiased, chemical fragment library to any unknown small molecule structure.

Taking advantage of the advancements in electron microscopy also requires that we employ those techniques to outstanding questions in biology. The structural and biochemical characterization of LECT2 presented in this thesis has advanced that effort by utilizing cryo-EM to determine the first structure of an amyloid fibril of LECT2. This work helps provide foundational structural knowledge about the conversion of LECT2 into an amyloid fold and its role in ALECT2 systemic amyloidosis. Identification of a stable LECT2 fibril core and the residues contributing to it offer a first step towards the rational design of targeted diagnostic agents and critically needed therapeutics for ALECT2 patients.

Appendix 1

Supplementary Information for Chapter 2: Determining Electron Diffraction Structures of Proteins using Fragment-Based Phasing

The work described in this section has been reproduced from:

Richards, L. S.; Millán, C.; Miao, J.; Martynowycz, M. W.; Sawaya, M. R.; Gonen, T.; Borges, R. J.; Usón, I.; Rodriguez, J. A. Fragment-Based Determination of a Proteinase K Structure from MicroED Data Using ARCIMBOLDO_SHREDDER. Acta Cryst D 2020, 76 (8), 703–712. https://doi.org/10.1107/S2059798320008049.

Volume 76 (2020)

Supporting information for article:

Fragment-based determination of a proteinase K structure from MicroED data using *ARCIMBOLDO_SHREDDER*

Logan S. Richards, Claudia Millán, Jennifer Miao, Michael W. Martynowycz, Michael R. Sawaya, Tamir Gonen, Rafael J. Borges, Isabel Usón and Jose A. Rodriguez

S2. Parameters for running ARCIMBOLDO_SHREDDER in spherical mode

[CONNECTION]:

distribute_computing: local_grid

setup_bor_path =

[GENERAL]

working_directory =

mtz_path =

hkl_path =

ent_path =

[ARCIMBOLDO-SHREDDER]

formfactors: FORMFACTORS ELECTRON

name_job =

molecular_weight = 28900

number_of_component = 1

i_label = IMEAN

sigi_label = SIGIMEAN

model_file =

rmsd_shredder = 0.8

savephs: True

alixe: False

shred_method = spherical

shelxe_line = -m15 -a8 -s0.45 -v0 -t6 -y1.60 -e1.10

shelxe_line_last = -m15 -a1 -s0.45 -v0 -t6 -y1.60 -e1.10

S1. Parameters for running ARCIMBOLDO_SHREDDER in sequential mode

[CONNECTION]

distribute_computing = multiprocessing

[GENERAL]

working_directory =

mtz_path =

hkl_path =

ent_path =

[ARCIMBOLDO-SHREDDER]

formfactors: FORMFACTORS ELECTRON

shred_range: 10 20 4 fragment

name_job = output

molecular_weight = 28900

number_of_component = 1

i_label = IMEAN

sigi_label = SIGIMEAN

model_file =

rmsd_shredder = 1.2

trim_to_polyala = True

shred_method = sequential

shelxe_line = -m15 -a8 -s0.5 -v0 -t10 -q -o -y1.60

shelxe_line_last = -m15 -a1 -s0.45 -v0 -t10 -q -o -y1.60 -e1.10

[LOCAL]

path_local_phaser =

path_local_shelxe =

S3. Parameters for running ARCIMBOLDO_LITE

[CONNECTION]

distribute_computing = multiprocessing

[GENERAL]

pdb_path =

working_directory =

mtz_path =

hkl_path =

ent_path =

[ARCIMBOLDO]

formfactors = FORMFACTORS ELECTRON

rmsd = 0.8

i_label = IMEAN

sigi_label = SIGIMEAN

name_job =

molecular_weight =

helix_length = 12

[LOCAL]

path_local_phaser =

path_local_shelxe =

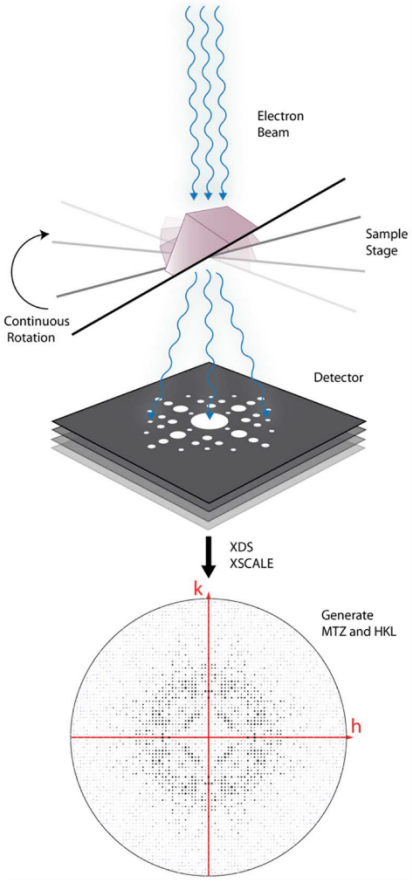


Figure S1 Schematic of continuous rotation MicroED data collection. An electron beam illuminates a protein crystal resting on a cryoEM grid. The microscope sample stage rotates continuously as the scattered electrons are detected below in a process analogous to the rotation method in macromolecular X-ray crystallography (Arndt & Wonacott, 1977; Nannenga *et al*., 2014). The recorded intensities are then processed to generate the MTZ and HKL data files for structure solution.

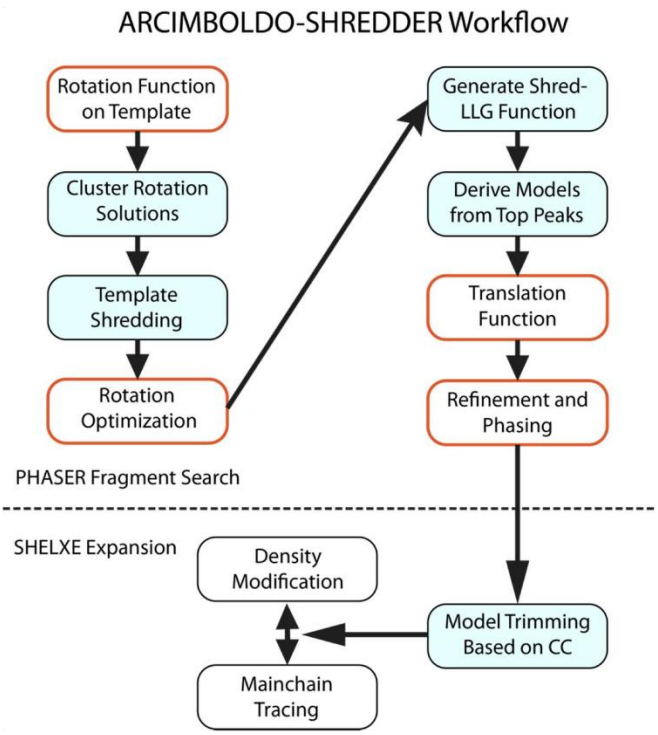


Figure S2 ARCIMBOLDO_SHREDDER workflow. A chosen template model undergoes rotational analysis in Phaser and top scoring peaks are clustered geometrically. These clusters are used to sequentially generate short fragments from the template which then individually undergo rotational analysis in Phaser, the results of which are used to define the Shred-LLG function. Based on the Shred-LLG function models are generated which combine top scoring fragments. These models undergo translational analysis, refinement, and phasing in Phaser. Finally, these solutions are sent into SHELXE where each model is further trimmed based on CC and is subjected to iterative cycles of density modification and mainchain auto-tracing until a final solution is achieved. ARCIMBOLDO operations are shown with a blue background and Phaser operations using electron scattering factors are shown outlined in red.

Table S1 Statistics of individual tilt series merged to obtain the 6V8R dataset

Movie	1	5	6	8	9	10
Resolution (Å)	55.79-1.6	55.79-1.6	55.79-1.6	55.79-1.6	55.79-1.6	55.79-1.6
Completeness	41.5%	34.0%	29.4%	32.6%	49.7%	29.6%
CC1/2	93.5	94.3	95.5	96.9	95.8	94.6
<I/σI>	1.93	2.06	2.81	3.10	2.47	2.14
Multiplicity	2.64	2.12	5.68	2.67	2.45	2.21
R-factor*	34.9%	23.5%	40.1%	18.3%	25.3%	26.6%

* As defined by XSCALE.

Table S2 Summary of HHpred results and structural alignments*

PDB	Identity	Prob.	E-value	Align length	RMSD_SSM	RMSD_GESAMT
5YL7	0.31	100	5.4e-30	344	1.7 Å / 245 res	1.43 Å / 245 res
5JXG	0.193	99.9	2.3e-29	332	2.03 Å /241 res	1.87 Å / 247 res
1GA6	0.208	99.9	6.4e-25	355	2.32 Å / 216 res	2.1 Å / 226 res

* Values obtained from HHpred reported as of March 2020.

Appendix 2

Supplementary Information for Chapter 3: Fragment-Based Phasing for Peptide Structures and de novo Fragment Generation

The work described in this section has been reproduced from:

Richards, L. S.; Flores, M. D.; Millán, C.; Glynn, C.; Zee, C.-T.; Sawaya, M. R.; Gallagher-Jones, M.; Borges, R. J.; Usón, I.; Rodriguez, J. A. Fragment-Based Ab Initio Phasing of Peptidic Nanocrystals by MicroED. ACS Bio Med Chem Au 2023. https://doi.org/10.1021/acsbiomedchemau.2c00082.

