

UCLA

UCLA Electronic Theses and Dissertations

Title

Developments of electron diffraction phasing methods and their applications

Permalink

<https://escholarship.org/uc/item/9j29759n>

Author

Zee, Chih-Te

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Developments of electron diffraction phasing methods and their applications

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Chemistry

by

Chih-Te Zee

2022

© Copyright by

Chih-Te Zee

2022

ABSTRACT OF THE DISSERTATION

Developments of electron diffraction phasing methods and their applications

by

Chih-Te Zee

Doctor of Philosophy in Chemistry

University of California, Los Angeles, 2022

Professor William M. Gelbart, Chair

Crystallization and phase retrieval are the two perennial problems that challenge any endeavor to retrieve structural information from a crystal. Micro-electron diffraction has emerged in recent years as a complementary method to X-ray diffraction by allowing the diffraction of crystals previously too small for routine X-ray techniques. While micro-electron diffraction has great potential to address barriers in crystallization, issues with phase retrieval precludes the technique's broad utilization. As a nascent technique, many X-ray phasing methods have not been replicated in electron diffraction. In this work, we demonstrate the development and application of racemic crystallography and fragment-based phasing methods for electron diffraction.

The dissertation of Chih-Te Zee is approved.

Todd O. Yeates

Yves F. Rubin

Jose A. Rodriguez

William M. Gelbart, Committee Chair

University of California, Los Angeles

2022

Dedicated to my grandmother, Zee Mao Ai-Hua.

May she rest in peace.

TABLE OF CONTENTS

Chapter 1: Introduction	1
1.1 Structural Studies of Macromolecules	2
1.2 Protein Crystallization: Practical Limitations	4
1.3 Crystallography and Phasing	7
1.4 Phasing Methods	9
1.5 Development of MicroED	13
1.6 Issues with MicroED	15
1.7 Racemic Crystallography – Development and Future	17
1.8 References	20
Chapter 2: Homochiral and racemic MicroED structures of a peptide repeat from the ice-nucleation protein InaZ	31
Chapter 3: Ab initio MicroED structure of racemic macrolittin 70	63
Chapter 4: Conclusion	108
Appendix to Dissertation	109
Appendix A: Structure of amyloid-β (20-34) with Alzheimer’s-associated isomerization at Asp23 reveals a distinct protofilament interface	110
Appendix B: Fjord-Edge Graphene Nanoribbons with Site-Specific Nitrogen Substitution	157
Appendix C: Scale-Dependent Photosaliency and Topotactic Reaction of Microcrystalline Benzylidenebutyrolactone Determined by Electron Microscopy and Electron Diffraction	319

LIST OF FIGURES

Chapter 1:

Figure 1. Optimal crystal dimensions for diffraction techniques 6

Chapter 2:

Figure 1. MicroED diffraction patterns of homochiral L-GSTSTA and racemic GSTSTA 36

Figure 2. Direct methods solution of L-GSTSTA and racemic GSTSTA 37

Figure 3. Phase analysis of MicroED racemic GSTSTA structure 38

Figure 4. Analysis of MicroED racemice GSTSTA structure 39

Chapter 3:

Figure 1. Analysis of macrolittin 70 structure 71

Figure 2. Directional distribution of hydrogen bond in melittin and macrolittin 70 unit cells 74

Figure 3. Refinement of macrolittin 70 structure obtained by using AlphaFold prediction 76

LIST OF TABLES

Chapter 1:

Table 1. Comparison of common phasing methods used for X-ray diffraction	12
Table 2. Atomic Scattering Cross Section: X-ray vs Electron	14

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my family for their love and support through the years.

I would also like to thank my friends and co-workers from my time at UCSD and Ferring Research Institute. I would not have achieved so much without their help and support.

University of California, San Diego

Alec Saitman

Alex Groisman

Arnold L. Rheingold

Edgar Gutierrez

Edward Ronan

Frank Chang

Haim Weizman

Janak D. Dadhaniya

Judy E. Kim

Kristy M. Elbel

Lawrence Wang

Mark Polinkovsky

Michael Galperin

Michael J. Erickstad

Michelle H. Lacoske

Ryan Nam

Silka Cheng

William R. Lotherington

Ferring Research Institute

Aleksandr Rabinovich

Antoine Henninot

Brian Ly

Claudio Schteingart

Daniel W. Carney

Guangcheng Jiang

Halina Wisniewska

Jacek Stalewski

James C. Collins

James T. Patterson

Javier Suarez-Diaz

John M. Nuss

Jolene L. Lau

Katrina Tjhung

Kazimierz Wisniewski

Kirk Kozimski

Monica Mares

Nicky Ferdyan

Wendy J. Hartsock

Last but not least, I would like to thank all the people I met and worked with at UCLA.

Alex Lisker	Gregory Khitrov
Alexander Stevens	Gretchen Guaglianone
Alexandria Brackbill	Hung Pham
Ayesha Hamid	Irena Mendez
Becky Jenkins	Isabel Usón
Brendan Mahoney	James Ziegenbein
Bruce Patterson	Janette Kropat
Calina A. Glynn	Janine Y. Fu
Carlos G. Santiago	Jennifer Miao
Christopher Sue	Jess Soule
Connor Short	Jimmy Haywood
Daniel Anderson	Jose A. Rodriguez
Daniel L. Sun	Jose G. Gonzalez
David S. Eisenberg	Justyna Glode
David Guo	Justin E. Miller
David J. Reilley	Kat Ellis-Guardiola
David J. Tranchemontagne	Kathleen W. Kurgan
Duilio D. Cascio	Kevin A. Murray
Einav Tayeb-Fligelman	Kyle Meador
Eric Z. Pang	Lana Martin
Erick F. Velasquez	Lee Joon Kim
Eva S. Rosker	Lily Taylor
Evelyn Hernandez	Logan S. Richards
Feng Guo	Luke Boralsky
Génesis Falcón	Maeve Nagle

Marcus Gallagher-Jones

Margot E. Quinlan

Maria Carmen

Maria D. Flores

Maria Gutierrez

Marie E. Grossett

Mark Arbing

Melinda Balbirnie

Meaghan A. Valliere

Michael R. Sawaya

Michael J. Collazo

Michael J. Ferracane

Natalie A. Schibrowsky

Nick Baerg

Niko Vlahakis

Nina Harpell

Orlando E. Martinez

Rebeccah A. Warmack

Ricky Medrano Ruiz

Roy D. Pan

Roger Castells Graells

Ryan Lannan

Saken Sherkhanov

Samuel H. Gellman

Samantha Zink

Stasik Popov

Steven G. Clarke

Sum Chan

Sydney J. Alairys

Todd O. Yeates

Tyler Bass

Vincent M. Hipwell

Wilberth A. Narvaez

William M. Gelbart

Wong Hoi Hui

Wonhyeuk Jung

Yolanda L. Li

Yu Chen

Yves F. Rubin

Zachary Hemminger

Chapter Two of this dissertation is a version of a published manuscript: Homochiral and racemic MicroED structures of a peptide repeat from the ice-nucleation protein InaZ. Zee, C., Glynn, C., Gallagher-Jones, M., Miao, J., Santiago, C.G., Cascio, D., Gonen, T., Sawaya, M.R., and Rodriguez, J.A. 2019. Reproduced with permission.

Chapter Three of this dissertation is a version of a manuscript in preparation: Ab initio MicroED structure of racemic macrolittin 70. Zee, C., Kurgan, K.W., Sawaya, M.R., Usón, I., Gellman, S. H., and Rodriguez, J.A.

Appendix A of this dissertation is a version of a published manuscript: Structure of amyloid- β (20-34) with Alzheimer's-associated isomerization at Asp23 reveals a distinct protofilament interface. Warmack, R.A., Boyer, D.R., Zee, C., Richards, L.S., Sawaya, M.R., Cascio, D., Gonen, T., Eisenberg, D.S., and Clarke, S.G. 2019. Reproduced with permission.

Appendix B of this dissertation is a version of a published manuscript: Fjord-edge graphene nanoribbons with site-specific nitrogen substitution. Li, Y., Zee, C., Lin, J., Basile, V., Muni, M., Flores, M., Munárriz Tabuena, J., Kaner, R., Alexandrova, A., Houk, K., Tolbert, S., Rubin, Y. 2020. Reproduced with permission.

Appendix C of this dissertation is a version of a published manuscript: Scale-Dependent Photosensitivity and Topotactic Reaction of Microcrystalline Benzylidenebutyrolactone Determined by Electron Microscopy and Electron Diffraction. Hipwell, V. M., Zee, C., Rodriguez, J. A., Garcia-Garibay, M.A. 2022. Reproduced with permission.

Finally, I would like to thank the University of California, Los Angeles for funding through the George Gregory Fellowship and the Dissertation Year Fellowship.

VITA

Education

- 2018 M.S. in Chemistry, University of California at Los Angeles, Los Angeles, CA
2014 B.S. in Molecular Synthesis, University of California at San Diego, La Jolla, CA

Honors and Awards

- 2021 Dissertation Year Fellowship, UCLA
2020 George Gregory Fellowship, UCLA

Selected Publications

- Hipwell, V. M., Zee, C., Rodriguez, J. A., Garcia-Garibay, M.A. Scale-Dependent Photosensitivity and Topotactic Reaction of Microcrystalline Benzylidenebutyrolactone Determined by Electron Microscopy and Electron Diffraction. *Cryst. Growth Des.* 2022, **22** (3):1533-1537.
- Zee, C., Saha, A., Sawaya, M.R., and Rodriguez, J.A. *Ab initio* determination of peptide structures by MicroED. *Methods Mol Biol.* 2021, **2215**: 329-348.
- Li, Y., Zee, C., Lin, J., Basile, V., Muni, M., Flores, M., Munárriz Tabuenca, J., Kaner, R., Alexandrova, A., Houk, K., Tolbert, S., Rubin, Y. Fjord-edge graphene nanoribbons with site-specific nitrogen substitution. *J. Am. Chem. Soc.* 2020, **142** (42): 18093-180102.
- Warmack, R.A., Boyer, D.R., Zee, C., Richards, L.S., Sawaya, M.R., Cascio, D., Gonen, T., Eisenberg, D.S., and 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**: 3357.
- Zee, C., Glynn, C., Gallagher-Jones, M., Miao, J., Santiago, C.G., Cascio, D., Gonen, T., Sawaya, M.R., and Rodriguez, J.A. Homochiral and racemic MicroED structures of a peptide repeat from the ice-nucleation protein InaZ. *IUCrJ.* 2019, **6** (2): 197-205.
- Hattne, J., Shi, D., Glynn, C., Zee, C., Gallagher-Jones, M., Martynowycz, M.W., Rodriguez, J.A., and Gonen, T. Analysis of global and site-specific radiation damage in cryo-EM. *Structure.* 2018, **26** (5): 759-766.
- Lee, K., Elliott, H.L., Oak, Y., Zee, C., Groisman, A., Tytell, J.D., and Danuser, G. Functional Hierarchy of Redundant Actin Assembly Factors Revealed by Fine-Grained Registration of Intrinsic Image Fluctuations. *Cell Systems.* 2015, **1** (1): 37-50.

Invited Oral Presentations

- Zee, C., Rodriguez, J.A. "MicroED: Developments in Electron Diffraction Phasing." American Peptide Society E-Seminar Series. Online, October 2021.
- Zee, C., Kurgan K.W., Rodriguez, J.A., and Gellman, S.H. "MicroED: Developments in Electron Diffraction Phasing." Gordon Research Seminar – Rising Tide of Peptides as Research Tools, Therapeutics and Biomaterials. Oxnard, CA, October 2022.

CHAPTER 1

Introduction

1.1 Structural Studies of Macromolecules

Determining the atomic structure of chemical compounds has provided science with great insight into materials, biological and otherwise. The three-dimensional (3D) structures of chemical species dictate the type of interactions that they can engage in, and knowing these structures help us understand the mechanism of action of molecules in the world around us. Understanding how these compounds interact or potentially interact with their environment has led humanity to numerous breakthroughs in medicine, energy, materials, etc, including our extensive use of zeolites for important industrial activities ranging from petroleum refinement to water filtration^{1,2}; development and analysis of semiconductors^{3,4}; and optimization of drug candidates^{5,6} to treat disease. Many complementary techniques that have been developed to elucidate structures. Some of these techniques are X-ray crystallography, electron microscopy (EM), electron paramagnetic resonance (EPR), mass spectrometry, and nuclear magnetic resonance (NMR), and each have their strengths and limitations (Table 1 – table of strengths, weaknesses, resolution limits). All of these techniques have made important contributions and all have garnered at least one, if not multiple Nobel prize awards^{7,8,9,10}. In particular, advances in structural biology have been repeatedly recognized by the Nobel committee. Rather than any favoritism in selection, it highlights the difficulties present in studying the atomic structure of biological molecules that set the sub-field apart. In fact, the 1946 Nobel Prize in Chemistry was awarded to James B. Sumner for the discovery that enzymes can be crystallized¹¹, a routine process today, which demonstrated the arduous nature of studying the structures of macromolecules and how each step in the process we take for granted today required immense effort to establish.

Through decades of scientific progress, two techniques have become the gold standards for structural biology: crystallography and cryo-electron microscopy (cryo-EM). Crystallography has been a mainstay of structural biology ever since Kendrew and Perutz' research groups reported the structures of myoglobin and hemoglobin. The first clear structural interpretation of myoglobin and hemoglobin at 6.0-Å and 5.5-Å resolution came in 1959 and 1960, respectively¹².

¹³. These structures were followed with a 2-Å resolution structure for myoglobin in 1960¹⁴ and a 2.8-Å structure for hemoglobin in 1968¹⁵. Besides crystallography, recent advances in cryo-EM have enhanced resolving power to approach atomicity^{16, 17}, increasing its utility in the field. Due to the power of these techniques, great effort has been made to make cryo-EM and crystallography accessible, namely: creation of a centralized database (Protein Data Bank), standardization of data formatting, and universalization of quality control requirements for data deposition prior to publication. This commitment to provide open access to structural data has led to a virtuous cycle of increased usage of cryo-EM and crystallography techniques, accompanied by methodological progression and improvement.

In conjunction with progress in wet lab technologies, the structural biology community is also pushing the envelope with computation to develop software that can predict folded structures of macromolecules from their amino acid or nucleic acid sequence. However, the creation of a folding algorithm is a difficult endeavor: assuming two configurations are possible for each amide bond in a 101 amino acid protein, there would be 10^{30} possible folded states. If each configuration takes a picosecond to sample, then it would take 10^{18} seconds to try all possible states, equivalent to billions of years, which is longer than Earth's existence¹⁸. Yet, in nature, many proteins, some much longer than 100 amino acids, spontaneously fold into their proper forms on millisecond or even microsecond timescales. The paradox of correct and spontaneous protein folding was first postulated by Cyrus Levinthal in 1969 and bears his name: Levinthal's paradox¹⁹. Ultimately, structural biologists continue to resolve this paradox in order to create folding algorithms that can predict a protein's three-dimensional structure based on its amino acid sequence²⁰. Since 1994, the community has gathered for a biannual worldwide experiment and competition called the Critical Assessment of protein Structure Prediction (CASP), where research groups test their modelling and prediction software and assess their performance objectively²¹. In 2018 and 2020, AlphaFold, an artificial intelligence (AI) program developed by Google took center stage by winning both CASP13 and CASP14, respectively^{22, 23}. Instead of using energy minimization to

predict structure, AlphaFold is a machine learning algorithm trained with structures available in the Protein Data Bank to fold an inputted primary amino acid sequence into a 3D protein structures^{24,25}. AlphaFold allows structural biologists to procure initial models to help process data recalcitrant to *ab initio* data processing²⁶ and some labs have already demonstrated the accuracy of AlphaFold models by using them as starting models to solve novel structures^{27,28}. Advances in wet lab and *in silico* methods will revolutionize structural biology by allowing the elucidation of more three-dimensional structures than ever before.

While powerful, all the above methods have their weaknesses. While AlphaFold offers never before seen prediction power, it has 1) difficulty predicting outlier structures that fall outside of the machine learning training set²⁹ and 2) can only predict a single state for a structure even if the data suggests the presence of multiple states.³⁰ Thus, AlphaFold's *in silico* structures still require validation through wet lab experiments³¹. To validate these *in silico* structure predictions, techniques such as electron microscopy can be used; however, cryo-EM has a lower limit of detection (LLOD) that hinders the collection of high-quality data from proteins below 100 kilodaltons (kDa)^{32,33}. While technological advancements have steadily decreased the LLOD over the years, this threshold still exists and prevents structural biologists from studying a large portion of both the prokaryotic and eukaryotic proteome³⁴. Efforts are being made to circumvent this problem, but these technologies are still being developed and elucidation of protein structures below 100 kDa by cryo-EM remains nonroutine^{34,35,36}.

1.2 Protein Crystallization: Practical Limitations

For crystallography, complications of retrieving structural information from diffraction data have centered on crystallization and solving the phase problem^{37,38}. In fact, the Protein Structure Initiative (PSI), a National Institutes of Health (NIH) program (2000-2015), was initiated as a collective effort to develop methods to elucidate “three-dimensional, atomic-level structures of most proteins easily obtainable from knowledge of their corresponding DNA sequences.”³⁹ A 2015

report, collating the results of this massive effort, estimated that only around 18% of the targeted, purified proteins produced diffraction-quality crystals, indicating that crystallization was the biggest bottleneck to the entire initiative pipeline⁴⁰. One of the most common shortfalls that resulted in such a low success rate is the inability to grow a crystal large enough to yield usable data³⁷.

Small crystals create problems for crystallographers for two major reasons: small scattering cross sections and radiation damage. First, a photon from an X-ray beam must collide with the electron cloud of a sample atom, a cross section; thus, the probability of any scattering event taking place is low because most X-ray photons pass through the sample without scattering⁴¹. Within this relatively low number of scattering events, even fewer photons scatter elastically⁴². Due to an X-ray beam's small diffraction cross section, scientists need an appropriately large crystal to achieve signal strong enough to constitute meaningful data, especially to high resolution. As a result, simply achieving crystallinity is insufficient. A crystal must be grown to a sufficient size while retaining sufficient regularity in its organization.

In addition to providing strong reflections, large crystal sizing also hedges against the degradation of diffraction signal due to radiation damage. Radiation damage arises from the usage of powerful radiation and long exposures to compensate for the weak observed signals. While the elastic Rayleigh scattering produces the desired signal, inelastic Compton scattering is more probable and results in incremental absorption of energy by the atoms within the crystal. These absorptions accumulate and eventually destroy covalent bonds, individual unit cells, and eventually the crystal^{43, 44}. As a result, diffraction data collection is often a race against time. Enough data must be collected from a crystal before excitations destroy the crystal lattice, potentially disallowing acquisition of a complete diffraction data set to high spatial resolution.

Rather than allowing crystal sizes to limit the outcome of projects, structural biologist have worked in conjunction with national labs to improve X-ray data collection capabilities. New technologies such as Microfocus X-ray beamlines have enabled the collection of data from

smaller crystals by significantly reducing the background scattering incurred by using an X-ray beam wider than the crystal^{45, 46}. This technology granted access to structures from crystals that would have been too small for conventional X-ray diffraction analyses. However, microfocus beams still do not yield usable diffraction data for crystals in the sub-micron scale⁴⁷.

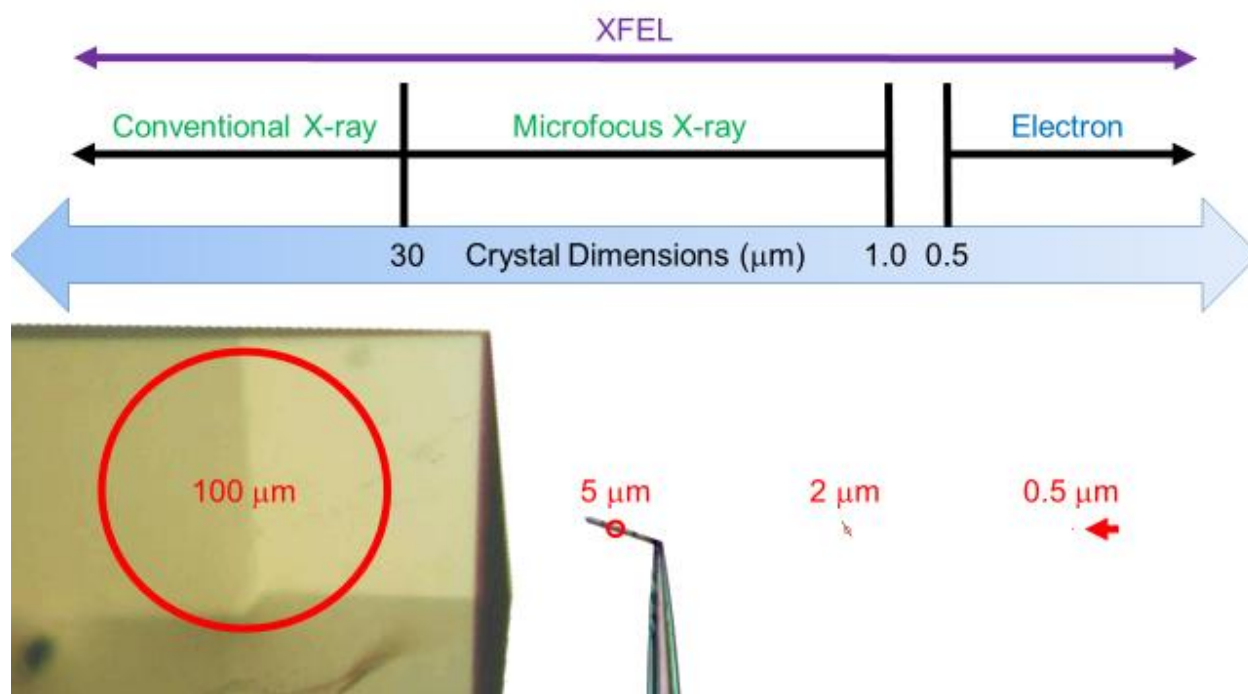


Figure 1: Diagram illustrating the ranges of crystal dimensions that each single-crystal diffraction techniques are effective in. Electron diffraction starts losing viability with crystals thicker than 0.5 μm ⁴⁸. FMX, the micro-focusing Frontier MX beamline in sector 17-ID-2 at NSLS-II is capable of a 1 μm x 1.5 μm X-ray beamline⁴⁹. Crystal Images (courtesy of Michael R. Sawaya) are to scale relative to each other. Additionally, the red circles and labels denote an appropriately sized condensed beam (X-ray or electron) for each crystal size.

In response to the desire to collect data from smaller crystals, stimulated/promoted by the inability to grow larger crystals, X-ray free-electron laser (XFEL) beam lines were developed and use powerful X-ray laser pulses to yield strong diffraction images from crystals far too small for traditional single-crystal data collection^{50, 51}. In a technique called serial crystallography, each frame collected represents a 2D projection of the reciprocal lattice points satisfying the Bragg condition at a single crystal orientation; a large number of these 2D projections are then collated to reconstruct a 3D diffraction data set. The diffraction intensities resulting from this reconstruction

can be used like a standard single-crystal diffraction data set to yield a 3D atomic structure. While XFEL beamlines enable data collection from “micro” crystals, serial crystallography experiments require thousands, sometimes millions, of homogeneously sized protein crystals. The requirement of XFEL experiments for abundant sample hinders our ability to study samples that are scarce or are difficult to produce with homogenous crystals at such scales. In summary, both radiation damage and a small diffraction cross section necessitate the usage and, thus, the growth of large crystals. And, while efforts have been made to broaden the spectrum of collectable crystal dimensions, a routine method to diffract nanometer scale crystals remain highly desirable⁵². Moreover, even if this challenge of obtaining larger crystals can be overcome, another bottleneck follows: phasing.

1.3 Crystallography and Phasing

Proteins imaged by electron microscopy require grouping and alignment via computational experiments. After identifying and classifying particles for different rigid body positions and orientations, single particle experiments have to account for translational, rotational, and vibrational movement of atoms within the protein that are not restricted by bonding geometries and intramolecular interactions from protein folding. For context, a carbon-carbon single bond of an ethane molecule rotates approximately 10^{11} times per second at room temperature⁵³. Even at cryogenic temperatures (-196°C), the same bond would be expected to rotate approximately 10^5 times per second (assuming a halving of rotation speed every 10°C). These atomic movements create minor incongruities between particles parsed from electron micrographs. When programs such as RELION and cryoSPARC group and align these particles in order to boost the data's signal to noise ratio, these incongruities propagate as a major source of error, preventing structural biologists from solving atomic-resolution structures via electron microscopy unless the protein is unusually rigid^{54, 55}.

In contrast to single particle experiments, crystallography utilizes crystals which form from the physical alignment of molecules into ordered lattices that minimally exhibit translational symmetry, if not rotational symmetries. Diffraction experiments exploit this physical attribute whereby electromagnetic radiation is scattered from copies of the molecule, ordered in regular lattices, and this configuration of molecules serves to amplify the diffraction signal in direct proportion to the number of unit cells that compose the crystal. Any incongruities between unit cells that result from atomic motion causes the signal to become averaged and blurred, while diffraction from atoms that are identically positioned in each unit cell leads to amplification. Additionally, crystal packing results in systemic intermolecular interactions and provides additional restraints on atomic positions beyond bonding geometries and intramolecular forces. As a result, the quality of protein structures elucidated via crystallography is dependent on the uniformity of a crystal's unit cells and structures can approach atomic-resolution if the lattice is highly organized.

While crystallography more frequently produces structures at higher resolutions than electron microscopy, a good dataset does not guarantee structure elucidation. Each measured intensity in a reciprocal lattice (notated with Miller indices, hkl) is a summation of all scattering vectors from every atom within the unit cell at that hkl coordinate. Although each atom contributes to all electromagnetic reflections, the relative contributions change according to its position (thus, scattering angle) in the unit cell. As it turns out, the amplitudes can be recorded as reflection intensities on the detector, but the phase of the electromagnetic wave is lost. There is no instrument available that can record the phases of scattered radiation, and these phases have a greater influence on our ability to locate atoms than do the amplitudes of the reflected radiation. This problem is known as the "phase problem" by crystallographers^{56, 57}. To resolve this issue, crystallographers have developed a myriad of powerful methods to carry out phase retrieval. In general, these phasing techniques seek to make an educated guess of potential atom positions, and these positions are assessed by checking the agreement between the model's Fourier

transform and the intensities observed. These methods include direct methods, molecular replacement (including fragment-based phasing), isomorphous replacement, and anomalous scattering, among others.

1.4 Phasing Methods

Direct methods combine brute force calculations, crystallographic statistics, and the Patterson function to determine atomic positions within a unit cell^{58, 59, 60, 61, 62}. This method usually requires a high-resolution dataset of at least 1.1 Å⁶³. High resolution data allows for the calculations to differentiate features that are spaced apart by the approximate length of an average covalent bond, which is the expected minimum distance between atoms in a unit cell. Direct methods are powerful for retrieving accurate phases; however, its success depends strongly on the crystal quality and size of the unit cell. First of all, larger proteins rarely diffract to high resolutions, making this method not useful for many macromolecules of interest. Additionally, even if the crystal of a large molecule diffracts to sufficient resolution, the number of calculations necessary to retrieve the phases by brute force calculation increases exponentially based on the number of atoms present in the unit cell. While powerful, these disadvantages limit direct methods in the study of large proteins.

Isomorphous replacement is the first phasing method proven to be successful for determining macromolecular crystal structures. It is an *ab initio* method, meaning it does not require any prior estimates of the atomic structure. Isomorphous replacement was used to solve the first three-dimensional macromolecular structure in 1956⁶⁴, and it remains in use today^{65, 66}. This phasing technique requires introducing elements, such as cesium, iodine, gold, or mercury into a crystal, with drastically higher scattering amplitudes compared to organic elements. The addition is usually achieved by soaking the crystal with a doped version of the sample's crystallization liquor, but covalent substitution can also be carried out by incorporating heavy atoms via synthesis or unique protein expression protocols⁶⁷. Differences between the diffraction

intensities of a native dataset and a doped dataset are called “isomorphous differences,” which are used for phasing. If the soaked or covalently modified crystal is isomorphous or has the same unit cell parameters as the unmodified crystal, isomorphous differences are used as coefficients in calculating an isomorphous-difference Patterson map. If successful, the map will display peaks related to position(s) of heavy atoms within the modified cell. These initial positions can then be used to estimate the phases of all the reflections in a dataset. This initial guess can be refined and has the potential to provide a reasonably accurate initial electron density map that can then be utilized to build and refine the correct structure.

Anomalous diffraction is another phasing technique that utilizes heavy atoms incorporated into a unit cell and is often used in conjunction with isomorphous replacement to enhance the accuracy of initial electron density solutions^{68,69}. Since the typical composition of macromolecular crystals consists of elements with low atomic mass, such as carbon, oxygen, and nitrogen, macromolecular crystals typically do not absorb light at the X-ray wavelength. The lack of absorption is the basis of a foundational assumption in X-ray crystallography: electron clouds are in their lowest energy state and have a centrosymmetric distribution around the nucleus. The centrosymmetry of electron clouds leads to centrosymmetric diffraction intensities expressed by Friedel’s law: the intensities of reflections observed at (h, k, l) equals the intensity observed at the centrosymmetrically-related lattice position $(-h, -k, -l)$. The pair of reflections (h, k, l) and $(-h, -k, -l)$ are called Friedel mates. Unlike the lighter elements, the excitation energy necessary to promote electrons from the innermost electron shell of heavy atoms overlaps with the photon wavelengths used for X-ray crystallography experiments. In fact, the absorption is so pronounced that the term K-edge was coined to denote the specific X-ray wavelengths that heavy elements absorb. When the electrons in the innermost electron shell, the K-shell, are excited, the electron cloud becomes polarized and no longer diffracts light isotropically. This aberration results in a measurable violation of Friedel’s law in the form of a non-centrosymmetric dataset at the K-edge wavelength of a heavy atom. Differences between the intensities of Friedel mates are called

anomalous differences, and these arise purely from the heavy atoms incorporated into the unit cell. Assuming the cell vectors remain congruent, the anomalous differences can be used to retrieve the phases with or without additional information from isomorphous replacement. Additionally, incorporating a plethora of heavy atoms and collecting more datasets at different wavelengths can give additional phasing power to this technique by reducing phase ambiguity and result in a more accurate initial phase solution.

Another method of phasing, called molecular replacement, starkly contrasts with the *ab initio* techniques above in that it requires prior knowledge of a reasonably accurate atomic search model^{70, 71}. This method tests different orientations and positions of the search model to find the one most consistent with the measured diffraction pattern. The rotated and translated model provides the starting phases for calculation of a density map. The success chance for phasing by molecular replacement increases in proportion to the search model's similarity to the crystallized molecule. However, since structural similarity is required, a similar structure must have been solved by another method. In fact, if traced back to its origin, most search models used in molecular replacement are derived from a structure solved by an *ab initio* method such as isomorphous replacement. So while molecular replacement is the most prevalent phasing method, its application may be limited when investigating with novel proteins and their structures.

Fragment-based methods, a relatively nascent form of molecular replacement, have been developed in the last decade for use in cases where no appropriate search model exists. Fragment-based methods were first used in 2009 by Rodriguez, *et al.*⁷² and have since been used to solve more than 100 of the structures deposited in the Protein Data Bank (PDB)^{73, 74}. Fragment-based phasing uses ligand libraries containing short fragments of well-established secondary structure elements, such as small alpha helices and beta sheets, to act as partial models. These fragments are placed within the unit cell by software to determine whether those atom positions, when Fourier transformed, would produce amplitudes (F_{calc}) similar to the observed (F_{obs}). As more and more fragments are added, the Fourier synthesis is iteratively checked to see if the

solution converges to the F_{obs} provided by the collected data. For smaller unit cells where the search models constitute a large percentage of a cell's content, fragment-based approaches are inherently more biased towards the initial search model. On the other hand, for larger cells where each fragment constitutes a much smaller percentage of the overall cell content, which may be one or multiple proteins, fragment-based methods can be argued to be an *ab initio* method with the benefit of less initial model bias associated with small fragments.

Comparison of Common Phasing Methods

Method	Strengths	Weaknesses	Model Bias
Direct Methods	<ul style="list-style-type: none"> Does not require additional wet lab experimentation Excellent for small molecules 	<ul style="list-style-type: none"> Requires high-resolution data (~1.1 Å or better) Low feasibility for larger molecules such as proteins 	none; <i>ab initio</i>
Molecular Replacement	<ul style="list-style-type: none"> Flexible: can be used with various data resolutions Does not require additional wet lab experimentation 	<ul style="list-style-type: none"> Relies on the usage of a previously solved structure that is similar to the unknown structure as a search model 	high
Fragment-based Phasing	<ul style="list-style-type: none"> Subset of molecular replacement; no additional wet lab experimentation Low model bias if fragments represent a low percentage of asymmetric unit 	<ul style="list-style-type: none"> Requires the generation of an appropriate fragment library Strong model bias if fragments represent a high percentage of asymmetric unit 	variable
Isomorphous Replacement	<ul style="list-style-type: none"> Can be combined with anomalous dispersion to improve phasing power Can be repeated with different heavy atoms to increase the accuracy of initial phases 	<ul style="list-style-type: none"> Requires incorporation of heavy atom(s) via additional wet lab work Unit cell may change during heavy atom incorporation, invalidating method 	none; <i>ab initio</i>
Anomalous Dispersion	<ul style="list-style-type: none"> Can be combined with isomorphous replacement to improve phasing power Can be repeated at multiple wavelengths to increase the accuracy of initial phases 	<ul style="list-style-type: none"> Requires incorporation of anomalous atom(s) via additional wet lab work Unit cell may change when incorporating anomalous atom, invalidating method 	none; <i>ab initio</i>

Table 1: Comparison of the commonly used X-ray phase retrieval methods.

Despite the variety of phase retrieval methods enumerated above, phasing remains a serious problem in crystallography. As alluded to before, statistical analysis of compiled data from

the PSI shows that phasing constitutes the second major bottleneck in their structural elucidation pipeline³⁸. Difficulties in retrieving phase information and producing crystals large enough for diffraction prevent many 3D structures from being solved via crystallography. The PSI program clearly reflects this conundrum: only 18% of targeted proteins produced crystals capable of high-resolution diffraction and structures of only 3% of all targeted proteins were solved and deposited into the PDB. Although not statistically analyzed, it is very likely that crystals were produced for a significant number of proteins but they were simply too small to be compatible with standard X-ray diffraction methods.

1.5 Development of MicroED

In response to the shortcomings of methods of X-ray diffraction, a complementary method has been expanded upon and optimized for macromolecular protein diffraction in the last decade: micro-electron diffraction (MicroED)^{75, 76}. Building upon 2D diffraction work by many scientists from various fields^{77, 78, 79, 80}, this method utilizes a transmission electron microscope (TEM), set up in diffraction mode, to diffract a 3D crystal in a similar manner compared to an X-ray diffraction experiment. However, due to the much larger diffraction cross section of electrons (Table 1), TEM diffraction experiments can produce usable diffraction patterns with crystals similar in size to the ones used in XFEL experiments⁸¹. Crystals with proportions in the sub-micron range can now be diffracted and analyzed by structural biologists, drastically decreasing the crystal size threshold for routine diffraction experiments. This advantage of MicroED makes it an extremely powerful technique, complementary to the X-ray methods that have been honed by crystallographers in the last century.

Atomic Scattering Cross Section: X-ray vs Electron

Energy	H	C	N	O	S
X-ray (8 KeV)	6.551×10^{-29}	9.127×10^{-27}	1.758×10^{-26}	3.090×10^{-26}	5.039×10^{-25}
Electron (200 KeV)	2.261×10^{-24}	5.005×10^{-23}	5.395×10^{-23}	5.683×10^{-23}	2.161×10^{-22}

Table 2: All cross section values in m^2 . All values calculated with NIST Standard Reference Databases for Photon or Electron Cross Sections. 8 KeV was selected for X-ray as it is close to Copper K- α . 200 KeV was selected for electron as it is a common max energy for TEMs used for electron diffraction.