Supporting Information

Fragment-based *ab initio* phasing of peptidic nanocrystals by MicroED

Authors: Logan S. Richards[1‡], Maria D. Flores[1‡], Claudia Millán[2‡], Calina Glynn[1], Chih-Te Zee[1], Michael R. Sawaya[3], Marcus Gallagher-Jones[4], Rafael J. Borges[2], Isabel Usón[2,5,*] and Jose A. Rodriguez[1,*]

Affiliations:

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

[2]Crystallographic Methods, Institute of Molecular Biology of Barcelona (IBMB–CSIC), Barcelona Science Park, Helix Building, Baldiri Reixach 15, 08028 Barcelona, Spain

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

[4] Correlated Imaging, The Rosalind Franklin Institute, Harwell Science & Innovation Campus, Rutherford Avenue, Harwell, Didcot OX11 0GD, United Kingdom

[5] ICREA, Institució Catalana de Recerca i Estudis Avançats, Passeig Lluís Companys 23, 08003 Barcelona, Spain.

‡ Denotes equal contribution

* Correspondence to Jose A. Rodriguez (jrodriguez@mbi.ucla.edu) and Isabel Usón (iufcri@ibmb.csic.es)

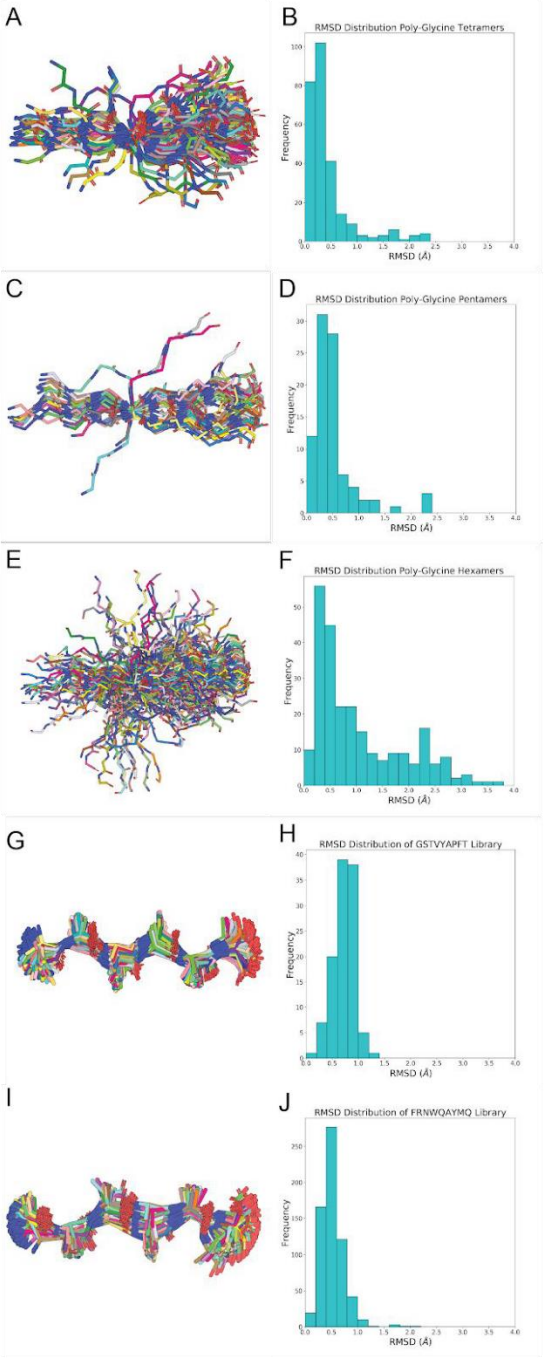
Table S1 Crystallographic data collection and refinement statistics. Values in parentheses are for the highest-resolution shell.

Direct Methods Structures	NYNNYQ (7N2H)	NYNNYQ	QIGLAQTQ Plate Polymorph (7N2E)	QYNNENNFV (7N2L)
Data Collection and Processing				
Molecular Formula	$C_{35}H_{46}N_{10}O_{13}$	$C_{35}H_{46}N_{10}O_{13}$	$C_{31}H_{55}N_9O_{11}$	$C_{49}H_{68}N_{14}O_{18}$
No. of crystals merged	2 (4 Scans)	4	3	6
Space Group	P 1	P 3_1	P 1 2_1 1	P 1
Cell dimensions				
a, b, c (Å)	4.86, 14.33, 15.75	27.20, 27.20, 4.83	4.83, 16.29, 29.02	4.87, 10.06, 30.66
α, β, γ (°)	87.369, 86.325, 88.501	90, 90, 120	90, 94.605, 90	94.850, 90.264, 99.988
Resolution Limit (Å)	1.05 (1.10-1.05)	1.10 (1.19-1.10)	1.00 (1.05-1.00)	0.90 (0.93-0.90)
Beam Type	x-ray	electron	electron	electron
Wavelength (Å)	0.979	0.0251	0.0251	0.0251
Electron Dose ($e^-/Å^2$)	N/A	<5	<5	<5
Rmerge (%)	16.1 (35.9)	21.1 (48.3)	15.2 (26.2)	19.7 (47.5)
I/σ_I	4.13 (1.97)	4.92 (1.21)	7.24 (4.03)	5.21 (2.56)
Completeness	90.2 (61.1)	83.4 (53.2)	79.6 (80.8)	81.0 (82.3)
$CC_{1/2}$	97.4 (89.1)	98.5 (60.4)	98.8 (96.8)	98.8 (78.3)
No. total reflections	4635 (314)	7893 (431)	11581 (1430)	18502 (1603)
No. unique reflections	1793 (154)	1357 (191)	1976 (252)	3437 (330)
Multiplicity	2.6 (2.0)	5.8 (2.2)	5.9 (5.7)	5.4 (4.9)
Refinement				
Resolution (Å)	15.70-1.05 (1.10 - 1.05)	13.60-1.10 (1.19-1.10)	7.10-1.00(1.05-1.00)	7.98-0.90 (0.93-0.90)
No. of Reflections (free)	1787 (180)	1356 (136)	1959 (197)	3125 (310)
Rwork (%)	14.92 (29.53)	13.83 (26.25)	18.60 (22.07)	20.01 (28.03)
Rfree(%)	16.49 (28.57)	16.06 (34.89)	20.92 (25.03)	20.60 (27.91)
CC(work)	96.9 (80.9)	98.2 (78.1)	95.5 (90.3)	95.8 (81.1)
CC(free)	94.7 (67.9)	97.2 (52.8)	92.2 (86.8)	95.0 (80.5)
Atoms				
No. peptide atoms	58	58	64	81
No. waters	5	2	0	2
No. ligand atoms	0	0	0	0
Average B-factor:	5.48	7.07	3.15	4.28
Protein	4.81	6.7	3.15	4.16
Water	13.16	17.8	N/A	8.88
Ligand	N/A	N/A	N/A	N/A
R.m.s.d. bonds (Å)	0.009	0.007	0.23	0.004
R.m.s.d. angles (°)	0.864	0.693	1.711	0.737
Ramachandran Outliers (%)	0	0	0	0
Clashscore	0	0	0	0

Table S2: Summary of peptide fragment-based phasing

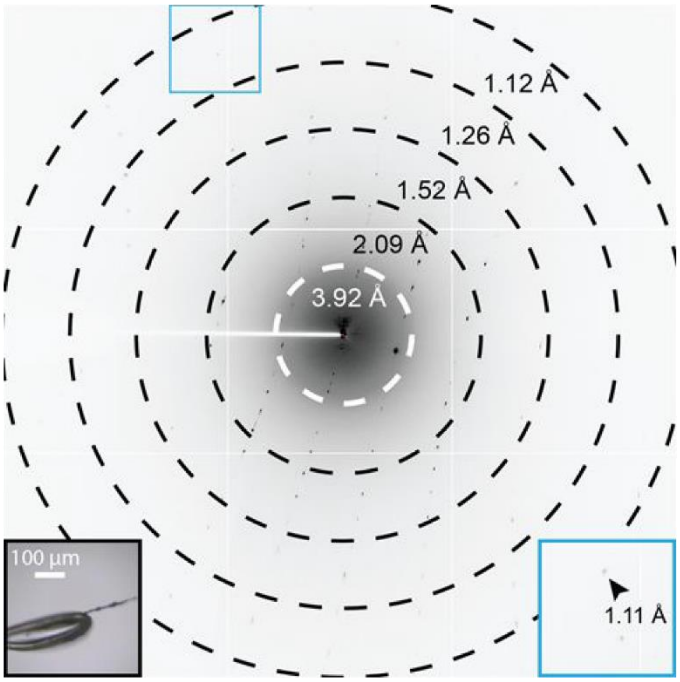
Peptide	**Resolution**	**# of probes**	**Best wMPE**	**Ca RMSD (output to final)**	**Final R_{work}/R_{free}**
QYNNENNFV	1.5Å	249	58.0°	0.522 Å^2	17.44 / 22.65
NYNNYQ	1.3Å	89	49.5°	0.561 Å^2	16.14 / 18.54
QIGLAQTQ_plate	1.2Å	89	54.8°	0.247 Å^2	19.8 / 22.9
QIGLAQTQ_needle	1.2Å	270	55.3°	2.395 Å^2	19.24 / 23.65
GSTVYAPFT	1.4Å	111	43.9°	0.509 Å^2	19.36 / 19.23
FRNWQAYMQ	1.5Å	640	53.9°	1.438 Å^2	21.20 / 23.87