Since scientists are executing electron diffraction on equipment that is optimized for imaging rather than diffraction, data collection problems have arisen. For example, the maximum tilt angle for many electron microscope stages is around $55 \pm 5^\circ$, drastically limiting amount of collectible data. Moreover, even if newer scopes expand the available tilt range, as long as crystals are deposited on a 2D support, collecting 360° of data from a single crystal is physically impossible. These physical restrictions mean that data set completion usually require merging diffraction patterns from multiple crystals. Fortunately, this problem has precedence since it is very similar to the problems faced by X-ray crystallographers before cryo-cooling was introduced for data collection. In those cases, where X-ray diffraction experiments were conducted at room temperature, the damage suffered by the sample from X-ray excitation was great enough where datasets from multiple crystals were merged together to reach high enough completion for a usable dataset. However, in electron diffraction, it is possible for some crystal wedges to remain uncollectable due to orientation bias of crystals, caused by the differences in growth rates of each crystal face⁸². For example, a crystal with plate morphology would lead to difficulty in collecting diffraction from Bragg planes orthogonal to the width of the plate on an electron microscope grid vs. in a loop for an X-ray experiment. As a result, the completion of a merged electron dataset will almost always be lower than that of an X-ray counterpart. Another hardware limitation stems from available detector sizes. Synchrotron detectors are typically made by tiling many small detector panels to form a large detector. These detectors are sized to allow the collection of high-resolution data at the detectors' edges while maintaining separation of reflections observed in the reciprocal

space of large unit cells. As it turns out, such methods to enlarge electron detectors are currently not being pursued for electron microscopes. Thus, for larger unit cells, sacrifices in resolution are necessary to maintain separation between reflections during MicroED. While these issues do affect data collection and processing, they are a result of physical and instrumental limitations. Just like with cryo-EM, these issues will be resolved over time with the improvement and customization of microscopes to better accommodate the evolving technique.

1.6 Issues with MicroED

On the other the hand, there are issues with MicroED that cannot be simply resolved by hardware advancements. For example, while X-rays have a small diffraction cross-section and most photons pass through a sample without scattering, electron beams have cross sections so large that dynamic scattering is frequently observed. Therefore, electrons that undergo multiple scattering events contribute significantly to the observed intensities^{83, 84}. The absence of dynamic scattering in X-ray experiments and the near monochromatic beam generated at synchrotron beamlines allows crystallographers to confidently assume that the observed X-ray intensities result from elastic scattering. On the other hand, during an electron diffraction experiment, crystallographers cannot confidently disambiguate between electrons that have inelastically or elastically scattered. Both dynamic scattering and inelastic scattering decrease the accuracy of observed intensities for electron crystallography. Moreover, both increase the calculated signal backgrounds during data processing, and when subtracted from the observed amplitudes these undesired scattering events create a phenomenon unseen in X-ray diffraction: negative intensities, impossible occurrences in a perfect diffraction experiment. Other than the coordinates of the reflections, their amplitudes are the only other information gathered in these experiments. Inaccuracies in the observed intensities complicate any attempt to develop methods for electron diffraction by replicating methods from X-ray experiments.

In addition to the inaccuracies in the intensities that result from dynamic and inelastic scattering, differences in scattering mechanics during electron experiments lead to issues in electron phasing vs X-ray phasing. Unlike the collision-based scattering observed in X-ray experiments, electrons scatter due to Coulombic repulsion and attraction in electron diffraction⁸⁵. So, while X-ray structure factors scale linearly with an element's atomic number, electron scattering is affected by relativistic considerations, most notably shielding. The standard way to convert from X-ray to electron scattering factors is to use the Mott-Bethe formula; however, this formula cannot, without modification, account for two important factors: atomic charge and molecular bonds^{86,87}. Due to the Coulombic nature of electron scattering, the change in scattering contributed by a change in charge is not a simple addition or subtraction of the number of electrons, as is the case for Rayleigh scattering in X-ray diffraction. Moreover, unlike in X-ray experiments, bonding electrons contribute significantly to the resulting diffraction pattern in electron experiments. On one hand, the scattering contribution of electrons in molecular orbitals allows electron diffraction to detect hydrogens at lower resolutions than X-ray diffraction^{88, 89}. On the other hand, this property completely derails a key estimation normally assumed in X-ray diffraction: atoms are discrete spheres. As a result, any hypothesis that an assumption in X-ray experiments translates to their electron equivalent cannot be taken for granted and must be tested.

As a result of these difficulties, the translation of X-ray phasing methods for usage in electron experiments has been slow. Prior to 2017, only two methods had been proven to be successful for phasing MicroED data: direct methods and molecular replacement⁹⁰. As previously discussed, both of these methods have significant limitations. Large proteins are not only more difficult to crystallize than small molecules, but the resulting crystals typically do not diffract to atomic resolution ($<1.2 \text{ \AA}$)⁹¹, which is necessary to achieve direct methods solutions. Molecular replacement can be useful but, as with X-ray experiments, molecular replacement cannot solve completely novel and unprecedented structures. As it turns out, the same reason a sample would

benefit from electron crystallography is the same reason why molecular replacement would be a poor phasing method: many of these samples have failed to grow into larger crystals and/or nucleate so rapidly as to prevent persistent crystal growth. Thus, these samples are unlikely to have been solved with X-ray techniques, meaning no search models would be available for usage in molecular replacement experiments.

To increase the utility of electron diffraction, my thesis work focused on two avenues of research. The first goal was to develop electron diffraction phasing methods that would allow us to phase diffraction data unamenable with direct methods or molecular replacement, particularly for macromolecules. Moreover, our secondary objective was to avoid using previously solved structures to confirm these phasing methods. We hoped that validating an electron phasing method and solving novel structures at the same time would best demonstrate the utility of electron diffraction. To improve our chances of achieving our goals, we did not start by experimenting with one of the core phasing techniques discussed previously. Instead, we first investigated racemic crystallography in the context of electron crystallography, which afforded us two major advantages: higher chance of crystallization and the simplification of the phase problem. We believed that these advantages would be key to actualizing the transfer of traditional X-ray phasing techniques to the realm of electron diffraction.

1.7 Racemic Crystallography – Development and Future

Suggested by Alan Mackay in 1989⁹² and first demonstrated by Laura Zawadzke and Jeremy Berg in 1993⁹³, racemic crystallography was developed through the 1990s to obtain crystals large enough for routine X-ray diffraction experiments and to help resolve the phase problem. As predicted by a novel mathematical treatment by Wukovitz and Yeates in 1995⁹⁴, and summarized later by Yeates and Kent in 2012⁹⁵, racemic mixtures present two mathematical advantages: first, racemic mixtures theoretically should crystallize preferentially into centrosymmetric crystal lattices rather than adopt the enantiomorphic lattices observed with

homochiral samples. This conformational preference occurs due to the additional degrees of freedom in which each protein molecule can form protein to protein contacts while still maintaining crystallographic symmetry in certain highly favored centrosymmetric space groups. Utilizing racemic mixtures would enable crystallographers to attempt crystallization in the 165 non-Sohncke space groups that otherwise are inaccessible by homochiral species. Notably, 92 of these space groups are centrosymmetric. Wukovitz and Yeates discovered, mathematically, that one particular centrosymmetric space group, $P1(\bar{1})$ offers more degrees of freedom in crystal formation than any other, leading to the prediction that the frequency of observing crystals in space group $P1(\bar{1})$ alone would exceed that of any other space group currently reported in the PDB. Utilizing this strategy, many scientists have obtained racemic structures that have been previously inaccessible in their homochiral form^{96, 97, 98}. While there have been questions on whether these racemic mixtures provide biologically relevant data, many labs have made efforts to probe the usability of racemic structures to infer information concerning the homochiral structures. So far, the published racemic structures suggest that racemic unit cells composed of separable subunits of each enantiomer that resemble those found in their homochiral analogues. Albeit, this conclusion is drawn from a limited sample size^{99, 100, 101}.

In addition to increasing the number of available crystallographic packing patterns and thus increasing the likelihood of crystal formation, racemic crystallography also aids in phasing⁹⁵, the other major bottleneck of crystallography. In single crystal diffraction experiments, each reflection results from an interference pattern with contributions from every single atom in a unit cell. Since diffraction patterns only capture the squared amplitude of the summed waves (vectors) of a series of constructive and deconstructive interferences, the phases of the summed vectors are lost. In an acentric crystal, these angles can range anywhere between 0° and 360° for every single reflection in the Ewald's sphere. On the other hand, a centric crystal, containing a lattice with a glide plane, is mathematically restricted to only two possible values, separated by 180° or π , for every reflection (typically 0° and 180° are used). This structure massively reduces the

mathematical choices faced by phasing software when considering the possible phase solutions to each reflection during phasing, from any angle between 0° and 360° to just two, 0° or 180° .

We believed that racemic crystallography would not only afford us the best chance of obtaining diffraction-ready samples, but would also reduce the complexity of phasing calculations for our datasets. As previously stated, there are many differences between X-ray and electron diffraction that complicate the transfer of X-ray techniques to MicroED such that it would be difficult to pinpoint the exact reason why a certain method did not work. By reducing the mathematical complexity of the phasing problem, we improve our ability to determine potential reasons why a phasing technique is not working, including: poor dataset completion, inaccurate intensities, and dynamic scattering. We could then address each deficiency separately until a structure is obtained; this procedure can also lead to great insight on which parameters are the most important for each technique. Finally, if no changes yield a structure, we would be more confident that the phasing method would not work with our current technology if even the phase restriction imposed on the dataset does not allow for a structure to be solved. So, while racemic crystallography is not a stand-alone phasing method, it potentially affords such immense and consequential advantages for our goals that we can ill-afford to ignore it. We hope to channel these advantages to not only develop electron phasing methods, but to also study novel systems concurrently during development. Thus, our first experiments were to validate racemic crystallography as a method transferable to electron diffraction experiments.

1.8 References

1. Li, Y.; Yu, J. New Stories of Zeolite Structures: Their Descriptions, Determinations, Predictions, and Evaluations. *Chem. Rev.* **2014**, *114* (14), 7268–7316.
2. Jiang, N.; Shang, R.; Heijman, S. G. J.; Rietveld, L. C. High-Silica Zeolites for Adsorption of Organic Micro-Pollutants in Water Treatment: A Review. *Water Research* **2018**, *144*, 145–161.
3. Rai, R. S.; Subramanian, S. Role of Transmission Electron Microscopy in the Semiconductor Industry for Process Development and Failure Analysis. *Progress in Crystal Growth and Characterization of Materials* **2009**, *55* (3), 63–97.
4. Nakamae, K. Electron Microscopy in Semiconductor Inspection. *Meas. Sci. Technol.* **2021**, *32* (5), 052003.
5. Guha, R. On Exploring Structure Activity Relationships. *Methods Mol Biol* **2013**, *993*, 81–94.
6. Barnes, C. O.; Jette, C. A.; Abernathy, M. E.; Dam, K.-M. A.; Esswein, S. R.; Gristick, H. B.; Malyutin, A. G.; Sharaf, N. G.; Huey-Tubman, K. E.; Lee, Y. E.; Robbiani, D. F.; Nussenzweig, M. C.; West, A. P.; Bjorkman, P. J. SARS-CoV-2 Neutralizing Antibody Structures Inform Therapeutic Strategies. *Nature* **2020**, *588* (7839), 682–687.
7. Crennell, K. *Nobel Prize winners*. <https://www.iucr.org/people/nobel-prize> (accessed 2022-11-29).
8. Boesch, C. Nobel Prizes for Nuclear Magnetic Resonance: 2003 and Historical Perspectives. *Journal of Magnetic Resonance Imaging* **2004**, *20* (2), 177–179.
9. Griffiths, J. A Brief History of Mass Spectrometry. *Anal. Chem.* **2008**, *80* (15), 5678–5683.
10. Cressey, D.; Callaway, E. Cryo-Electron Microscopy Wins Chemistry Nobel. *Nature* **2017**, *550* (7675), 167–167.

11. Jaskolski, M.; Dauter, Z.; Wlodawer, A. A Brief History of Macromolecular Crystallography, Illustrated by a Family Tree and Its Nobel Fruits. *The FEBS Journal* **2014**, *281* (18), 3985–4009.
12. Bodo, G.; Dintzis, H. M.; Kendrew, J. C.; Wyckoff, H. W.; Bragg, W. L. The Crystal Structure of Myoglobin, V. A Low-Resolution Three-Dimensional Fourier Synthesis of Sperm-Whale Myoglobin Crystals. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **1959**, *253* (1272), 70–102.
13. Perutz, M. F.; Rossmann, M. G.; Cullis, A. F.; Muirhead, H.; Will, G.; North, A. C. T. Structure of Hæmoglobin: A Three-Dimensional Fourier Synthesis at 5.5-Å. Resolution, Obtained by X-Ray Analysis. *Nature* **1960**, *185* (4711), 416–422.
14. Kendrew, J. C.; Dickerson, R. E.; Strandberg, B. E.; Hart, R. G.; Davies, D. R.; Phillips, D. C.; Shore, V. C. Structure of Myoglobin: A Three-Dimensional Fourier Synthesis at 2 Å. Resolution. *Nature* **1960**, *185* (4711), 422–427.
15. Perutz, M. F.; Muirhead, H.; Cox, J. M.; Goaman, L. C. G. Three-Dimensional Fourier Synthesis of Horse Oxyhaemoglobin at 2.8 Å Resolution: The Atomic Model. *Nature* **1968**, *219* (5150), 131–139.
16. Yip, K. M.; Fischer, N.; Paknia, E.; Chari, A.; Stark, H. Atomic-Resolution Protein Structure Determination by Cryo-EM. *Nature* **2020**, *587* (7832), 157–161.
17. 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.
18. Zwanzig, R.; Szabo, A.; Bagchi, B. Levinthal's Paradox. *Proceedings of the National Academy of Sciences* **1992**, *89* (1), 20–22.

19. Levinthal, C. How to Fold Graciously. In *Mossbauer Spectroscopy in Biological Systems*, Proceedings of a meeting held at Allerton House, Monticello, Illinois; DeBrunner, J.T.P. and Munck, E., Eds.; University of Illinois Press: Illinois, **1969**; 22-24.
20. Ivankov, D. N.; Finkelstein, A. V. Solution of Levinthal's Paradox and a Physical Theory of Protein Folding Times. *Biomolecules* **2020**, *10* (2), 250.
21. Moult, J.; Pedersen, J. T.; Judson, R.; Fidelis, K. A Large-Scale Experiment to Assess Protein Structure Prediction Methods. *Proteins: Structure, Function, and Bioinformatics* **1995**, *23* (3), ii–iv.
22. Service, R.F. *Google's DeepMind aces protein folding*. <https://www.science.org/content/article/google-s-deepmind-aces-protein-folding> (accessed 2022-11-29).
23. Callaway, E. 'It Will Change Everything': DeepMind's AI Makes Gigantic Leap in Solving Protein Structures. *Nature* **2020**, *588* (7837), 203–204.
24. 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. Protein Structure Prediction Using Multiple Deep Neural Networks in the 13th Critical Assessment of Protein Structure Prediction (CASP13). *Proteins: Structure, Function, and Bioinformatics* **2019**, *87* (12), 1141–1148.
25. 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.
26. Callaway, E. 'It Will Change Everything': DeepMind's AI Makes Gigantic Leap in Solving Protein Structures. *Nature* **2020**, *588* (7837), 203–204.

27. Langer, L. M.; Bonneau, F.; Gat, Y.; Conti, E. Cryo-EM Reconstructions of Inhibitor-Bound SMG1 Kinase Reveal an Autoinhibitory State Dependent on SMG8. *eLife* **2021**, *10*, e72353.
28. Chai, L., Zhu, P., Chai, J., Pang, C., Andi, B., McsWeeney, S., Shanklin, J., Liu, Q. *7SEV: Crystal structure of E coli contaminant protein YadF co-purified with a plant protein*. <https://www.rcsb.org/structure/7SEV> (accessed 2022-11-29).
29. Luciano, A. *AlphaFold-based databases and fully-fledged, easy-to-use, online AlphaFold interfaces poised to revolutionize biology*. <https://towardsdatascience.com/alphafold-based-databases-and-fully-fledged-easy-to-use-alphafold-interfaces-poised-to-baf865c6d75e> (accessed 2022-11-29).
30. Evans, R.; O'Neill, M.; Pritzel, A.; Antropova, N.; Senior, A.; Green, T.; Žídek, A.; Bates, R.; Blackwell, S.; Yim, J.; Ronneberger, O.; Bodenstein, S.; Zielinski, M.; Bridgland, A.; Potapenko, A.; Cowie, A.; Tunyasuvunakool, K.; Jain, R.; Clancy, E.; Kohli, P.; Jumper, J.; Hassabis, D. Protein Complex Prediction with AlphaFold-Multimer. bioRxiv March 10, 2022, p 2021.10.04.463034.
31. Mullard, A. What Does AlphaFold Mean for Drug Discovery? *Nature Reviews Drug Discovery* **2021**, *20* (10), 725–727.
32. Glaeser, R. M.; Hall, R. J. Reaching the Information Limit in Cryo-EM of Biological Macromolecules: Experimental Aspects. *Biophys J* **2011**, *100* (10), 2331–2337.
33. Henderson, R. The Potential and Limitations of Neutrons, Electrons and X-Rays for Atomic Resolution Microscopy of Unstained Biological Molecules. *Quarterly Reviews of Biophysics* **1995**, *28* (2), 171–193.
34. Yeates, T. O.; Agdanowski, M. P.; Liu, Y. Development of Imaging Scaffolds for Cryo-Electron Microscopy. *Current Opinion in Structural Biology* **2020**, *60*, 142–149.
35. Herzik, M. A.; Wu, M.; Lander, G. C. High-Resolution Structure Determination of Sub-100 KDa Complexes Using Conventional Cryo-EM. *Nat Commun* **2019**, *10* (1), 1032.

36. Fan, X.; Wang, J.; Zhang, X.; Yang, Z.; Zhang, J.-C.; Zhao, L.; Peng, H.-L.; Lei, J.; Wang, H.-W. Single Particle Cryo-EM Reconstruction of 52 KDa Streptavidin at 3.2 Angstrom Resolution. *Nat Commun* **2019**, *10* (1), 2386.
37. Holcomb, J.; Spellmon, N.; Zhang, Y.; Doughan, M.; Li, C.; Yang, Z. Protein Crystallization: Eluding the Bottleneck of X-Ray Crystallography. *AIMS Biophys* **2017**, *4* (4), 557–575.
38. Cowtan, K. Phase Problem in X-Ray Crystallography, and Its Solution. In *eLS*; John Wiley & Sons, Ltd, 2003.
39. Burley, S. K.; Joachimiak, A.; Montelione, G. T.; Wilson, I. A. Contributions to the NIH-NIGMS Protein Structure Initiative from the PSI Production Centers. *Structure* **2008**, *16* (1), 5–11.
40. Kirkwood, J.; Hargreaves, D.; O’Keefe, S.; Wilson, J. Analysis of Crystallization Data in the Protein Data Bank. *Acta Cryst F* **2015**, *71* (10), 1228–1234.
41. X-Ray Mass Attenuation Coefficients. *NIST* **2009**.
42. Duvauchelle, P.; Peix, G.; Babot, D. Effective Atomic Number in the Rayleigh to Compton Scattering Ratio. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* **1999**, *155* (3), 221–228.
43. Garman, E. F.; Weik, M. Radiation Damage in Macromolecular Crystallography. In *Protein Crystallography: Methods and Protocols*; Wlodawer, A., Dauter, Z., Jaskolski, M., Eds.; Methods in Molecular Biology; Springer: New York, NY, 2017; pp 467–489.
44. Holton, J. M. A Beginner’s Guide to Radiation Damage. *J Synchrotron Rad* **2009**, *16* (2), 133–142.
45. Smith, J. L.; Fischetti, R. F.; Yamamoto, M. Micro-Crystallography Comes of Age. *Current Opinion in Structural Biology* **2012**, *22* (5), 602–612.
46. Sanishvili, R.; Nagarajan, V.; Yoder, D.; Becker, M.; Xu, S.; Corcoran, S.; Akey, D. L.; Smith, J. L.; Fischetti, R. F. A 7 Mm Mini-Beam Improves Diffraction Data from Small or Imperfect Crystals of Macromolecules. *Acta Cryst D* **2008**, *64* (4), 425–435.

47. Holton, J. M.; Frankel, K. A. The Minimum Crystal Size Needed for a Complete Diffraction Data Set. *Acta Cryst D* **2010**, *66* (4), 393–408.
48. Shi, D.; Nannenga, B. L.; Iadanza, M. G.; Gonen, T. Three-Dimensional Electron Crystallography of Protein Microcrystals. *eLife* **2013**, *2*, e01345.
49. Schneider, D. K.; Shi, W.; Andi, B.; Jakoncic, J.; Gao, Y.; Bhogadi, D. K.; Myers, S. F.; Martins, B.; Skinner, J. M.; Aishima, J.; Qian, K.; Bernstein, H. J.; Lazo, E. O.; Langdon, T.; Lara, J.; Shea-McCarthy, G.; Idir, M.; Huang, L.; Chubar, O.; Sweet, R. M.; Berman, L. E.; McSweeney, S.; Fuchs, M. R. FMX - the Frontier Microfocusing Macromolecular Crystallography Beamline at the National Synchrotron Light Source II. *J Synchrotron Radiat* **2021**, *28* (Pt 2), 650–665.
50. Liu, H.; Lee, W. The XFEL Protein Crystallography: Developments and Perspectives. *International Journal of Molecular Sciences* **2019**, *20* (14), 3421.
51. Barends, T. R. M.; Foucar, L.; Botha, S.; Doak, R. B.; Shoeman, R. L.; Nass, K.; Koglin, J. E.; Williams, G. J.; Boutet, S.; Messerschmidt, M.; Schlichting, I. De Novo Protein Crystal Structure Determination from X-Ray Free-Electron Laser Data. *Nature* **2014**, *505* (7482), 244–247.
52. Mishin, A.; Gusach, A.; Luginina, A.; Marin, E.; Borshchevskiy, V.; Cherezov, V. An Outlook on Using Serial Femtosecond Crystallography in Drug Discovery. *Expert Opinion on Drug Discovery* **2019**, *14* (9), 933–945.
53. Loudon, M. and Parise, J. *Organic Chemistry*, 5th ed.; Roberts and Company Publishers, 2009.
54. Nakane, T.; Kotecha, A.; Sente, A.; McMullan, G.; Masiulis, S.; Brown, P. M. G. E.; Grigoras, I. T.; Malinauskaite, L.; Malinauskas, T.; Miehl, 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.

55. Yip, K. M.; Fischer, N.; Paknia, E.; Chari, A.; Stark, H. Atomic-Resolution Protein Structure Determination by Cryo-EM. *Nature* **2020**, *587* (7832), 157–161.
56. Taylor, G. The Phase Problem. *Acta Cryst D* **2003**, *59* (11), 1881–1890.
57. Taylor, G. L. Introduction to Phasing. *Acta Cryst D* **2010**, *66* (4), 325–338.
58. Cochran, W. A Relation between the Signs of Structure Factors. *Acta Cryst* **1952**, *5* (1), 65–67.
59. Sayre, D. The Squaring Method: A New Method for Phase Determination. *Acta Cryst* **1952**, *5* (1), 60–65.
60. Zachariasen, W. H. A New Analytical Method for Solving Complex Crystal Structures. *Acta Cryst* **1952**, *5* (1), 68–73.
61. Karle, J.; Hauptman, H. A Theory of Phase Determination for the Four Types of Non-Centrosymmetric Space Groups 1P222, 2P22, 3P12, 3P22. *Acta Cryst* **1956**, *9* (8), 635–651.
62. Karle, I. L.; Karle, J. The Crystal and Molecular Structure of the Alkaloid Jamine, C₂₁H₃₅N₃, from *Ormosia Jamaicensis*. *Acta Cryst* **1964**, *17* (11), 1356–1360.
63. Usón, I.; Sheldrick, G. M. Advances in Direct Methods for Protein Crystallography. *Current Opinion in Structural Biology* **1999**, *9* (5), 643–648.
64. Perutz, M. F. Isomorphous Replacement and Phase Determination in Non-Centrosymmetric Space Groups. *Acta Cryst* **1956**, *9* (11), 867–873.
65. Yamashita, K.; Pan, D.; Okuda, T.; Sugahara, M.; Kodan, A.; Yamaguchi, T.; Murai, T.; Gomi, K.; Kajiyama, N.; Mizohata, E.; Suzuki, M.; Nango, E.; Tono, K.; Joti, Y.; Kameshima, T.; Park, J.; Song, C.; Hatsui, T.; Yabashi, M.; Iwata, S.; Kato, H.; Ago, H.; Yamamoto, M.; Nakatsu, T. An Isomorphous Replacement Method for Efficient de Novo Phasing for Serial Femtosecond Crystallography. *Sci Rep* **2015**, *5* (1), 14017.
66. Hanashima, S.; Nakane, T.; Mizohata, E. Heavy Atom Detergent/Lipid Combined X-Ray Crystallography for Elucidating the Structure-Function Relationships of Membrane Proteins. *Membranes* **2021**, *11* (11), 823.

67. Pike, A. C. W.; Garman, E. F.; Krojer, T.; von Delft, F.; Carpenter, E. P. An Overview of Heavy-Atom Derivatization of Protein Crystals. *Acta Cryst D* **2016**, *72* (3), 303–318.
68. Hendrickson, W. A. Determination of Macromolecular Structures from Anomalous Diffraction of Synchrotron Radiation. *Science* **1991**, *254* (5028), 51–58.
69. Liu, Q.; Hendrickson, W. A. Contemporary Use of Anomalous Diffraction in Biomolecular Structure Analysis. In *Protein Crystallography: Methods and Protocols*; Wlodawer, A., Dauter, Z., Jaskolski, M., Eds.; Methods in Molecular Biology; Springer: New York, NY, 2017; pp 377–399.
70. Huber, R. Die automatisierte Faltmolekülmethode. *Acta Cryst* **1965**, *19* (3), 353–356.
71. Evans, P.; McCoy, A. An Introduction to Molecular Replacement. *Acta Cryst D* **2008**, *64* (1), 1–10.
72. 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.
73. McCoy, A. J.; Oeffner, R. D.; Wrobel, A. G.; Ojala, J. R. M.; Tryggvason, K.; Lohkamp, B.; Read, R. J. Ab Initio Solution of Macromolecular Crystal Structures without Direct Methods. *Proceedings of the National Academy of Sciences* **2017**, *114* (14), 3637–3641.
74. 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. bioRxiv September 14, 2021, p 2021.09.13.459692.
75. Shi, D.; Nannenga, B. L.; Iadanza, M. G.; Gonen, T. Three-Dimensional Electron Crystallography of Protein Microcrystals. *eLife* **2013**, *2*, e01345.
76. Mu, X.; Gillman, C.; Nguyen, C.; Gonen, T. An Overview of Microcrystal Electron Diffraction (MicroED). *Annual Review of Biochemistry* **2021**, *90* (1), 431–450.
77. Glaeser, R. M. Electron Crystallography of Biological Macromolecules. *Annual Review of Physical Chemistry* **1985**, *36* (1), 243–275.

78. Henderson, R. From Electron Crystallography to Single Particle CryoEM (Nobel Lecture). *Angewandte Chemie International Edition* **2018**, *57* (34), 10804–10825.
79. Fujiyoshi, Y.; Unwin, N. Electron Crystallography of Proteins in Membranes. *Current Opinion in Structural Biology* **2008**, *18* (5), 587–592.
80. De Rosier, D. J.; Klug, A. Reconstruction of Three Dimensional Structures from Electron Micrographs. *Nature* **1968**, *217* (5124), 130–134.
81. Zatsepin, N. A.; Li, C.; Colasurd, P.; Nannenga, B. L. The Complementarity of Serial Femtosecond Crystallography and MicroED for Structure Determination from Microcrystals. *Current Opinion in Structural Biology* **2019**, *58*, 286–293.
82. Nannenga, B. L.; Shi, D.; Hattne, J.; Reyes, F. E.; Gonen, T. Structure of Catalase Determined by MicroED. *eLife* **2014**, *3*, e03600.
83. Dorset, D. L.; Tivol, W. F.; Turner, J. N. Dynamical Scattering and Electron Crystallography - Ab Initio Structure Analysis of Copper Perbromophthalocyanine. *Acta Cryst A* **1992**, *48* (4), 562–568.
84. Blum, T. B.; Housset, D.; Clabbers, M. T. B.; van Genderen, E.; Bacia-Verloop, M.; Zander, U.; McCarthy, A. A.; Schoehn, G.; Ling, W. L.; Abrahams, J. P. Statistically Correcting Dynamical Electron Scattering Improves the Refinement of Protein Nanocrystals, Including Charge Refinement of Coordinated Metals. *Acta Cryst D* **2021**, *77* (1), 75–85.
85. Peng, L.-M. Electron Atomic Scattering Factors and Scattering Potentials of Crystals. *Micron* **1999**, *30* (6), 625–648.
86. Yonekura, K.; Matsuoka, R.; Yamashita, Y.; Yamane, T.; Ikeguchi, M.; Kidera, A.; Maki-Yonekura, S. Ionic Scattering Factors of Atoms That Compose Biological Molecules. *IUCrJ* **2018**, *5* (3), 348–353.
87. Palatinus, L.; Brázda, P.; Boullay, P.; Perez, O.; Klementová, M.; Petit, S.; Eigner, V.; Zaarour, M.; Mintova, S. Hydrogen Positions in Single Nanocrystals Revealed by Electron Diffraction. *Science* **2017**, *355* (6321), 166–169.

88. 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. *Proceedings of the National Academy of Sciences* **2016**, *113* (40), 11232–11236.
89. 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.
90. Wang, H.-W.; Wang, J.-W. How Cryo-Electron Microscopy and X-Ray Crystallography Complement Each Other. *Protein Science* **2017**, *26* (1), 32–39.
91. Theis, K., Namba, K., and Martz, E. *Resolution - Proteopedia, life in 3D*. https://proteopedia.org/wiki/index.php/Resolution#cite_note-atomsiz-3 (accessed 2022-11-29).
92. Mackay, A. L. Crystal Enigma. *Nature* **1989**, *342* (6246), 133–133.
93. Zawadzke, L. E.; Berg, J. M. A Racemic Protein. *J. Am. Chem. Soc.* **1992**, *114* (10), 4002–4003.
94. Wukovitz, S. W.; Yeates, T. O. Why Protein Crystals Favour Some Space-Groups over Others. *Nat Struct Mol Biol* **1995**, *2* (12), 1062–1067.
95. Yeates, T. O.; Kent, S. B. H. Racemic Protein Crystallography. *Annual Review of Biophysics* **2012**, *41* (1), 41–61.
96. Pentelute, B. L.; Gates, Z. P.; Tereshko, V.; Dashnau, J. L.; Vanderkooi, J. M.; Kossiakoff, A. A.; Kent, S. B. H. X-Ray Structure of Snow Flea Antifreeze Protein Determined by Racemic Crystallization of Synthetic Protein Enantiomers. *J. Am. Chem. Soc.* **2008**, *130* (30), 9695–9701.

97. Pentelute, B. L.; Mandal, K.; Gates, Z. P.; Sawaya, M. R.; Yeates, T. O.; Kent, S. B. H. Total Chemical Synthesis and X-Ray Structure of Kaliotoxin by Racemic Protein Crystallography. *Chem. Commun.* **2009**, *46* (43), 8174–8176.
98. Banigan, J. R.; Mandal, K.; Sawaya, M. R.; Thammavongsa, V.; Hendrickx, A. P. A.; Schneewind, O.; Yeates, T. O.; Kent, S. B. H. Determination of the X-Ray Structure of the Snake Venom Protein Omwaprin by Total Chemical Synthesis and Racemic Protein Crystallography. *Protein Sci* **2010**, *19* (10), 1840–1849.
99. Kurgan, K. W.; Kleman, A. F.; Bingman, C. A.; Kreitler, D. F.; Weisblum, B.; Forest, K. T.; Gellman, S. H. Retention of Native Quaternary Structure in Racemic Melittin Crystals. *J. Am. Chem. Soc.* **2019**, *141* (19), 7704–7708.
100. Bunker, R. D.; Mandal, K.; Bashiri, G.; Chaston, J. J.; Pentelute, B. L.; Lott, J. S.; Kent, S. B. H.; Baker, E. N. A Functional Role of Rv1738 in Mycobacterium Tuberculosis Persistence Suggested by Racemic Protein Crystallography. *Proceedings of the National Academy of Sciences* **2015**, *112* (14), 4310–4315.
101. Mandal, P. K.; Collie, G. W.; Srivastava, S. C.; Kauffmann, B.; Huc, I. Structure Elucidation of the Pribnow Box Consensus Promoter Sequence by Racemic DNA Crystallography. *Nucleic Acids Research* **2016**, *44* (12), 5936–5943.

CHAPTER 2

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

The following is a reprint of a research article from

International Union of Crystallography Journal

6(2), 197-205 (2019)

DOI: 10.1107/S2052252518017621

Reprinted by permission from IUCr Journals.

This is an open access article under CC-BY License

(<https://creativecommons.org/licenses/by/4.0/>)



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

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

Received 28 August 2018
Accepted 12 December 2018

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

Keywords: amyloid; racemic; electron diffraction; ice nucleation; intermolecular interactions; co-crystals; electron crystallography; structural biology.

PDB references: racemic GSTSTA, X-ray structure, 6m7m; MicroED structure, 6m9j; L-GSTSTA, X-ray structure, 6eex; MicroED structure, 6m9i

Supporting information: this article has supporting information at www.iucr.org

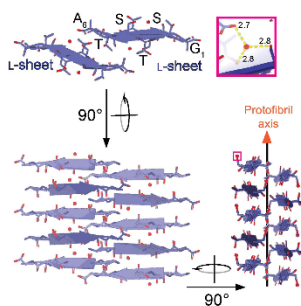
^aDepartment of Chemistry and Biochemistry, UCLA-DOE Institute for Genomics and Proteomics, University of California Los Angeles, Los Angeles, CA 90095, USA, ^bDepartment of Biological Chemistry, UCLA-DOE Institute for Genomics and Proteomics, University of California Los Angeles, Los Angeles, CA 90095,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

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

1. Introduction

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

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



OPEN ACCESS

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

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

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

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

2. Methods

2.1. Sequence analysis of ice-nucleation proteins

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

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

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

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

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

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

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

2.3. Microfocus X-ray data collection and structure determination

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

Diffraction data were collected from both homochiral and racemic GSTSTA crystals under cryogenic conditions (100 K) on beamline 24-ID-E at the Advanced Photon Source (APS) equipped with an ADSC Q315 CCD detector using a 5 µm beam with wavelength 0.979 Å. Signal was only limited by the edge of our detector at approximately 1.1 Å; as such, higher resolution data could perhaps be achieved by modifying the experimental geometry and/or adjusting the energy of the X-ray beam in the experiment. Data from both homochiral and racemic crystals were reduced in *XDS* (Kabsch, 2010) and yielded *ab initio* solutions using *SHELXT* and *SHELXD* (Sheldrick, 2015). The phases obtained from these coordinates produced maps of sufficient quality for subsequent model building in *Coot* (Emsley *et al.*, 2010). The resulting models were refined against the measured data using *PHENIX* (Adams *et al.*, 2010).