Supplementary Figure 1: Visualization of fragment libraries and RMSD analysis



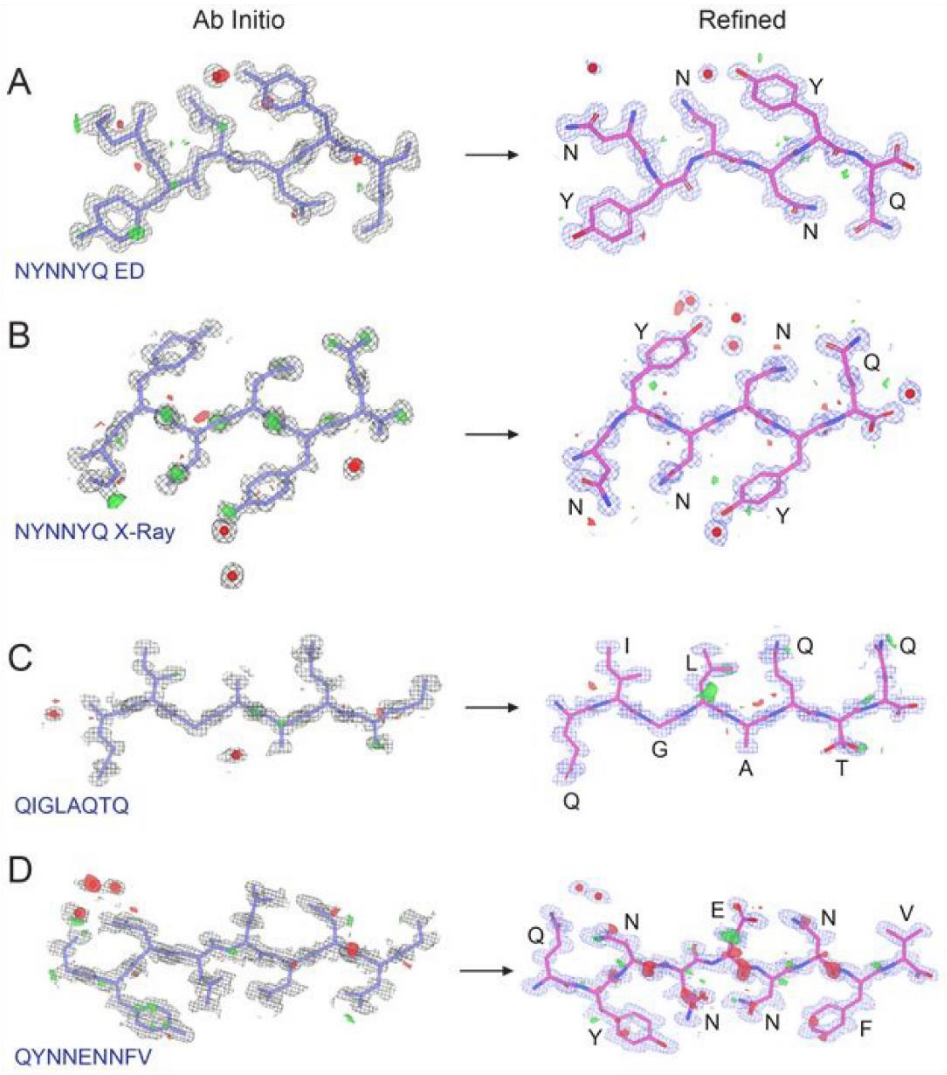
The poly-glycine tetrapeptide (A), poly-glycine pentapeptide (C), poly-glycine hexapeptide (E), poly-alanine Rosetta GSTVYAPFT (G), and poly-alanine Rosetta FRNWQAYMQ (I) libraries are shown aligned in pymol. B, D, F, H, J) Distribution of RMSD values calculated for each fragment relative to an idealized poly-glycine or poly-alanine strand of equivalent length.

Supplementary Figure 2: X-ray diffraction of the NYNNYQ peptide



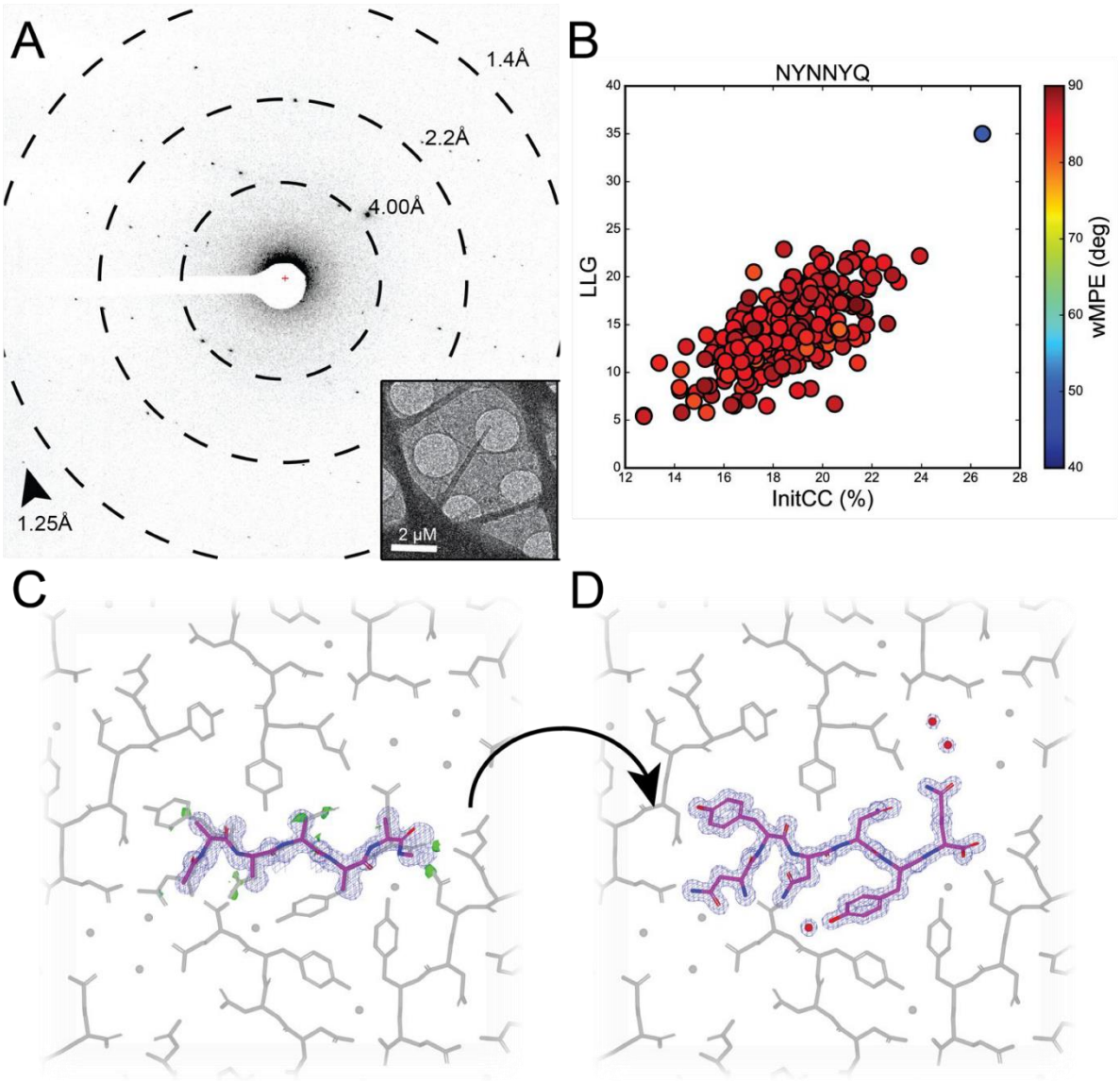
Single diffraction pattern of NYNNYQ measured during manual continuous raster scanning, microfocus x-ray data collection. The pattern corresponds to a 5° wedge of reciprocal space. Black inset shows in-line crystal image; the area boxed by the blue square corresponds to the magnified region (blue inset) of the pattern, which shows diffraction near the detector edge at approximately 1.1 Å resolution (black arrow). Resolution rings are indicated.

Supplementary Figure 3: *Ab initio* solutions and refined peptide structures



Ab initio structures and maps of NYNNYQ (A, electron), NYNNYQ (B, x-ray), QIGLAQTQ (C, plate polymorph), and QYNNENNFV (D). The initial atomic coordinates on the left, calculated by *SHELXD*, are overlaid by their corresponding initial maps. The refined atomic structures on the right are overlaid by their corresponding final maps. The $2F_O$-F_C maps, represented by the black mesh (left) or blue mesh (right), is contoured at 1.2σ. Green and red surfaces represent F_o-F_c maps contoured at 3.0σ and -3.0σ, respectively. Modeled waters are present as red spheres. The waters in the *ab initio* solution of B and the refined structure are related by symmetry. The waters in the *ab initio* solution of C were later determined to be backbone atoms during refinement.

Supplementary Figure 4: Diffraction pattern, overfocused diffraction image, density maps of initial ARCIMBOLDO outputs and final structures for New1p peptide.



A) Diffraction pattern for NYNNYQ B) Post-mortem analysis for ARCIMBOLDO-BORGES plotting LLG vs. InitCC. C) Initial output potential maps expanded by SHELXE for NYNNYQ overlaid on final solution (grey). D) Final potential maps of NYNNYQ with symmetry mates shown in grey.

Supplementary References:

Apostol, M. I.; Wiltzius, J. J. W.; Sawaya, M. R.; Cascio, D.; Eisenberg, D. Atomic Structures Suggest Determinants of Transmission Barriers in Mammalian Prion Disease. *Biochemistry* **2011**, *50* (13), 2456–2463. https://doi.org/10.1021/bi101803k.

Brumshtein, B.; Esswein, S. R.; Sawaya, M. R.; Rosenberg, G.; Ly, A. T.; Landau, M.; Eisenberg, D. S. Identification of Two Principal Amyloid-Driving Segments in Variable Domains of Ig Light Chains in Systemic Light-Chain Amyloidosis. *J. Biol. Chem.* **2018**, *293* (51), 19659–19671. https://doi.org/10.1074/jbc.RA118.004142.

Colletier, J.-P.; Laganowsky, A.; Landau, M.; Zhao, M.; Soriaga, A. B.; Goldschmidt, L.; Flot, D.; Cascio, D.; Sawaya, M. R.; Eisenberg, D. Molecular Basis for Amyloid- Polymorphism. *Proceedings of the National Academy of Sciences* **2011**, *108* (41), 16938–16943. https://doi.org/10.1073/pnas.1112600108.