2.4. Electron microscopy, MicroED data collection and structure determination

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

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

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

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

2.5. Analysis of homochiral and racemic GSTSTA structures

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

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

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

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

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

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

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

3. Results

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

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

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

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

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

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

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

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

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

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

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

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

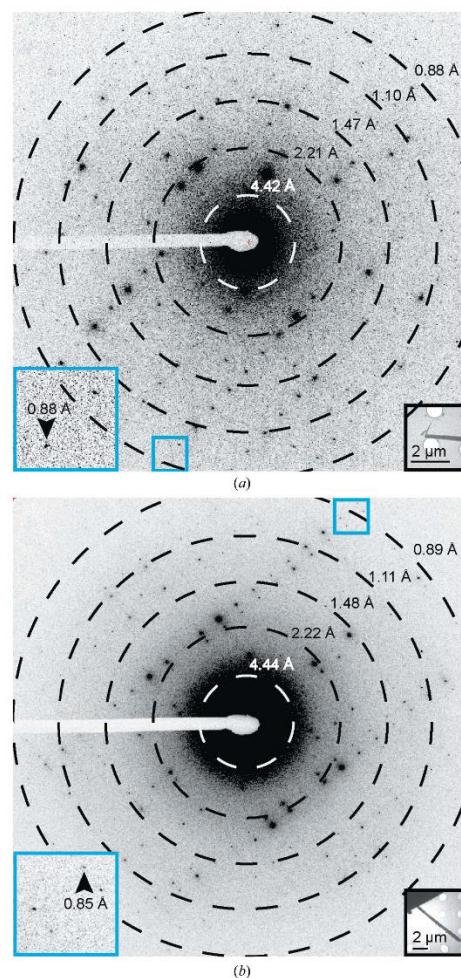


Figure 1
Single diffraction patterns of homochiral L-GSTSTA (*a*) and racemic GSTSTA (*b*) measured during continuous-rotation MicroED data collection. Each pattern corresponds to a 0.6° wedge (*a*) or a 0.9° wedge (*b*) of reciprocal space. Black insets show overfocused diffraction images of the crystals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

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

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

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

3.5. Phase restriction in centrosymmetric crystals evaluated by MicroED

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

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

3.6. Structure of L-GSTSTA

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

3.7. Structure of racemic GSTSTA

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

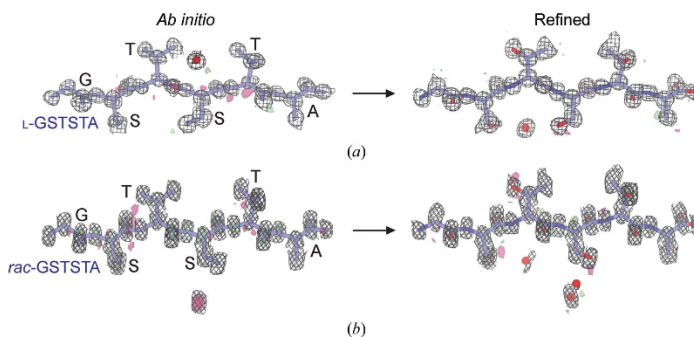


Figure 2 *Ab initio* structures and electrostatic potential maps of L-GSTSTA (a) and racemic GSTSTA (b). Each map in (a) is overlaid onto the initial atomic coordinates calculated by *SHELXD* from MicroED data. Each map in (b) is overlaid onto its corresponding refined model. The $2F_o - F_c$ map represented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

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

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

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

4. Discussion

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

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

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

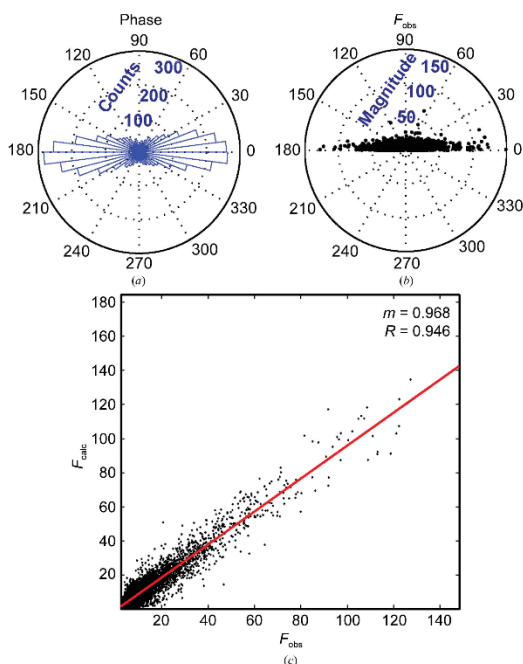


Figure 3
(a) The calculated phase associated with each reflection in the *P1* refinement of racemic GSTSTA data obtained by MicroED was analyzed and plotted as a histogram along the unit circle. (b) The magnitude of each reflection is plotted as a function of the absolute value of its associated phase.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

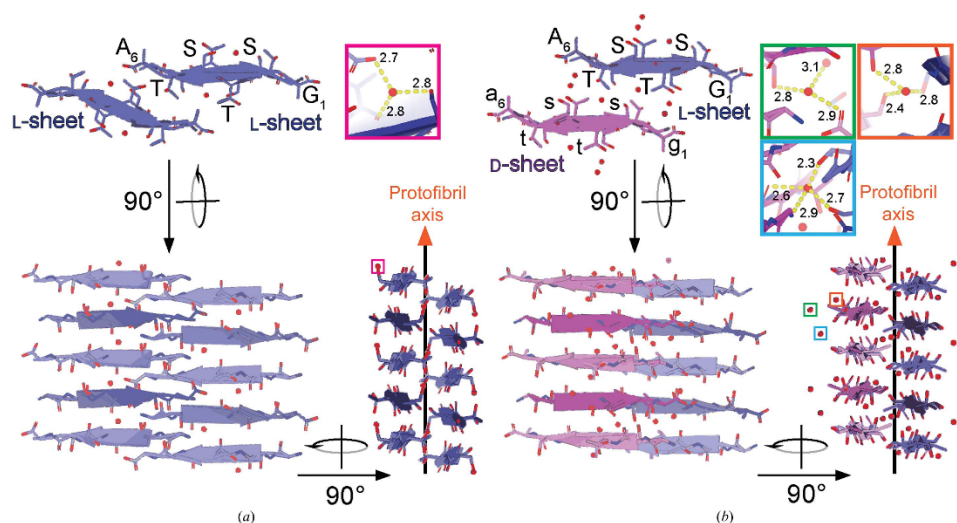


Figure 4

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

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

Acknowledgements

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

Funding information

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

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

References

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

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

IUCrJ

Volume 6 (2019)

Supporting information for article:

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

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

Table S1 Crystallographic data collection and refinement statistics.

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

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

Rms deviations

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

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

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

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

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

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

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

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

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

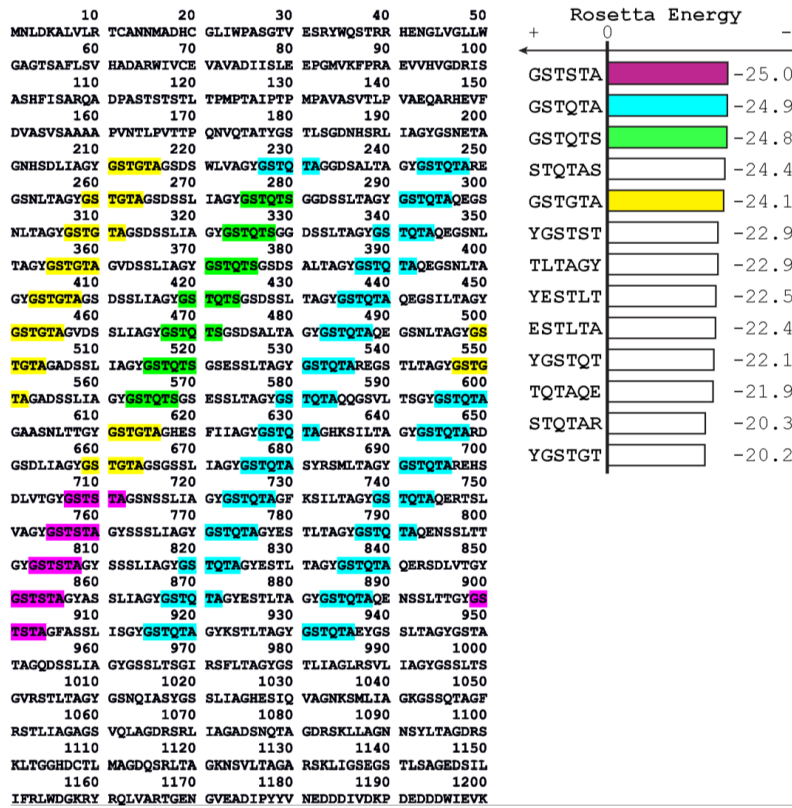


Figure S1 Sequence of the ice nucleation protein InaZ is shown with its degenerate hexameric repeats highlighted as follows: GSTGTA (yellow), GSTQTA (cyan), GSTSTA (magenta), and GSTQTS (green). The propensity for the hexamers to form steric zippers is shown on the left as Rosetta energy scores, determined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

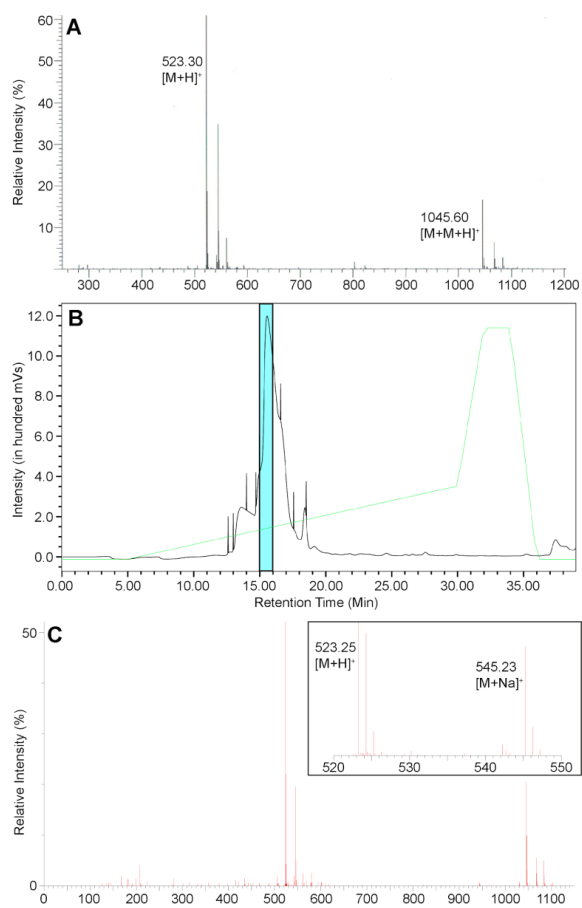


Figure S2 The mass trace of L-GSTSTA (A), purchased from Genscript, shows a [M+H]⁺ peak of 523.30 g/mol (expected 523.22) and a dimer [M+M+H]⁺ peak of 1045.6 g/mol (expected 1045.44). D-GSTSTA was synthesized and purified in-house by reverse-phase HPLC (B). The cyan shaded region highlights the HPLC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

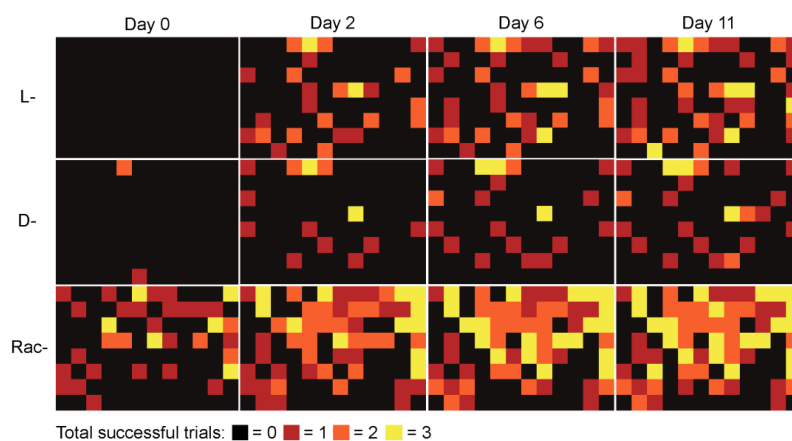


Figure S3 The number of successful crystallization trials in the 96-well crystal screens for L-GSTSTA, D-GSTSTA, and racemic GSTSTA were monitored and converted to heatmaps. Each well contained three hanging drops, each with different protein to buffer ratios. A count of crystals found in each of these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

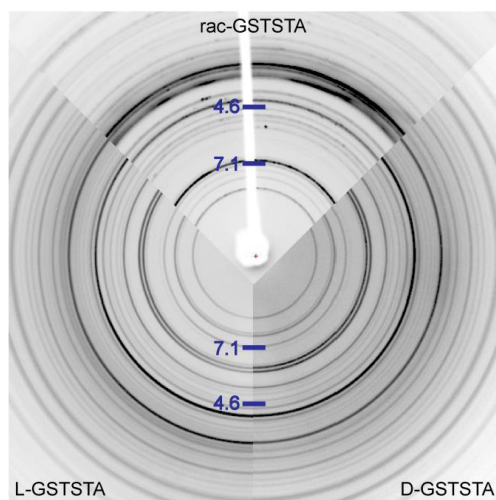


Figure S4 Comparison of powder diffraction patterns measured from L-GSTSTA (left), D-GSTSTA (right), and racemic GSTSTA (top) slurries show differences across all resolutions. Both patterns show faint rings at approximately 4.6 Å, representing the approximate distance between strands along the fibril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

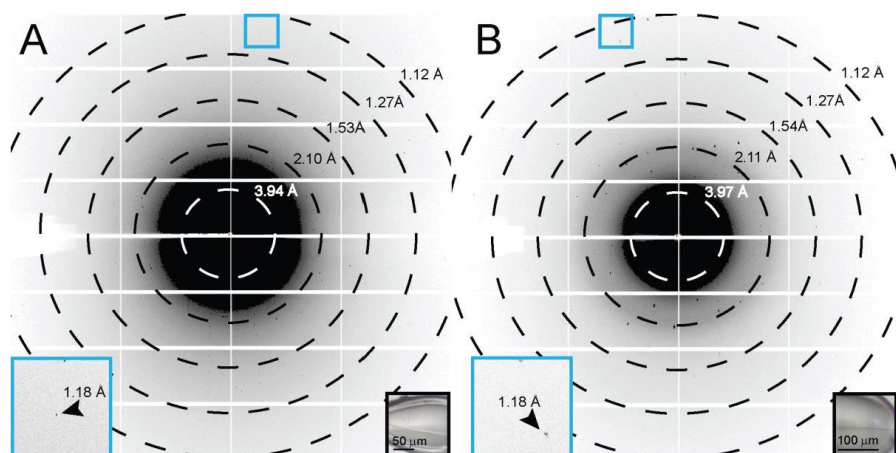


Figure S5 Single diffraction patterns of homochiral L-GSTSTA (A) or racemic GSTSTA (B) measured during continuous vector scanning, microfocus x-ray data collection. Each pattern corresponds to a 5° wedge (A) or 2.5° wedge (B) of reciprocal space. Black insets show in-line images of the crystals that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

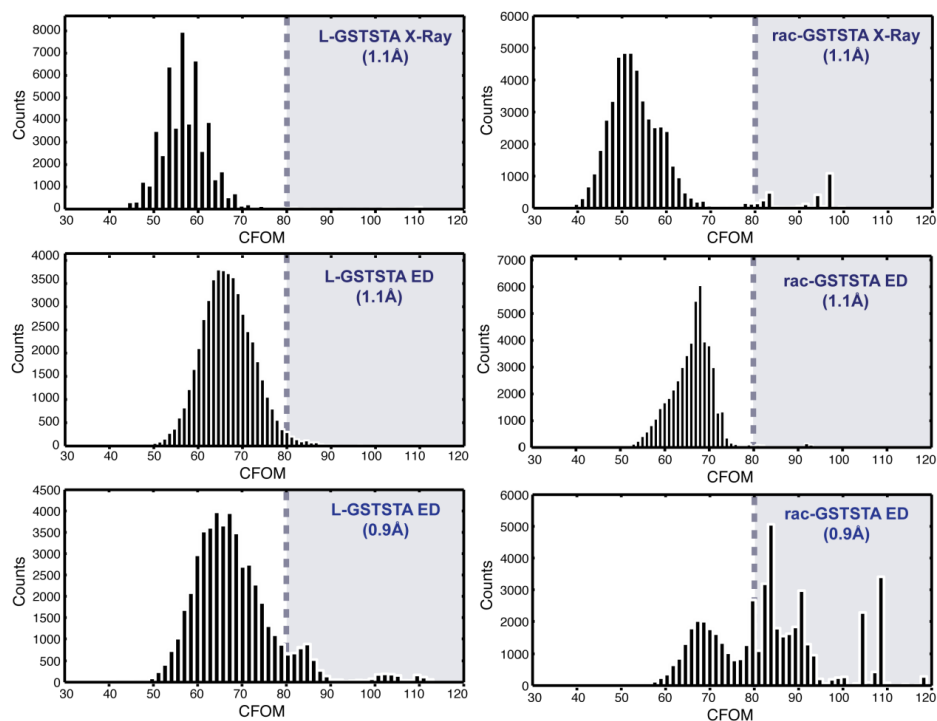


Figure S6 Histograms of the combined figure of merit (CFOM) scores for 50,000 trials by SHELXD indicate the approximate frequency of correct solutions. The shaded region in each plot, where CFOM scores are greater than 80, represents an area in which solutions have a high probability of being correct.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

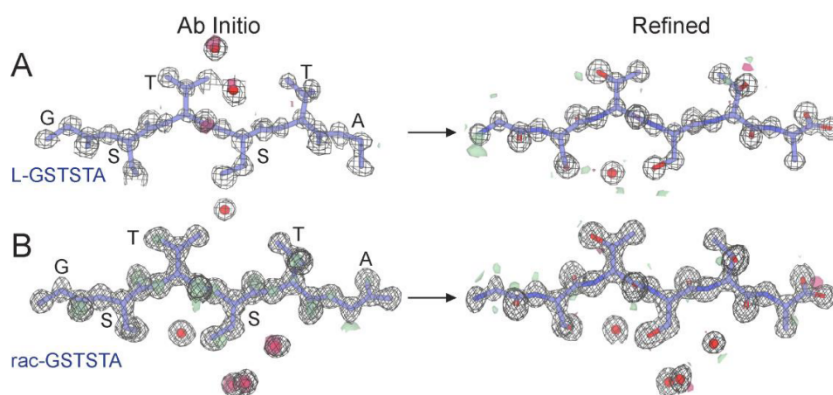


Figure S7 *Ab Initio* structures and electron density maps of L-GSTSTA (A) or racemic GSTSTA (B). Each map in A is overlaid onto the initial atomic coordinates calculated by SHELXD from x-ray diffraction data. Each map in B is overlaid onto its corresponding refined model. The $2F_o - F_c$ map represented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

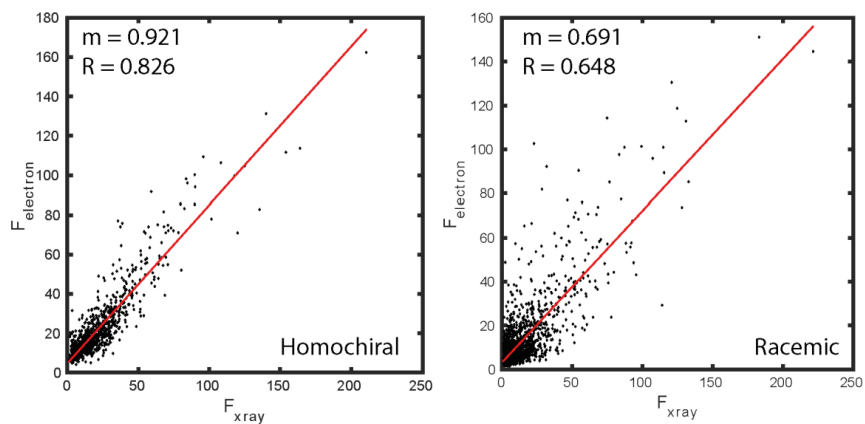


Figure S8 A plot of magnitudes (F) compared between reflections in datasets collected from homochiral crystals (left) or racemic crystals (right) by either electron or x-ray scattering shows a distribution that can be fit by linear regression, indicated by red lines with slope (m) and R-value.