Colvin, M. T.; Silvers, R.; Ni, Q. Z.; Can, T. V.; Sergeyev, I.; Rosay, M.; Donovan, K. J.; Michael, B.; Wall, J.; Linse, S.; Griffin, R. G. Atomic Resolution Structure of Monomorphic $A\beta_{42}$ Amyloid Fibrils. *J. Am. Chem. Soc.* **2016**, *138* (30), 9663–9674. https://doi.org/10.1021/jacs.6b05129.

Do, T. D.; LaPointe, N. E.; Sangwan, S.; Teplow, D. B.; Feinstein, S. C.; Sawaya, M. R.; Eisenberg, D. S.; Bowers, M. T. Factors That Drive Peptide Assembly from Native to Amyloid Structures: Experimental and Theoretical Analysis of [Leu-5]-Enkephalin Mutants. *J. Phys. Chem. B* **2014**, *118* (26), 7247–7256. https://doi.org/10.1021/jp502473s.

Evans, K. C.; Berger, E. P.; Cho, C. G.; Weisgraber, K. H.; Lansbury, P. T. Apolipoprotein E Is a Kinetic but Not a Thermodynamic Inhibitor of Amyloid Formation: Implications for the Pathogenesis and Treatment of Alzheimer Disease. *Proceedings of the National Academy of Sciences* **1995**, *92* (3), 763–767. https://doi.org/10.1073/pnas.92.3.763.

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. Cryo-EM Structures of Tau Filaments from Alzheimer's Disease. *Nature* **2017**, *547* (7662), 185–190. https://doi.org/10.1038/nature23002.

Gremer, L.; Schölzel, D.; Schenk, C.; Reinartz, E.; Labahn, J.; Ravelli, R. B. G.; Tusche, M.; Lopez-Iglesias, C.; Hoyer, W.; Heise, H.; Willbold, D.; Schröder, G. F. Fibril Structure of Amyloid-β(1–42) by Cryo–Electron Microscopy. *Science* **2017**, *358* (6359), 116–119. https://doi.org/10.1126/science.aao2825.

Guenther, E. L.; Cao, Q.; Trinh, H.; Lu, J.; Sawaya, M. R.; Cascio, D.; Boyer, D. R.; Rodriguez, J. A.; Hughes, M. P.; Eisenberg, D. S. Atomic Structures of TDP-43 LCD Segments and Insights into Reversible or Pathogenic Aggregation. *Nat Struct Mol Biol* **2018**, *25* (6), 463–471. https://doi.org/10.1038/s41594-018-0064-2.

Hughes, M. P.; Sawaya, M. R.; Boyer, D. R.; Goldschmidt, L.; Rodriguez, J. A.; Cascio, D.; Chong, L.; Gonen, T.; Eisenberg, D. S. Atomic Structures of Low-Complexity Protein Segments Reveal Kinked β Sheets That Assemble Networks. *Science* **2018**, *359* (6376), 698–701. https://doi.org/10.1126/science.aan6398.

Ivanova, M. I.; Sievers, S. A.; Guenther, E. L.; Johnson, L. M.; Winkler, D. D.; Galaleldeen, A.; Sawaya, M. R.; Hart, P. J.; Eisenberg, D. S. Aggregation-Triggering Segments of SOD1 Fibril Formation Support a Common Pathway for Familial and Sporadic ALS. *Proceedings of the National Academy of Sciences* **2014**, *111* (1), 197–201. https://doi.org/10.1073/pnas.1320786110.

Krotee, P.; Rodriguez, J. A.; Sawaya, M. R.; Cascio, D.; Reyes, F. E.; Shi, D.; Hattne, J.; Nannenga, B. L.; Oskarsson, M. E.; Philipp, S.; Griner, S.; Jiang, L.; Glabe, C. G.; Westermark, G. T.; Gonen, T.; Eisenberg, D. S. Atomic Structures of Fibrillar Segments of HIAPP Suggest Tightly Mated β-Sheets Are Important for Cytotoxicity. *eLife* **2017**, *6*, e19273. https://doi.org/10.7554/eLife.19273.

Laganowsky, A.; Liu, C.; Sawaya, M. R.; Whitelegge, J. P.; Park, J.; Zhao, M.; Pensalfini, A.; Soriaga, A. B.; Landau, M.; Teng, P. K.; Cascio, D.; Glabe, C.; Eisenberg, D. Atomic View of a Toxic Amyloid Small Oligomer. *Science* **2012**, *335* (6073), 1228–1231. https://doi.org/10.1126/science.1213151.

Li, D.; Jones, E. M.; Sawaya, M. R.; Furukawa, H.; Luo, F.; Ivanova, M.; Sievers, S. A.; Wang, W.; Yaghi, O. M.; Liu, C.; Eisenberg, D. S. Structure-Based Design of Functional Amyloid Materials. *J. Am. Chem. Soc.* **2014**, *136* (52), 18044–18051. https://doi.org/10.1021/ja509648u.

Liu, J.; Doty, T.; Gibson, B.; Heyer, W.-D. Human BRCA2 Protein Promotes RAD51 Filament Formation on RPA-Covered Single-Stranded DNA. *Nat Struct Mol Biol* **2010**, *17* (10), 1260–1262. https://doi.org/10.1038/nsmb.1904.

Lu, J.-X.; Qiang, W.; Yau, W.-M.; Schwieters, C. D.; Meredith, S. C.; Tycko, R. Molecular Structure of β-Amyloid Fibrils in Alzheimer's Disease Brain Tissue. *Cell* **2013**, *154* (6), 1257–1268. https://doi.org/10.1016/j.cell.2013.08.035.

Luo, F.; Gui, X.; Zhou, H.; Gu, J.; Li, Y.; Liu, X.; Zhao, M.; Li, D.; Li, X.; Liu, C. Atomic Structures of FUS LC Domain Segments Reveal Bases for Reversible Amyloid Fibril Formation. *Nat Struct Mol Biol* **2018**, *25* (4), 341–346. https://doi.org/10.1038/s41594-018-0050-8.

Murray, D. T.; Kato, M.; Lin, Y.; Thurber, K. R.; Hung, I.; McKnight, S. L.; Tycko, R. Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. *Cell* **2017**, *171* (3), 615-627.e16. https://doi.org/10.1016/j.cell.2017.08.048.

Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A. Ø.; Riekel, C.; Grothe, R.; Eisenberg, D. Structure of the Cross-β Spine of Amyloid-like Fibrils. *Nature* **2005**, *435* (7043), 773–778. https://doi.org/10.1038/nature03680.

Rodriguez, J. A.; Ivanova, M. I.; Sawaya, M. R.; Cascio, D.; Reyes, F. E.; Shi, D.; Sangwan, S.; Guenther, E. L.; Johnson, L. M.; Zhang, M.; Jiang, L.; Arbing, M. A.; Nannenga, B. L.; Hattne, J.; Whitelegge, J.; Brewster, A. S.; Messerschmidt, M.; Boutet, S.; Sauter, N. K.; Gonen, T.; Eisenberg, D. S. Structure of the Toxic Core of α-Synuclein from Invisible Crystals. *Nature* **2015**, *525* (7570), 486–490. https://doi.org/10.1038/nature15368.

Saelices, L.; Chung, K.; Lee, J. H.; Cohn, W.; Whitelegge, J. P.; Benson, M. D.; Eisenberg, D. S. Amyloid Seeding of Transthyretin by Ex Vivo Cardiac Fibrils and Its Inhibition. *Proc Natl Acad Sci USA* **2018**, *115* (29), E6741–E6750. https://doi.org/10.1073/pnas.1805131115.

Saelices, L.; Nguyen, B. A.; Chung, K.; Wang, Y.; Ortega, A.; Lee, J. H.; Coelho, T.; Bijzet, J.; Benson, M. D.; Eisenberg, D. S. A Pair of Peptides Inhibits Seeding of the Hormone Transporter Transthyretin into Amyloid Fibrils. *J. Biol. Chem.* **2019**, *294* (15), 6130–6141. https://doi.org/10.1074/jbc.RA118.005257.

Sangwan, S.; Sawaya, M. R.; Murray, K. A.; Hughes, M. P.; Eisenberg, D. S. Atomic Structures of Corkscrew-Forming Segments of SOD1 Reveal Varied Oligomer Conformations: Structures of Variants of a Toxic Segment from SOD1. *Protein Science* **2018**, *27* (7), 1231–1242.

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. Atomic Structures of Amyloid Cross-β Spines Reveal Varied Steric Zippers. *Nature* **2007**, *447* (7143), 453–457. https://doi.org/10.1038/nature05695.

Schütz, A. K.; Vagt, T.; Huber, M.; Ovchinnikova, O. Y.; Cadalbert, R.; Wall, J.; Güntert, P.; Böckmann, A.; Glockshuber, R.; Meier, B. H. Atomic-Resolution Three-Dimensional Structure of Amyloid β Fibrils Bearing the Osaka Mutation. *Angew. Chem. Int. Ed.* **2015**, *54* (1), 331–335. https://doi.org/10.1002/anie.201408598.

Seidler, P. M.; Boyer, D. R.; Rodriguez, J. A.; Sawaya, M. R.; Cascio, D.; Murray, K.; Gonen, T.; Eisenberg, D. S. Structure-Based Inhibitors of Tau Aggregation. *Nature Chem* **2018**, *10* (2), 170–176. https://doi.org/10.1038/nchem.2889.

Sgourakis, N. G.; Yau, W.-M.; Qiang, W. Modeling an In-Register, Parallel "Iowa" Aβ Fibril Structure Using Solid-State NMR Data from Labeled Samples with Rosetta. *Structure* **2015**, *23* (1), 216–227. https://doi.org/10.1016/j.str.2014.10.022.

Sievers, S. A.; Karanicolas, J.; Chang, H. W.; Zhao, A.; Jiang, L.; Zirafi, O.; Stevens, J. T.; Münch, J.; Baker, D.; Eisenberg, D. Structure-Based Design of Non-Natural Amino-Acid Inhibitors of Amyloid Fibril Formation. *Nature* **2011**, *475* (7354), 96–100. https://doi.org/10.1038/nature10154.

Soragni, A.; Janzen, D. M.; Johnson, L. M.; Lindgren, A. G.; Thai-Quynh Nguyen, A.; Tiourin, E.; Soriaga, A. B.; Lu, J.; Jiang, L.; Faull, K. F.; Pellegrini, M.; Memarzadeh, S.; Eisenberg, D.

S. A Designed Inhibitor of P53 Aggregation Rescues P53 Tumor Suppression in Ovarian Carcinomas. *Cancer Cell* **2016**, *29* (1), 90–103. https://doi.org/10.1016/j.ccell.2015.12.002.

Soragni, A.; Yousefi, S.; Stoeckle, C.; Soriaga, A. B.; Sawaya, M. R.; Kozlowski, E.; Schmid, I.; Radonjic-Hoesli, S.; Boutet, S.; Williams, G. J.; Messerschmidt, M.; Seibert, M. M.; Cascio, D.; Zatsepin, N. A.; Burghammer, M.; Riekel, C.; Colletier, J.-P.; Riek, R.; Eisenberg, D. S.; Simon, H.-U. Toxicity of Eosinophil MBP Is Repressed by Intracellular Crystallization and Promoted by Extracellular Aggregation. *Molecular Cell* **2015**, *57* (6), 1011–1021. https://doi.org/10.1016/j.molcel.2015.01.026.

Soriaga, A. B.; Sangwan, S.; Macdonald, R.; Sawaya, M. R.; Eisenberg, D. Crystal Structures of IAPP Amyloidogenic Segments Reveal a Novel Packing Motif of Out-of-Register Beta Sheets. *J. Phys. Chem. B* **2016**, *120* (26), 5810–5816. https://doi.org/10.1021/acs.jpcb.5b09981.

Tuttle, M. D.; Comellas, G.; Nieuwkoop, A. J.; Covell, D. J.; Berthold, D. A.; Kloepper, K. D.; Courtney, J. M.; Kim, J. K.; Barclay, A. M.; Kendall, A.; Wan, W.; Stubbs, G.; Schwieters, C. D.; Lee, V. M. Y.; George, J. M.; Rienstra, C. M. Solid-State NMR Structure of a Pathogenic Fibril of Full-Length Human α-Synuclein. *Nat Struct Mol Biol* **2016**, *23* (5), 409–415. https://doi.org/10.1038/nsmb.3194.

Wiltzius, J. J. W.; Landau, M.; Nelson, R.; Sawaya, M. R.; Apostol, M. I.; Goldschmidt, L.; Soriaga, A. B.; Cascio, D.; Rajashankar, K.; Eisenberg, D. Molecular Mechanisms for Protein-Encoded Inheritance. *Nat Struct Mol Biol* **2009**, *16* (9), 973–978. https://doi.org/10.1038/nsmb.1643.

Wiltzius, J. J. W.; Sievers, S. A.; Sawaya, M. R.; Cascio, D.; Popov, D.; Riekel, C.; Eisenberg, D. Atomic Structure of the Cross-β Spine of Islet Amyloid Polypeptide (Amylin). *Protein Sci.* **2008**, *17* (9), 1467–1474. https://doi.org/10.1110/ps.036509.108.

Xiao, Y.; Ma, B.; McElheny, D.; Parthasarathy, S.; Long, F.; Hoshi, M.; Nussinov, R.; Ishii, Y. Aβ(1–42) Fibril Structure Illuminates Self-Recognition and Replication of Amyloid in Alzheimer's Disease. *Nat Struct Mol Biol* **2015**, *22* (6), 499–505. https://doi.org/10.1038/nsmb.2991.

Yu, L.; Lee, S.-J.; Yee, V. C. Crystal Structures of Polymorphic Prion Protein B1 Peptides Reveal Variable Steric Zipper Conformations. *Biochemistry* **2015**, *54* (23), 3640–3648. https://doi.org/10.1021/acs.biochem.5b00425.

Appendix 3

Supplementary Information for Chapter 4: Structural Characterization of a LECT2 Amyloid Fibril Core

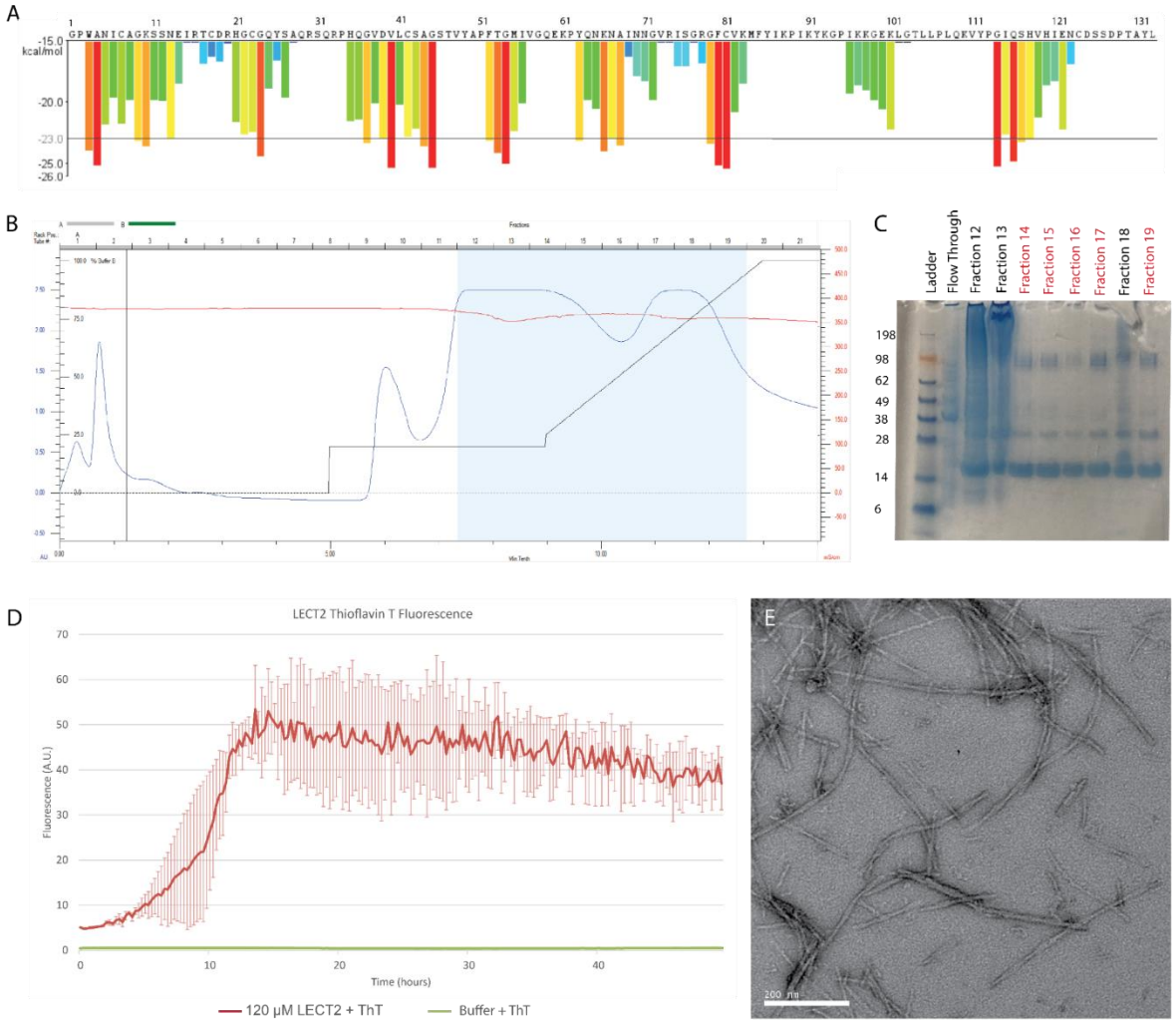
The work described in this section has been reproduced from:

Richards, L. S.; Flores, M. D.; Zink, S.; Schibrowsky, N. A.; Sawaya, M. R.; Rodriguez, J. A. Cryo-EM Structure of a Human LECT2 Amyloid Fibril Reveals a Network of Polar Ladders at Its Core. bioRxiv February 11, 2023, p 2023.02.08.527771. https://doi.org/10.1101/2023.02.08.527771.

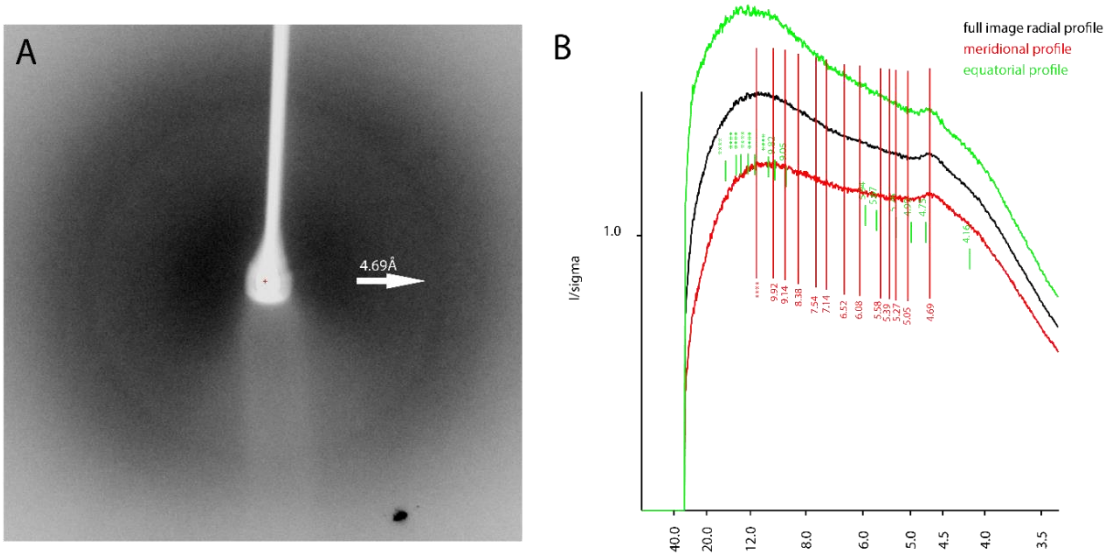
Supplementary Table 1: Cryo-EM data collection and structure information.