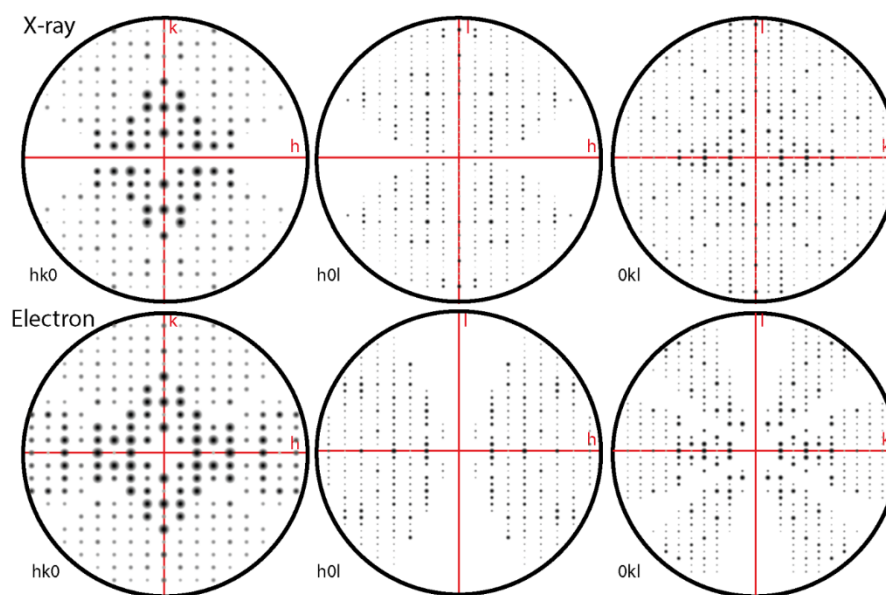


Figure S9 HKL Zone analysis for crystals of L-GTSTA. Fourier magnitudes are shown as displayed by the HKL view software for reflections along principal zones of the reciprocal lattice. Zones where $l=0$ (left), $k=0$ (middle), $h=0$ (right) are shown for merged data collected by x-ray (top) and electron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

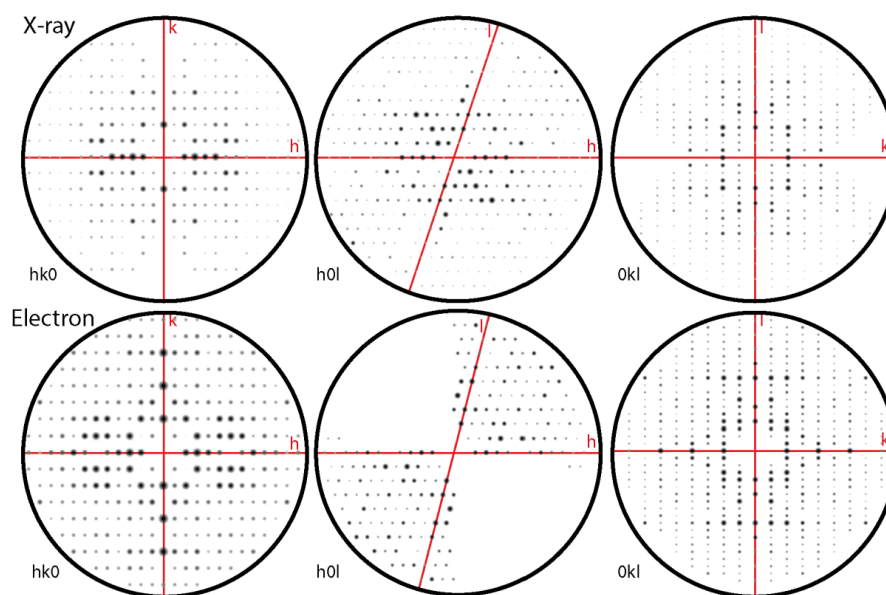


Figure S10 HKL Zone analysis for crystals of racemic GSTSTA. Fourier magnitudes are shown as displayed by the HKL view software for reflections along principal zones of the reciprocal lattice. Zones where $l=0$ (left), $k=0$ (middle), $h=0$ (right) are shown for merged data collected by x-ray (top) and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

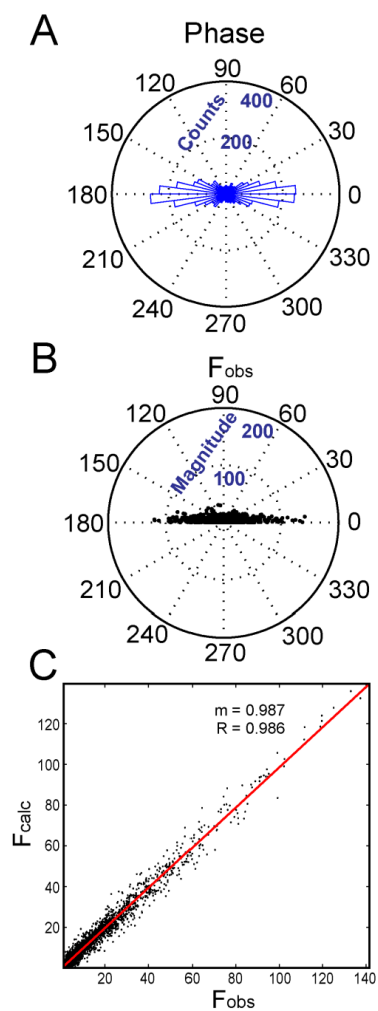


Figure S11 (A) The calculated phase associated with each reflection in the P1 refinement of racemic GSTSTA data obtained by x-ray diffraction was analyzed and plotted as a histogram along the unit circle. (B) The magnitude of each reflection is plotted as a function of the absolute value of its associated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

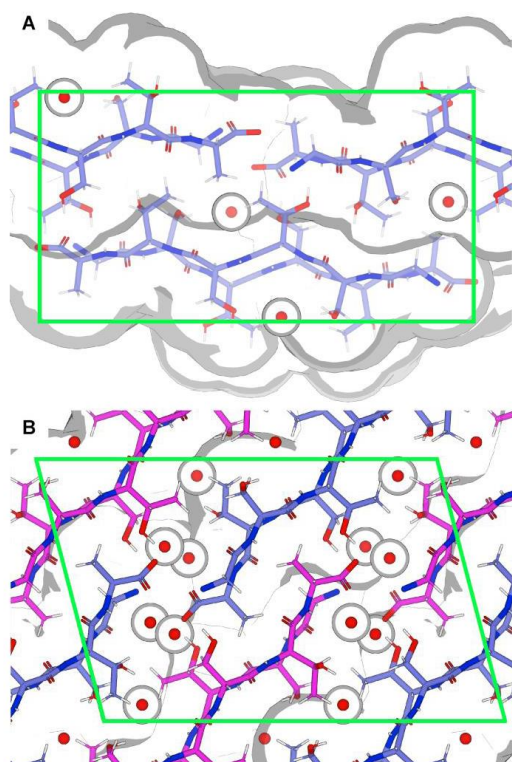


Figure S12 Views down the a-axis of L-GSTSTA (A) and racemic GSTSTA (B) structures are enclosed in green by their respective unit cells. L-GSTSTA strands are colored blue while D-GSTSTA strands are colored magenta. The volumes occupied by each structure are shown in white with edges defined by the shaded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

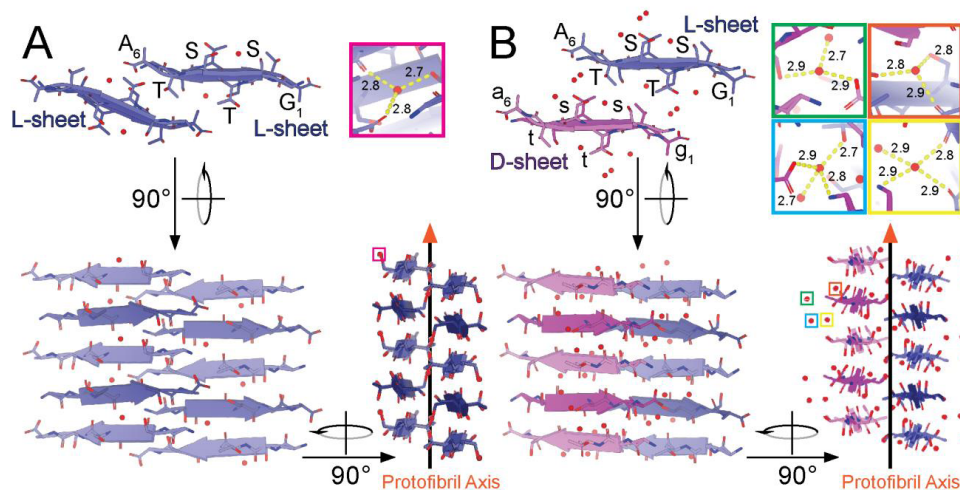


Figure S13 Views of protofibrils of L-GSTSTA (A) and racemic GSTSTA (B) represented by pair of sheets with a view down the protofibril axis; both structures derived by x-ray diffraction. A 90° rotation shows a side view of the protofibril with strands stacked along each sheet in an antiparallel fashion.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

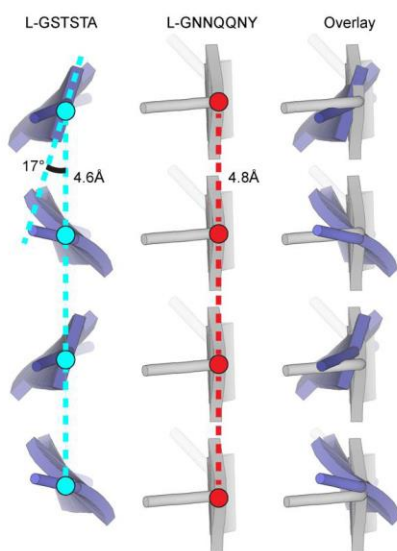


Figure S14 Beta sheets in the structure of L-GSTSTA have an inter-strand distance of 4.6Å with amide carbonyls angled away from the protofibril axis by approximately 17° (left). Canonical beta sheets formed by the yeast prion segment L-GNNQNNY (RCSB PDB: 1YJP) show inter-strand distance of 4.8Å and a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

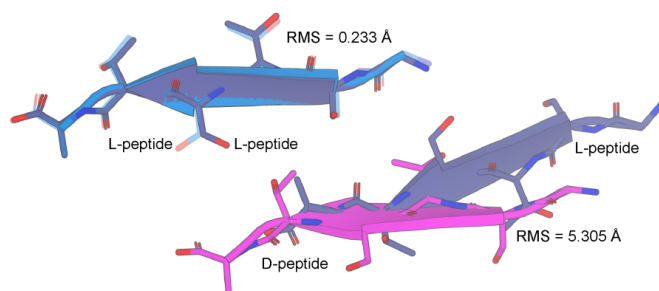


Figure S15 Pairs of mated strands representing the homochiral and racemic protofibrils of GSTSTA. Alignment of these protofibrils based on a common L-GSTSTA sheet shows a displacement of their paired sheet. The RMSD between the common L-GSTSTA sheets is 0.23 Å, while that between the pairing L and D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

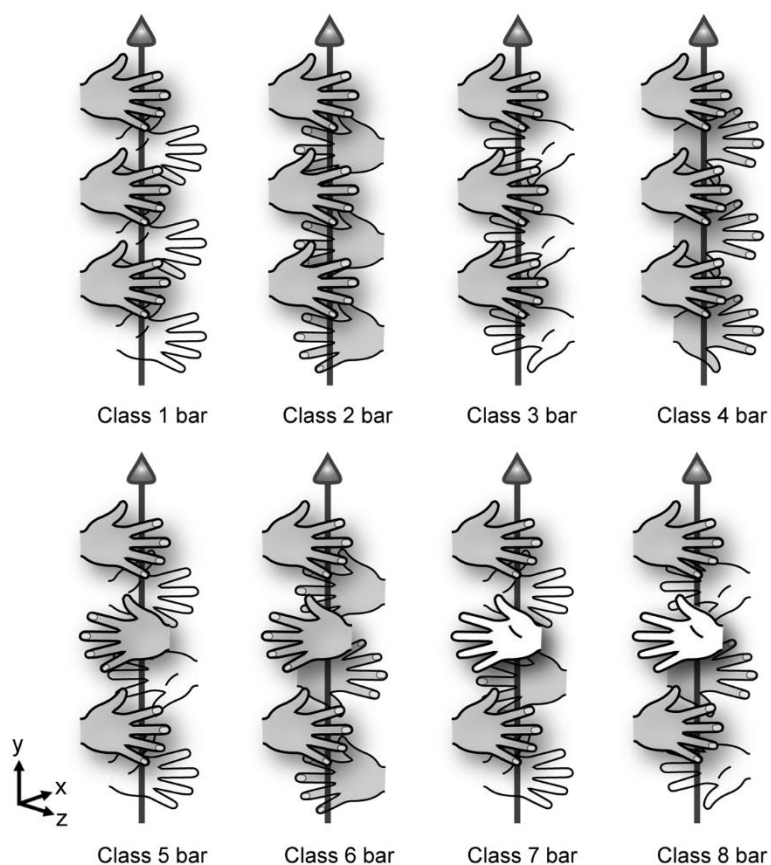


Figure S16 Eight new potential steric zipper classes are enabled by racemic assemblies. All new classes are based on those originally described in Sawaya *et al.* (Sawaya *et al.*, 2007) but now contain a mirror plane, a glide plane, or an inversion center. Strands are represented by left and right hands,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

CHAPTER 3

***Ab initio* MicroED structure of racemic macrolittin 70**

The following is a manuscript in preparation for submission.

Ab Initio MicroED structure of racemic macrolittin 70

Chih-Te Zee^{1*}, Kathleen W. Kurgan^{2*}, Michael R. Sawaya¹, Isabel Usón^{3#}, Sam H. Gellman^{2#}, Jose A. Rodriguez^{1#}

*Denotes equal contribution as co-first author

#Correspondence: ufcri@ibmb.csic.es, rodriguez@chem.ucla.edu, gellman@chem.wisc.edu

Abstract

The routine structural characterization of proteins and peptides using MicroED would be a great advance in the biophysical field. Toward this end, we hope to develop methods for structural determination using MicroED that can be employed with a standard instrumental set-up. Here we determine the structure of a 26-residue pore-forming peptide, macrolittin 70, using racemic crystallization, MicroED, and fragment-based phasing methods.

Introduction

MicroED is a powerful tool whose potential for routinely acquiring high-resolution structures of proteins and peptides from imperfect crystals, e.g., mosaic or micro crystals, has yet to be fully realized. Recent advances¹⁻³, notably work from Gonen, T. and co-workers⁴, highlight this potential. However, the protocol implemented in the aforementioned work required advanced sample preparation, data collection, and data analysis; more advanced than would be expected of a general user. Here we combine racemic crystallization and fragment-based phasing to determine the structure of a 26-residue peptide, macrolittin 70. The structure of this peptide is novel and to our knowledge, the longest *ab initio* and novel peptide or protein structure to be determined using MicroED methods.

Macrolittin 70 is a peptide developed by the Hristova and Wimley labs who aimed to design peptides with high pore-forming activity. The sequence of macrolittin 70 is a variant of the canonical example of a cytotoxic peptide, melittin. Starting from the melittin sequence, the Wimley and Hristova labs have discovered a number of peptides that induce pore-formation in neutral

conditions and in a pH-dependent manner using “synthetic molecular evolution⁵⁻⁸.” The synthetic molecular evolution process consists of iterative combinatorial library design and high-throughput screening to identify “gain-of-function” variants. Wimley, W. C. and co-workers first used this approach to assess the pore-forming activity of a 7776-member rational combinatorial library of melittin variants where 10 of the 26 residues were randomized⁸. Employment of the “split and combine” method⁹ allowed for the randomization of amino acid composition at defined positions in the peptide sequence. This study led to the identification of MelP5, a peptide that is able to induce the leakage of small molecules at extremely low peptide:lipid ratios (~1:1000) in vesicles with various lipid compositions where melittin is inactive.

Macrolittins are a class of the peptides that are able to induce small molecule and macromolecule leakage at neutral pH⁸. The macrolittin sequences are derived from the Mel P5 peptides with 9 randomized positions. This library was screened for small molecule and macromolecular leakage in POPC vesicles. Small molecule leakage was assessed by incubating the peptides in the presence of vesicles containing encapsulated ANTS, a small molecule fluorophore, and DPX, an ANTS quencher. Leakage is measured by an increase in ANTS fluorescence as a result of the dilution of DPX which occurs when it escapes from the vesicle. The peptides were also assessed for leakage of the large molecule TAMRA-biotin-dextran 40 kDa (TBD) in the presence of fluorophore-labelled streptavidin. Leakage was quantified by measuring FRET between the TBD and fluorophore-labelled streptavidin. Macrolittins are able to achieve 100% leakage of TBDs at peptide to lipid ratios as low as 1:5000⁷. One of the most highly active peptides from this library is macrolittin 70.

Previously, we were able to elucidate the structure of melittin by co-crystallizing the racemic variants of melittin, L- and D-melittin¹⁰, a technique called racemic crystallography¹¹. As macrolittin 70 and melittin share many chemical similarities, we set about using racemic crystallography to determine the structure of macrolittin 70 in order to potentially identify structural differences between the two peptides that could account for the difference in pore-forming activity.

Many crystals of the 1:1 mixture of L- and D-macrolittin 70 were obtained but we observed high lattice disorder during X-ray diffraction data collection that precluded structural elucidation. We turned to MicroED and were able to successfully determine the structure of macrolittin 70 using fragment-based phasing and molecular replacement methods. The structural determination of macrolittin 70 has presented a great case-study for: elucidating novel structures using MicroED, allowing the comparison of fragment-based phasing versus molecular replacement with models generated with AlphaFold, and identifying crystal qualities that may lead to higher versus lower crystal stability.

Results

Attempts with X-ray Crystallography and Batch Crystallization

Racemic crystallization has been harnessed to acquire the structures of systems that are recalcitrant to homomeric crystallization. It has been applied to antimicrobial and cytotoxic peptides including: baboon Θ -defensin-2 (BTD-2),¹² derivatives of the host-defense peptide, magainin 2,^{13,14} and melittin.¹⁰ Our report of the structure of racemic melittin demonstrated that the co-incubation of the peptide racemates did not disrupt the homomeric association of the peptides as demonstrated by the overlay of the homomeric and racemic structures.¹⁰ We reasoned that the chemically similar macrolittin 70 would be a good candidate for structural elucidation via racemic crystallization. We screened both homochiral and racemic macrolittin 70 for crystal hits. We found that the racemic mixture of macrolittin 70 crystallized much more readily than its homochiral counterpart. Well-defined crystals of the racemic system were observed in 14 of the 192 conditions screened versus 4 for the enantiomerically pure, L-macrolittin 70. In screening the crystals for X-ray diffraction, we were only able to observe diffraction for some of the racemic crystals and thus only moved forward with the optimization of the racemic system (Supplementary Figure 1).

Initial hits from 96-well screens were optimized using hanging-drop vapor diffusion methods in 24-well trays. We obtained crystals of various morphologies, which were appropriately sized for X-ray diffraction. Multiple conditions yielded such crystals; however, regardless of which condition we used, the single crystals diffracted X-rays poorly, displaying sparse, fuzzy patterns indicating disorder in the crystal lattice (Supplementary Figure 2). We observed an interesting phenomenon when transferring the crystal trays from the room in which they were prepared (average temperature of 22°C) to the room in which the 16°C environment: crystals suitable for electron diffraction was observed to form in the hanging drop. The temperature dependence of the formation of these crystals was made evident by the melting of the crystals when equilibrated back to 22°C (Supplementary Figure 3).

As a response to observing the formation of these small crystals in the 24-well optimization trays, we attempted batch crystallization with one of the crystallization conditions that had previously produced large crystals: 25% isopropanol, 0.2M NH₄OAc, 0.1M Tris-HCl, pH 8.5. However, to better replicate the observed crystal formation at lower temperatures, we chilled both the racemic peptide solution and the crystallization reagent to 4°C. When the two chilled solutions were mixed, needle crystals of appropriate size for electron diffraction immediately formed. However, when these crystals were deposited and vitrified, they diffracted poorly in a similar manner to the diffraction attempts with X-ray. We hypothesize that the crystallization reagent, isopropanol, was evaporating at such a rapid rate that the crystal lattice was becoming damaged during crystal extraction or vitrification, for X-ray and electron diffraction, respectively.

We then switched to a PEG3350-based crystallization condition for batch crystallization to avoid the adverse effects of the isopropanol. Again, a crystallization solution, chilled to 4°C was mixed with a chilled racemic peptide solution. Unlike the isopropanol condition, amorphous aggregates formed instead of needle crystals. Despondent, the batch was left in an ice bath overnight. Surprisingly, needle crystals, previously unobserved, started to form. The batch was stored at 4°C, and the number and size of the crystals grew (Supplementary Figure 4). The

diffraction quality of these crystals were much better than the crystal grown in isopropanol, and we decided to move forward with this batch condition.

Data Collection, Phasing, and Space Group Assignment

After optimizing the batch crystallization, the deposition and vitrification of the macrolittin 70 crystals were optimized for ED data collection. Unlike the tightly packed unit cells of small molecules and amyloidogenic peptides that we had handled in the past^{15, 16}, the macrolittin 70 crystal is expected to be packed loosely (owing to the relatively large size of macrolittin 70) and thus contain a large amount of solvent (estimated to be 37.61% if C2 asymmetric unit contains 4 helices). Since suboptimal packing is expected, it is reasonable to expect greater disorder and, thus, weaker diffraction. In an effort to boost the signal to noise ratio, we decided to increase our exposure time and decrease the oscillation angle per image (Supplementary Figure 5).

While the resolution of the data only reached around 2.3 Å, higher resolution would have been difficult to accommodate as 2.3 Å is already near the detector edge; a shorter camera distance would inevitably cause overlaps in the already proximal spots and prevent proper indexing and integration of the dataset. On a related note, beam drift at a longer camera distance is much more significant. Compounded with the slower rate of oscillation (overall longer data collection time), the accumulated beam drift significantly affected data processing after collection. XDS does not support frame to frame beam correction, but we circumvented the issue by separating the diffraction movies into halves and indexed and integrated them as separate datasets. Data from 4 different crystals, indexed and integrated as 6 data sets, were merged to achieve a reflection file with 82.5% completeness at 2.3 Å resolution (Table S1). The space group of the unit cell was assigned as C2/c due to the clear presence of zonal absences.

The resolution of our data was insufficient to use direct methods successfully for phasing. Alternatively, molecular replacement seemed plausible; however, the quality of search models was at the time was relatively poor, and this molecular replacement is subject to model bias. So,

we turned to fragment based phasing. Many of the publicly available versions of fragment-based phasing programs could not handle the apparently centrosymmetric (C2/c) space group of our crystal. Isabel Usón, with developer access to one of the phasing programs, recoded it to accept the centrosymmetric C2/c space group. A standard α -helical fragment library, consisting of polyalanines, was used. Other than the usage of a specialized space group, no other significant roadblock was observed: initial phasing solution presented the expected difference density for the polyalanines to be mutated to their appropriate identities.

After phases were successfully retrieved with fragment-based methods in space group C2/c, the atomic model was refined. Despite best efforts, the lowest R_{work} achieved while enforcing C2/c symmetry and maintaining a R_{work} to R_{free} split of less than 5-7% was 27.85%. Additionally, when binned by resolution, high R-factors were observed at high-resolutions (Table S2). The inability to refine the structure to R-values less than 25% and the high R-factors observed in the high-resolution bins led us to suspect pseudo-symmetry. Since there is precedence with racemic melittin⁶, we suspected that there is some breakage in the glide symmetry. In order to confirm our suspicions, we expanded the atomic coordinates of the asymmetric unit from C2/c to C2 and refined the structure. To compare the coordinates of the L- and D- macrolittin 70 molecules, we inverted the asymmetric unit and superimposed the inverted coordinate with the original. The α -carbons of the superimposed asymmetric units had an RMSD value of 0.358 Å, signifying only small deviations in the backbone of L- and D-molecules. However, alignment using all atoms, produced an RMSD value of 0.947 Å. This result suggests that the bulk of the structural differences between L- and D- molecules are contributed by differences in side chain conformations. These differences in fine features in the unit cell was reflected in the high R-factors in the high resolution bins when C2/c was enforced during structure refinement. Releasing the glide plane restriction by expanding the atomic coordinates to C2 allowed the refinement process to account for these differences and resulted in lower R-factors. So while the original space group assignment was C2/c, we carried out the final refinements with C2 symmetry.

Assessment of the Racemic Macrolittin 70 Structure

The racemic macrolittin 70 structure shares many commonalities with the racemic melittin structure described in the previous communication. Similar to the racemic melittin (PDB 64oM) structure, the asymmetric unit consists of an antiparallel dimer. This dimer consists of hydrophobic “concave” face and hydrophilic “convex” face. Upon generation of symmetry partners in the racemic macrolittin 70 crystal lattice one sees that these dimers form tetrameric assemblies resulting from the burying of the hydrophobic concave faces of two macrolittin 70 dimers. However, there are marked differences observed between the structures. The residue modifications made to generate macrolittin 70 from melittin include drastic changes to the side chains such as: hydrophobic to anionic mutations (A⁴_E, V⁸_E, and A¹⁵_E), cationic to hydrophobic mutations (R²²_A and K²³_A) and a polar to hydrophobic mutation (T¹⁰_A). The cationic to hydrophobic mutations are particularly noteworthy as they are clustered around the C-terminus and disrupt the amphiphilic nature of melittin, a characteristic that is considered highly important to its mode of action. In the macrolittin 70 structure, the hydrophilic side chains are more evenly dispersed in the sequence. These modifications lead to noticeable differences in the peptides’ hydrophobic surfaces (Figure 1). While both melittin and macrolittin 70 exhibit a α -helix-bend- α -helix motif, the helices in the melittin structures adopt a pronounced bend of 130° initiated by the presence of proline, while macrolittin 70 adopts a subtler 160° kink. This more minor bend of the macrolittin 70 peptide could be partially attributed to the G¹²_L mutation which may render the macrolittin 70 structure less flexible.

Of note there is a comparative tetrameric quaternary structure observed in the crystal structure of the host-defense peptide, LL-37 (PDB 7PDC).¹⁷ Similarly to the racemic melittin and macrolittin 70 structure, LL-37 adopts a tetrameric arrangement consisting of two antiparallel dimers.¹⁷ The dimers can also be characterized by a convex surface and a concave surface observed for the tetrameric assemblies of melittin and macrolittin 70. Dissimilarly, the convex

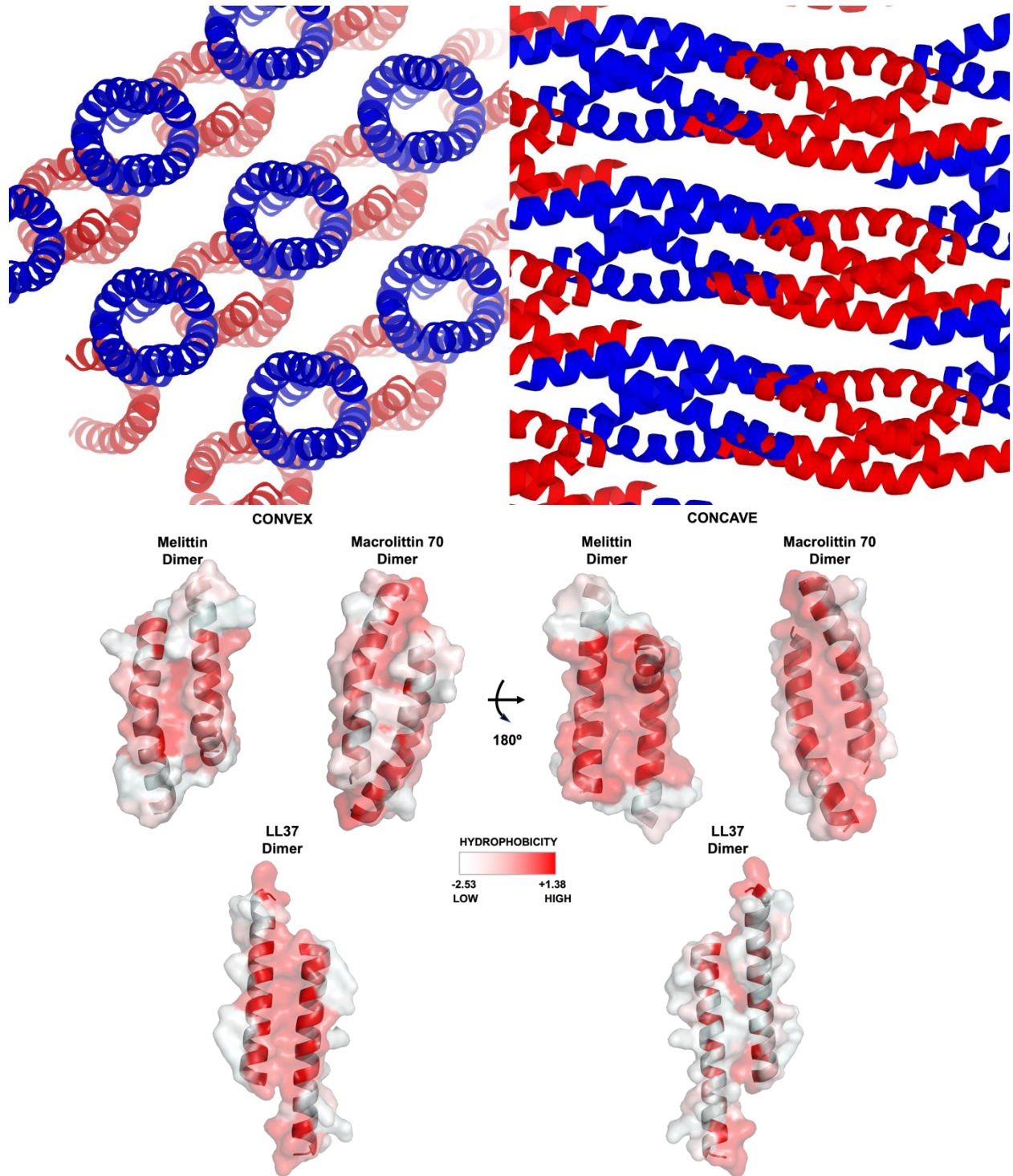


Figure 1: Top panel shows profile and top-down view of racemic macrolittin 70 crystal lattice. L-peptides are colored red while D-peptides are colored blue. Bottom panel shows comparison of the concave and convex hydrophobic surfaces of the antiparallel dimers observed in the structures of racemic macrolittin 70, racemic melittin, and LL-37. The surfaces are colored according to the Eisenberg scale, which quantifies the hydrophobicity of the 20 proteinogenic amino acids.¹⁸

surface of the LL-37 dimer is markedly more hydrophobic than those of the melittin and macrolittin 70 structures (Figure 1). The concave surface of the L-L37 dimer is ionic and not highly hydrophobic (Figure 1). The authors of this structure suggest that hydrophobic nature of the convex surface lends itself to being able to associate in the membrane. They thus hypothesize that the tetrameric structure of LL-37 is representative of the active pore.¹⁷ The relevance of the quaternary contacts observed in these structures to biological activity is remains difficult to discern.

In addition to differences in the homochiral contacts in the racemic melittin and macrolittin 70 structures, there is significant differences in the heterochiral contacts that make up the crystal lattices. Both crystal lattices are composed of L-peptide layers and D-peptide layers. In the racemic melittin structure the convex faces of the L- and D-tetramers align against one another while in the macrolittin 70 structure, the convex faces of the homochiral tetramers align against the convex faces of tetramers of the same chirality. We had previously observed hydrogen-bonding contacts between the L- and D-melittin peptides. While most of the residues that participated in hydrogen bond interactions between L- and D-peptides in the melittin structure are retained in the macrolittin 70 sequence, we do not observe these same hydrogen-bonding interactions in the racemic macrolittin 70 crystal. In fact, we do not observe any intimate contacts between the L- and D-peptides in the racemic macrolittin 70 structure.

Compared to our previous work with racemic melittin, much more effort was necessary to isolate single crystals of macrolittin 70 for diffraction. This observation led us to hypothesize that the racemic macrolittin 70 crystal lattice could be weaker and more susceptible to mechanical perturbation than the racemic melittin lattice. To assess our hypothesis, we calculated the buried surface area (BSA) and $\Delta G_{\text{formation}}$ using PDBePISA (Table S3). If the BSA is lower and the $\Delta G_{\text{formation}}$ higher for macrolittin 70, it would give us more confidence that the difficulties in isolating diffraction-capable crystals is due to differences in crystal lattice susceptibility to external mechanical forces or solvent evaporation during extraction or vitrification.

The results did not entirely support our hypothesis. On one hand, the melittin helical bundle is evaluated to be much more stable; its energy of formation is very favorable and its molecular interfaces bury a large amount of surface area within the bundle. On the other hand, the BSA and ΔG values indicate that the interactions of the helical bundle within its own enantiomeric layer and with adjacent bundles of the opposite chirality are much weaker for melittin than for macrolittin. This was a surprising result as we have observed many more close contacts between the L- and D-bundles in the melittin structure in comparison to the macrolittin 70 structure. This disagreement between the observed contacts and the ΔG scores could be attributed to the fact that ΔG of solvation estimated by PDBePISA do not consider the contributions of salt bridging and hydrogen bonds. This is especially egregious since the interfaces that were calculated to contribute most significantly to the positive ΔG values of the enantiomeric interfaces are heavily affected by the presence of Arg24, which is mutated to a glutamine in macrolittin 70 and thus absent. Unfortunately, the best algorithms we have for evaluating energy from atomic coordinates are insufficient to explain the difference in crystal mechanical strength.