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

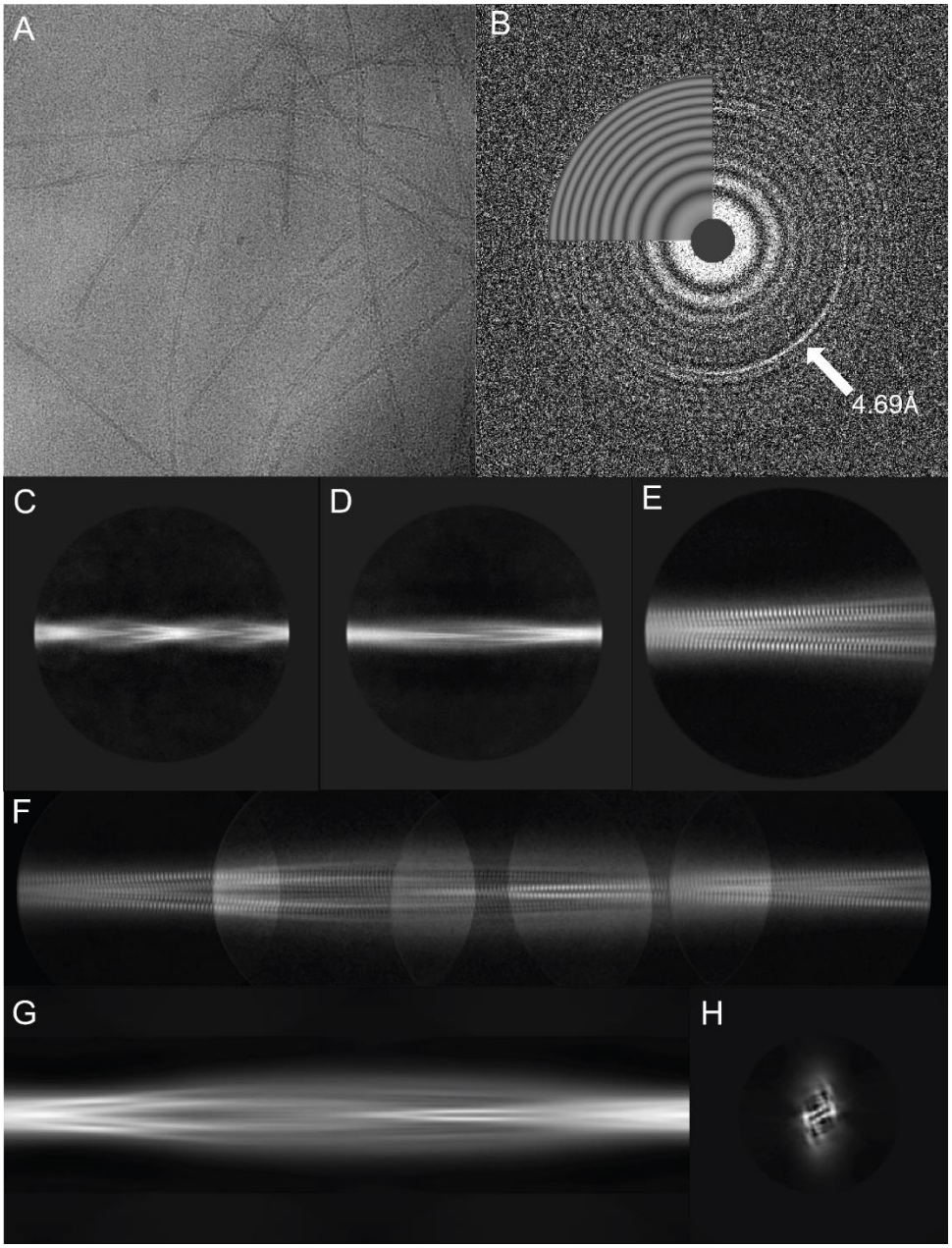
Supplementary Figure 1: Production of recombinant ALECT2 amyloid fibrils. A) Predicted amyloid forming six-residue segments (kcal/mol < -23.0) are shown for the LECT2 sequence, calculated using ZipperDB. B) A His-tagged, full-length LECT2 polypeptide was expressed and purified using a His-Trap column under the illustrated gradient. C) Purity of the fractions was evaluated by gel electrophoresis. The purest fractions: 14-17, 19 (red) were pooled and concentrated. D) The recombinant protein was used to form fibrils and Thioflavin T fluorescence was monitored over 48 hours of growth. E) The resulting fibrils from Thioflavin T analysis were imaged by negative stain TEM.



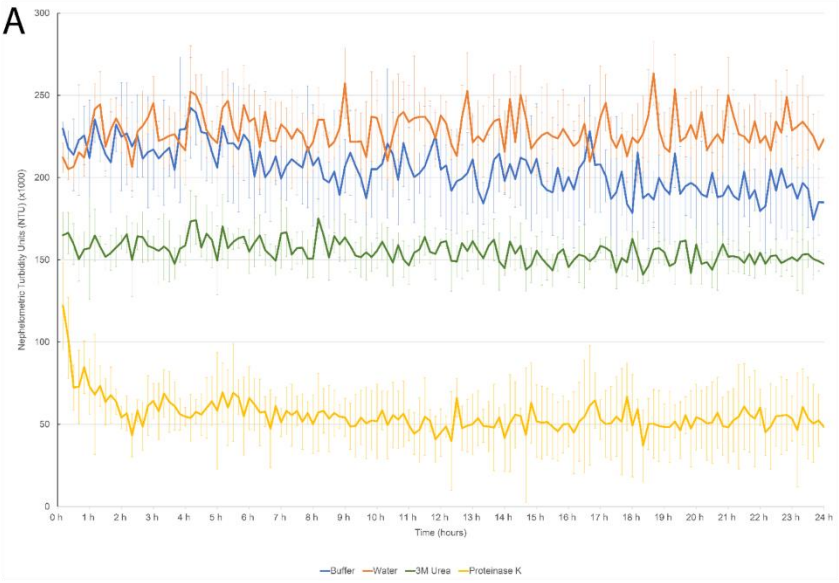
Supplementary Figure 2: X-ray Diffraction from ALECT2 fibrils A) X-ray diffraction exposure of bundles recombinant ALECT2 fibrils. B) Radial profile generated from the diffraction image indicates a strong diffraction ring at 4.69Å.

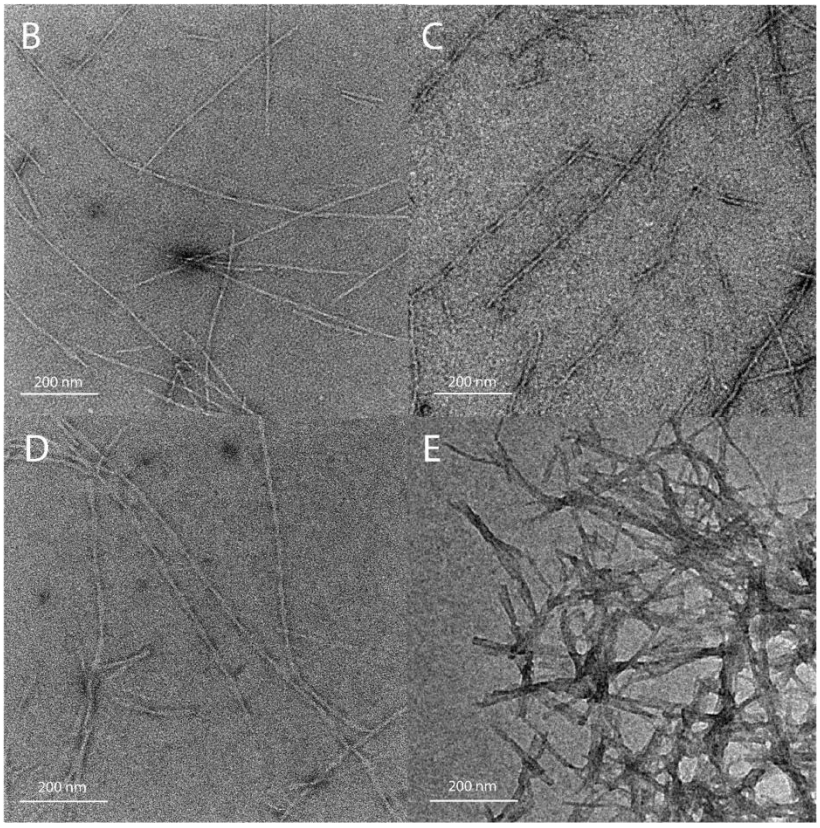


Supplementary Figure 3: Vitrified ALECT2 fibrils, CTF analysis, and 2D classification. A) Example image of the vitrified fibrils used for cryo-EM data collection. B) Example CTF image showing the signature fibril diffraction profile at 4.69Å. C) Representative 2D class of the fast-twisting fibril species from a 1024 pixel box. D) Representative 2D class of the slow-twisting fibril species from a 1024 pixel box. E) A slow-twisting fibril image from 2D class after classification using a 384 pixel box in Relion. F) Composite image of 384 pixel box size 2D classes from the slow-twisting fibril polymorph stitched together to show its full twist. G) *De novo* 3D model generated from 2D classes in Relion viewed from the side and H) as a cross section.

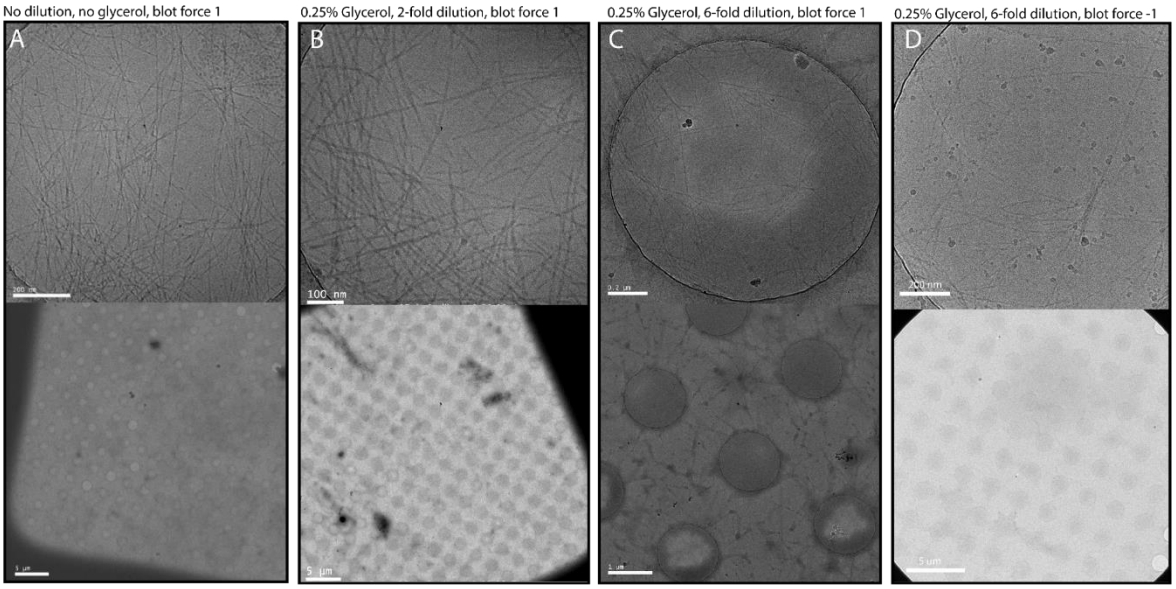


Supplementary Figure 4: ALECT2 fibril stability. A) Nephelometer turbidity readings for the incubated fibril samples over 24 hours. B-E) Negative stain TEM images of the fibrils remaining after 24 hours incubation in fibrillization buffer (B), water (C), 3M urea (D), and 40nM Proteinase K (E).

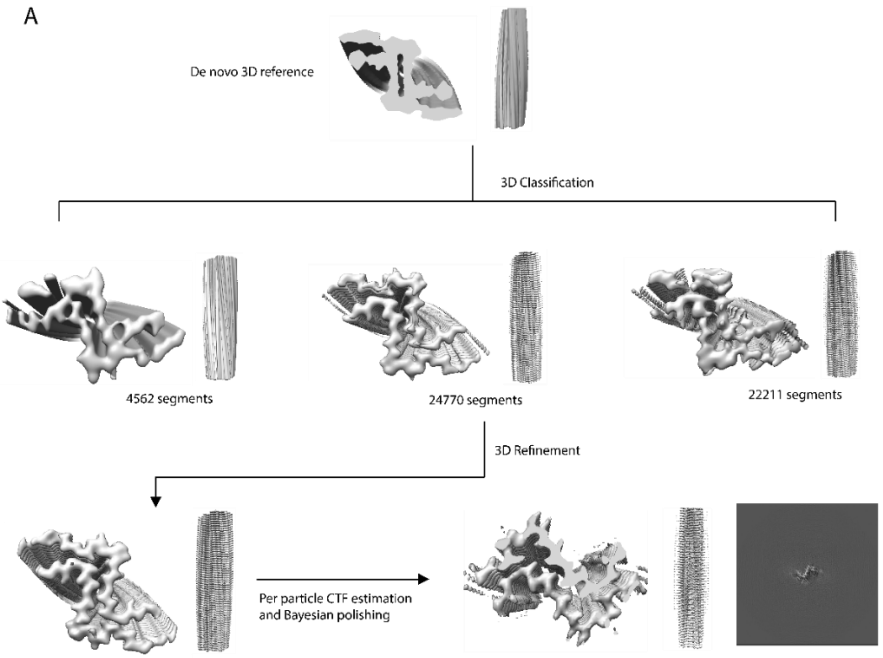


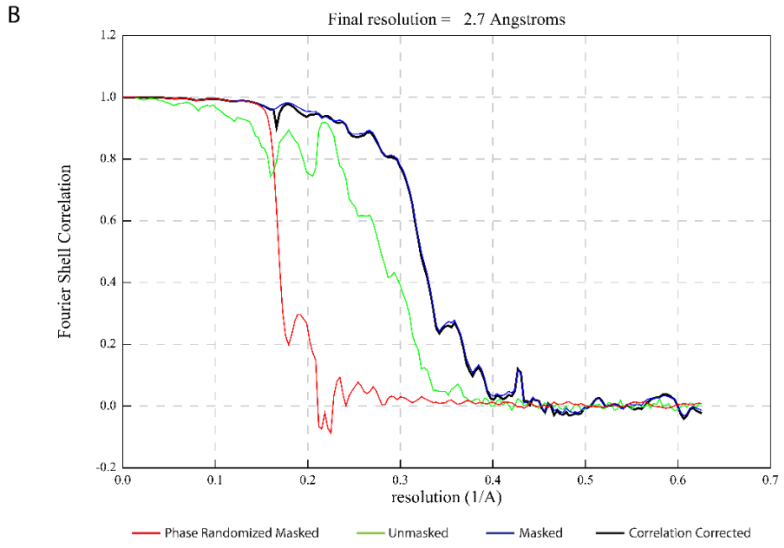


Supplementary Figure 5: Cryo-EM grid optimization. Recombinant fibrils were plunge frozen in buffer at different dilutions and Vitrobot blot settings for optimal particle dispersion, representative micrographs are shown. A) No dilution results in clumping and overcrowding. B) Addition of 0.25% glycerol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

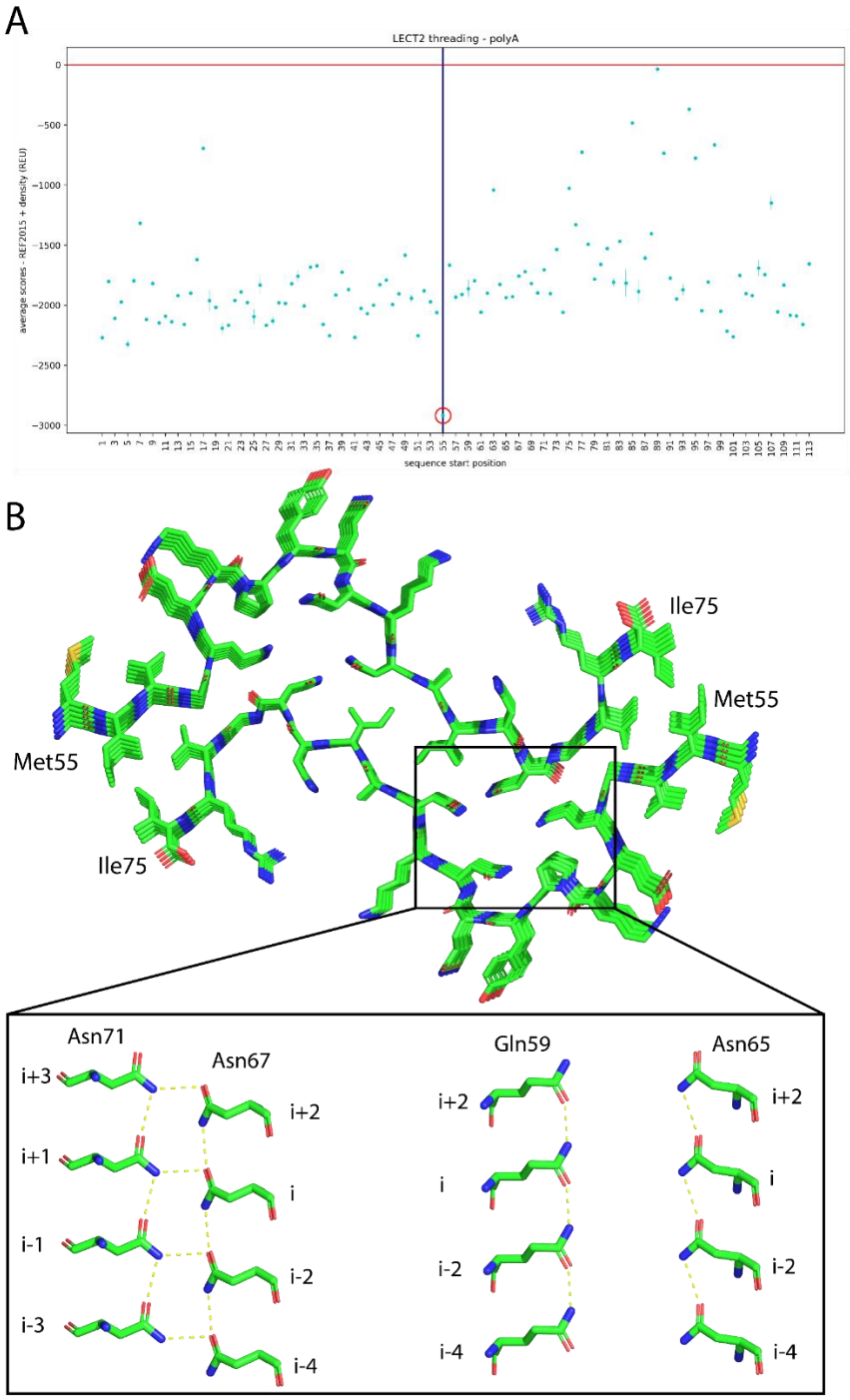


Supplementary Figure 6: Cryo-EM data processing and model comparisons. A) Overview schematic representation of helical reconstruction of fibril species performed in RELION showing maps of fibril sides and cross sections. B) FSC curves between independently refined half-maps which were phase randomized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

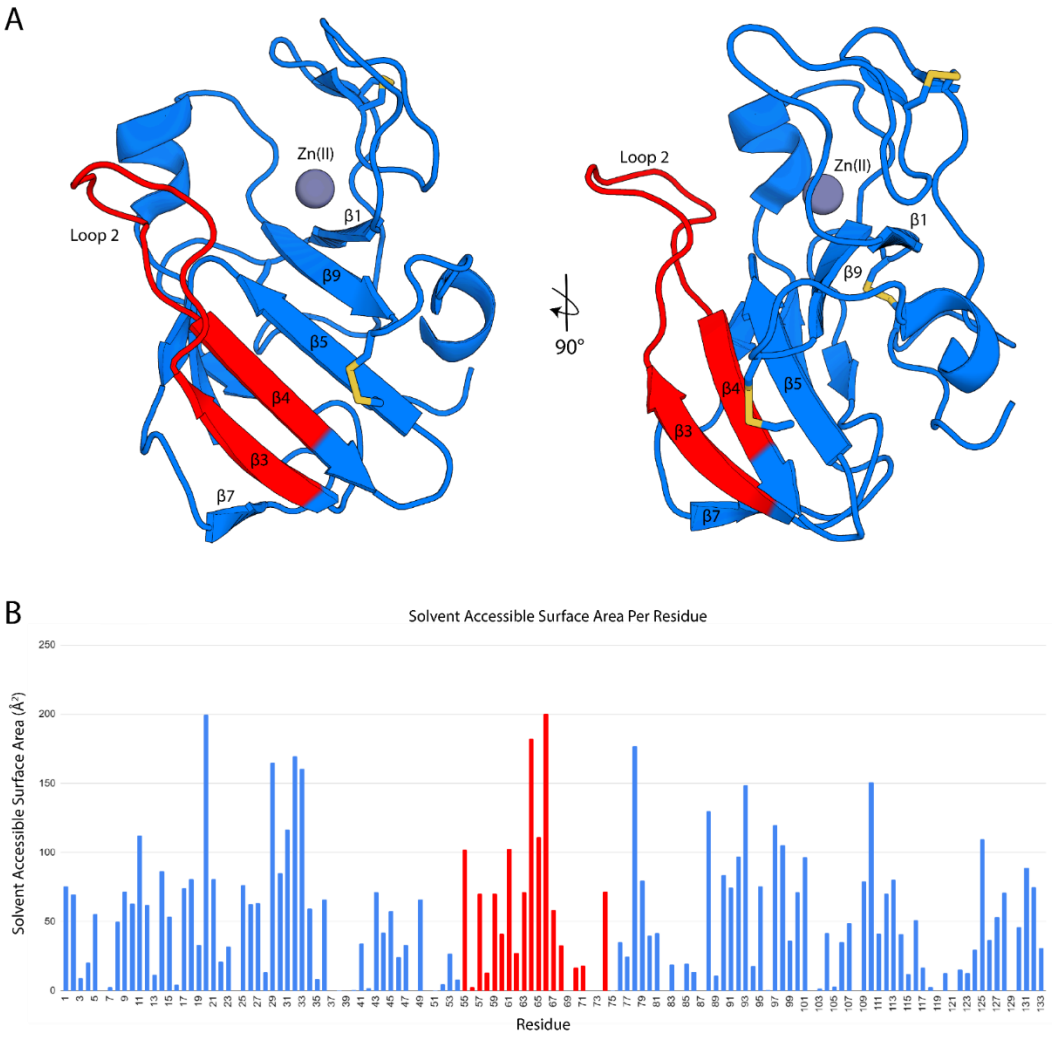




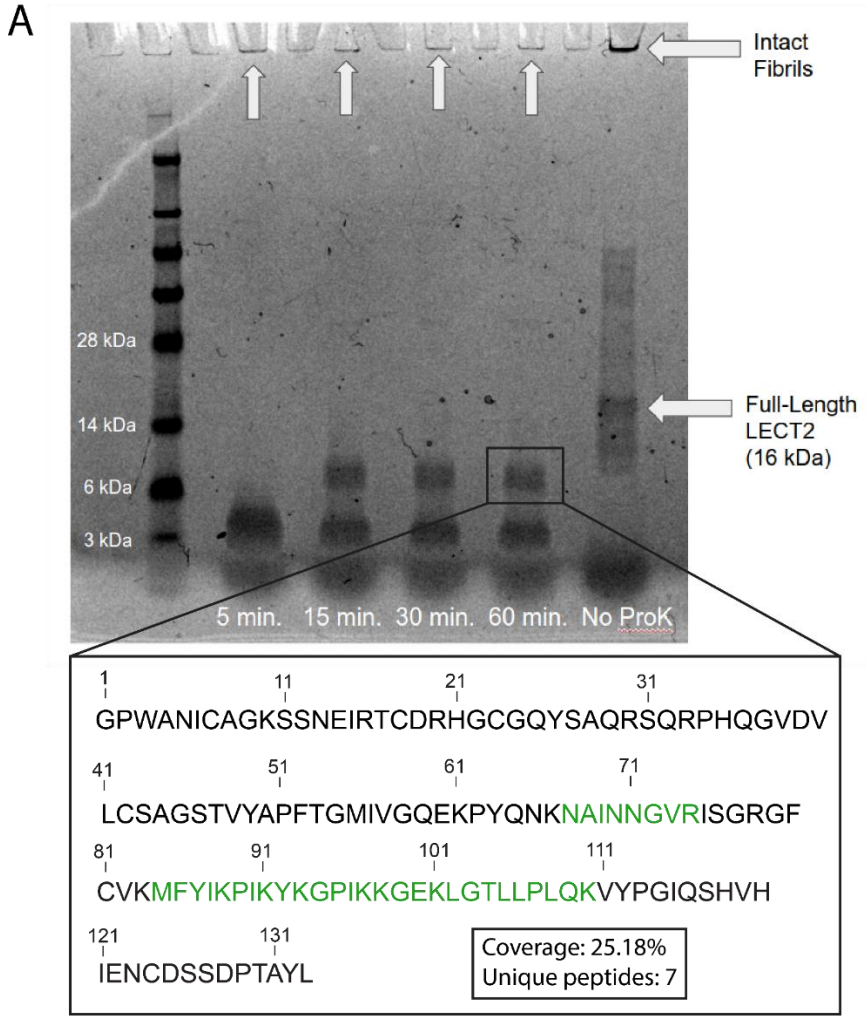
Supplementary Figure 7: Identification of sequence matching ALECT2 fibril structure by threading and energy minimization in Rosetta. A) Threading analysis indicates that the sequence beginning at Met55 is the most energetically favorable among possible assignments for the cryo-EM map (red circle). B) Energy-minimized, fast relaxed model from Rosetta predicts overall fibril core morphology as well as the stabilizing polar ladders in the hydrophilic pockets (inset).



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



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

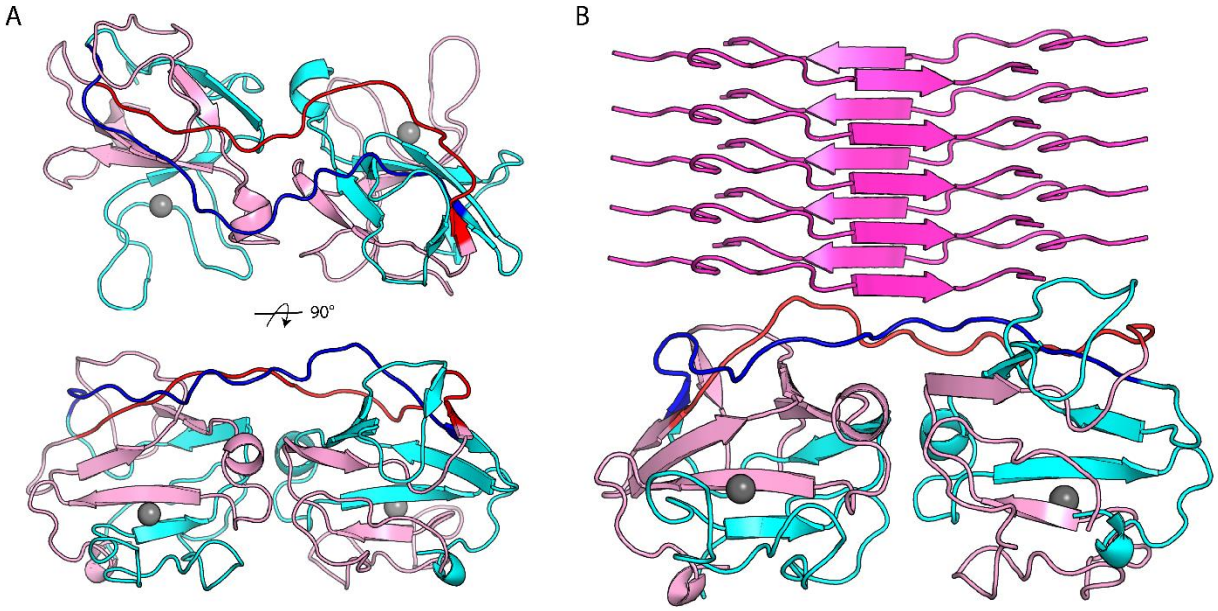


Supplementary Table 2: Summary of Proteinase K digested ALECT2 peptides identified by GeLC-MS/MS.

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

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

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



Supplementary Code: PyRosetta Sequence Threading

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

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

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

np.set_printoptions(threshold=sys.maxsize)

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

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

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

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

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

### sliding window size
window_size = 21

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

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

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

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

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

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

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

    return pose

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

    start_pos = 0
    end_pos = start_pos + wnd_sz

    num_windows = len(x) - wnd_sz + 1

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

    return window_list

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

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

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

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

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

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

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

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

    return scores

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

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

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

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