Vectorization of Hydrogen Bonds

Perplexed by disagreement between the calculated and the observed strength of the crystal lattice, we decided to explore a key structural difference that we observed between the secondary structures of the two crystal structures: the proline kink. Intuitively, we hypothesized that hydrogen bonds within a more kinked helix would provide greater resistance against tensile or compressive forces in more directions, compared to hydrogen bonds in a more rectilinear helix (Supplementary Figure 6). To test this hypothesis, the full unit cells of melittin and macrolittin 70 were recapitulated by applying symmetry operators, and all valid hydrogen bonding pairs were identified (Supplementary Figure 7). The hydrogen bond vectors, forward and reverse, were summed to assess the directional isotropy of hydrogen bonding. For comparison, we executed

the same analysis with Proteinase K, a globular protein, TMEM106B¹⁹, an amyloid fiber (Supplementary Figure 8), and PE/PEE²⁰, an α -helical protein complex, infamous for its anisotropic lattice and diffraction.

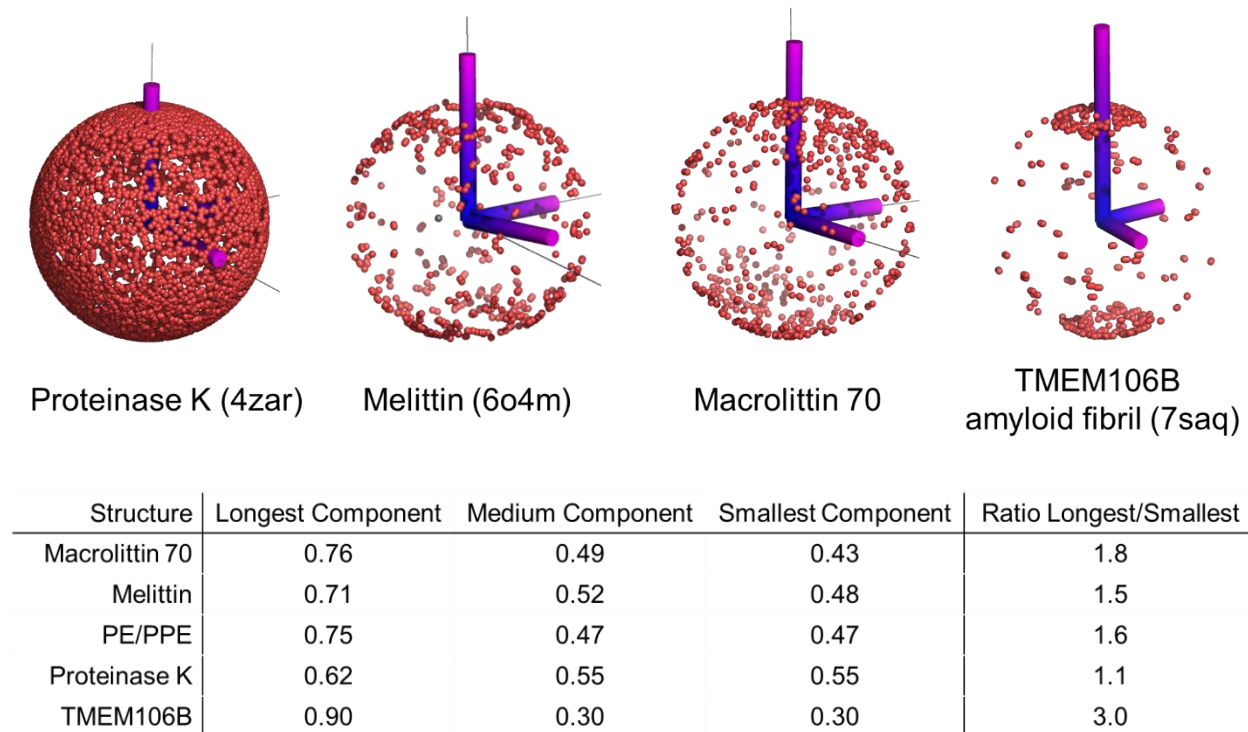


Figure 2: Directional distribution of hydrogen bonds in multiple structures represented as points on a sphere based on their direction. The blue and magenta rods represent the length of the axial components contributing to a unit vector that represents the summation of all the unit vectors created from the vectorization of the hydrogen bonds. The table below quantifies the axial components of the unit vector.

The hydrogen bond vectors, forward and reverse, were scaled to unit length so the tip of each vector is plotted as a dot on a unit sphere (Figure 2). The three principle components of the vector sum are shown as magenta cylinders, magnified 2x beyond the unit length for visibility. Our calculations indicate that macrolittin 70 has the most biased component along the helical axis, i.e. the vector in the z direction is significantly larger than those in the x and y directions, surpassed only by the beta sheets observed in TMEM106B. Not only is the bias in macrolittin 70 the strongest, it also has the highest ratio of longest to shortest axial component among the helical structures. PE/PPE, a structure noted for strong diffraction anisotropy, ranked a close second

while melittin exhibit the least polarization. The difference in strength between the melittin and macrolittin 70 lattices potentially arises in part from this biased directionality of the hydrogen bonds.

Phasing with AlphaFold Model

After solving the structure using fragment-based methods, AlphaFold v2 and its subsequent updated versions were released^{21, 22}. We decided to compare our fragment-based phasing results with the results we may achieve with an AlphaFold model. To this end, two models were generated with AlphaFold v2.1: a monomer and a tetramer²³. With a superimposition RMSD value of 3.084 Å, the AlphaFold's predicted tetramer overlaid poorly with the helical bundle observed in the crystal structure. The AlphaFold model predicted that all helices align side by side with the heads and tails aligned on the same top and bottom plane. On the other hand, the heads and tail of the crystal structure are offset from one another along the helical bundle's rise (Supplementary Figure 9). When used as a search model for molecular replacement phasing, the tetramer did not yield a usable electron density map.

Discarding the tetramer as a valid molecular replacement search model, we moved on to testing AlphaFold's monomer prediction as a valid MR search model. To this end, we tested three versions of the AlphaFold monomer with PHENIX Phaser: the monomer with all native R-groups, the monomer stripped to a polyalanine chain, and the monomer stripped to a polyglycine chain. Phasing with the native monomer yielded 10 solutions with a top TFZ score of 4.9 and top LLG score of 32.638, and the polyalanine chain resulted in 24 solutions with a top TFZ score of 6.2 and top LLG score of 60.765. Neither yielded a usable electron density map. Surprisingly, the polyglycine chain yielded a single MR solution with a top LLG score of 139.207 and top TFZ score of 9.5. The output of a single solution and the corresponding high scores indicate a solution was

found. However, even a brief inspection of the electron density brought concerns immediately: little to no difference density was found near the α -carbons to indicate the presence of side chain atoms. While not unprecedented, the slow addition of atoms to the positive difference density followed by structure refinement usually slowly increases the difference density observed. However, this did not happen. After hours of testing, we determined that the only way to converge the initial phasing solution achieved with the AlphaFold model and the structure phased by with fragment-based methods was to: 1) mutate all residues to their native amino acid identities without the positive (green) electron density to justify their mutation and 2) disregard the backbone information suggested by AlphaFold and carry out a head to tail real space refinement based on electron density and geometric restraints. Molecular replacement with the AlphaFold predicted

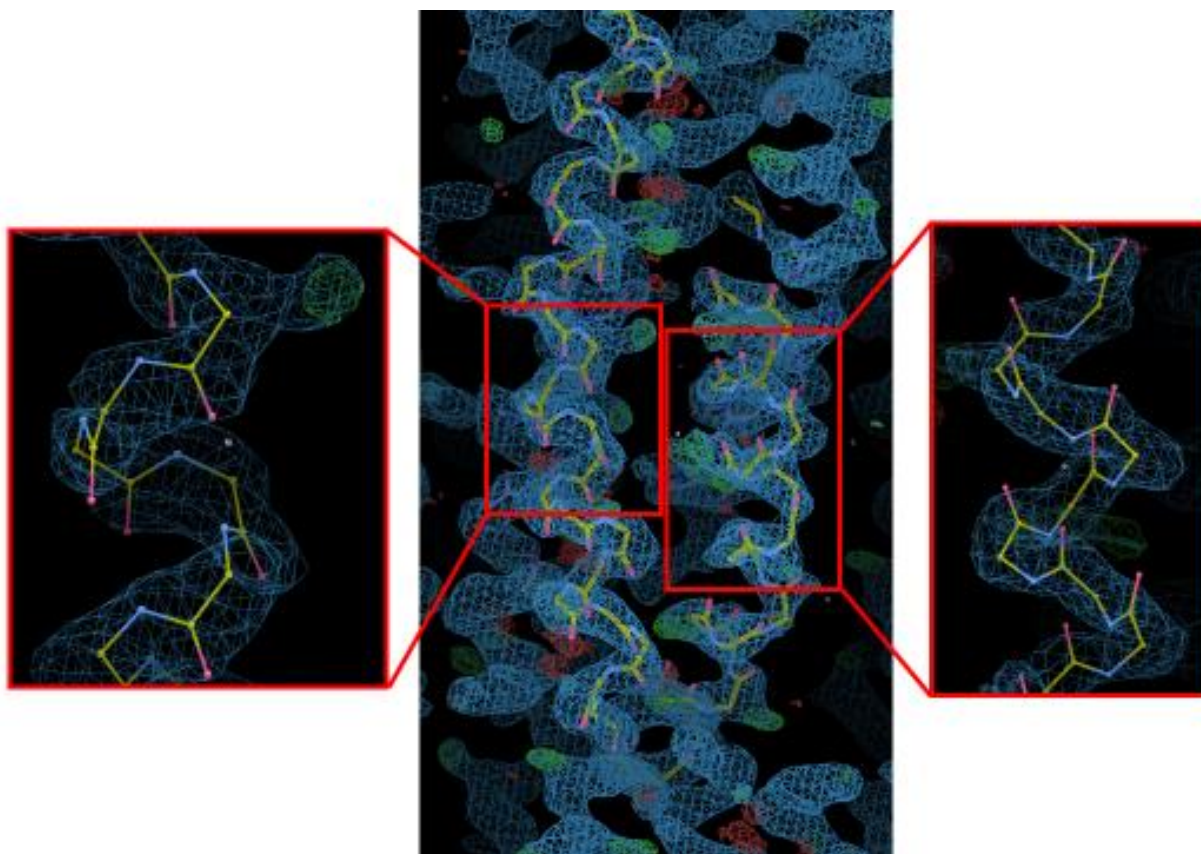


Figure 3: The initial solution outputted by PHENIX Phaser when using the AlphaFold monomer as a search model. The red insets correspond to the boxed regions in the center image. Very little positive difference density was present at the side chain locations to justify the mutation of the polyglycines in the chain to their proper native amino acids.

models proved to be a less advantageous strategy than using the fragment-based phasing approach.

Discussion

In this work, we combined racemic crystallization methods with MicroED data collection to acquire a 2.3-Å structure of racemic macrolittin 70, phased by fragment-based methods. Racemic crystallization and MicroED are methods that excel in rescuing samples recalcitrant to traditional X-ray techniques. In addition to providing the potential for simplifying the phase problem, racemic crystallization is a valid alternative when homochiral crystallization falls short.

Micro-ED is employed to acquire high resolution diffraction data from crystals too small for routine X-ray methods. Both these methods present a path toward acquiring structures of difficult systems. Of note, no unique sample preparation was necessary in solving this structure; any facility with access to vitrification and a diffraction capable electron microscope should find our methods accessible.

The lack of success when utilizing traditional X-ray techniques to extract and diffract macrolittin 70 crystals was surprising due to the wide variety of condition in which crystals formed. The poor outcome could be explained by 1) poor unit cell organization during crystal growth, 2) the crystal extraction process resulted in the destruction of the unit cell, or 3) a combination of both. In the instance of poor unit cell organization, MicroED is expected to give better results since a smaller area of a crystal is being sampled. The smaller area sampled by ED consists of fewer crystal domains, lowering the potential for high mosaicity and the amount of crystal disorder captured in diffraction. There may have been an advantage in the crystal extraction process for MicroED versus X-ray diffraction as well. The vitrification process used to prepare electron diffraction samples decreased the amount of time the crystals are exposed to atmosphere comparative to X-ray sample preparation, possibly sparing the crystal from total collapse as a result of solvent evaporation and/or tensile forces. The structural determination of racemic

macrolittin 70 is an example of how MicroED can be used to acquire high resolution data with a crystal of high mosaicity and/or fragility.

We carried out buried surface area and $\Delta G_{\text{solvation}}$ calculations in an effort to quantify the heightened frailty observed for the racemic macrolittin 70 crystal lattice in comparison to the racemic melittin crystal lattice. Surprisingly, not only does the macrolittin 70 structure have a higher calculated buried surface area than melittin, the calculated ΔG values suggest that macrolittin 70 lattice is thermodynamically more favorable than its racemic melittin counterpart. Despite these quantities, a visual inspection of the two structures intuitively led us to hypothesize that the directionality of the hydrogen bonds plays a key role in explaining the fragility of the macrolittin 70 lattice. Indeed, the vectorization of the hydrogen bonds in the unit cell show a strong polarization, parallel to the α -helical axes, for both melittin and macrolittin 70. When scaled to a unit vector and divided into its contributing components, the polarization found in the macrolittin 70 cell is higher than both melittin and the PE/PPE protein complex, a helical bundle complex notorious for its diffraction anisotropy. The observed differences between the melittin and macrolittin 70 can be compared to the difference between diamond (tetrahedral carbon) and graphite (trigonal planar carbon). The bonding enthalpy for diamond is higher than graphite and, thus, numerically less stable. Yet the directionality of the bonds in diamond create a much more stable superstructure than the 2-dimensional planar sheets that constitutes graphite, which allows the sliding of planar carbon layers. The more unidirectional nature of the hydrogen bonding observed in macrolittin 70, analogous to the sheet defined graphite structure, makes the lattice susceptible to alteration by external forces. On the other hand, the higher direction variability observed for hydrogen bonds in melittin yield a much stronger superstructure despite the numerically favorable buried surface area and $\Delta G_{\text{formation}}$ values for macrolittin 70.

The publication of AlphaFold presented a good opportunity to compare molecular replacement strategies using models generated from fragment-based methods versus AlphaFold. AlphaFold was unable to predict tetrameric α -helical bundle structure we solved for macrolittin 70.

We were eventually able to solve the structure by MR with AlphaFold's monomer model; however, multiple leaps of faith were necessary from getting the initial phasing solution to the final refined structure: 1) stripping the search model to a polyglycine chain rather than to a standard polyalanine one prior to phasing; 2) mutating the amino acids in the initial phasing solution to match the peptide sequence without the presence of electron density to justify the placement of the R-groups; and 3) real space refinement of the entire chain in COOT based on electron density and geometric restraints prior to refinement in Phenix, disregarding the backbone information. The necessity to mutate the search model to a polyglycine chain is bizarre and is not standard practice during MR phasing. The mutation of side chains without density justification and head to tail backbone refinement are sometimes executed, but the need to perform such refinement practices are common indicators of a poor or even incorrect initial solution. This leads us to conclude that the fragment-based approach out-performed AlphaFold in this case. Indeed, without prior knowledge provided by solving the structure via fragment-based methods, it is unlikely the structure would have been solved by using the monomer model generated by AlphaFold.

In conclusion, our experiments show that electron diffraction techniques have strong utility in studying crystals constituted by weak lattices. The vitrification process may have significantly reduced the likelihood of lattice collapse by a combination of maintaining sample hydration and exerting less mechanical forces on the crystal. The racemic macrolittin 70 unit cell, with symmetry C2, contains 416 amino acids and ample space for disorganized solvent molecules to reside, not unlike a protein crystal lattice. This structure not only further exhibits the utility of fragment-based phasing methods for structural elucidation using Micro-ED methods but also demonstrates the potential of electron diffraction for solving novel protein structures. Additionally, our experiments demonstrate our increased ability to obtain *ab initio* phases for electron diffraction datasets and elucidate structures without the usage of specialized procedures to make a crystal more amendable for electron diffraction.

Acknowledgements

We would like to thank Todd O. Yeates for his insightful suggestions about space group assignment and discussion concerning symmetry breakage. We would like to thank Duilio D. Cascio and for his management of the UCLA-DOE crystallography facilities and discussion on the project in general.

Methods

Synthesis and Purification of L- and D- Macrolittin70

The L- and D-peptides were prepared by microwave-assisted solid-phase peptide synthesis using a Liberty Blue™ automatic synthesizer (CEM). Macrolittin 70's sequence is H₂N-GIGEVLKELATTPELQSWIKAAQQL-CONH₂.² Dry nova PEG rink amide resin (100 μmol, EMD Millipore) was added to the sample loader and swelled according to the Liberty Blue high swell method with N,N-dimethylformamide (DMF, Sigma Aldrich). An Fmoc-protected PAL linker (ChemImpex International Inc.) was coupled to the resin via addition of 2.5 mL of 0.2 M Fmoc-PAL linker (5 equiv.) in DMF, 1.0 mL of 1 M *N,N*-diisopropylcarbodiimide (DIC, ChemImpex International Inc., 10 equiv.) in DMF and 0.5 mL of 1.0 M ethyl cyano(hydroxyamino)acetate (oxyma pure, ChemImpex International Inc., 5 equiv.) in DMF, as recommended by CEM, to the reaction vessel. The reaction vessel was heated to 75°C for 15 seconds, and then the heat was increased to 90°C for 110 seconds. Deprotection of the Fmoc group was performed by addition of 20% piperidine (Sigma-Aldrich) in DMF into the reaction vessel and heating the reaction vessel to 75°C for 180 seconds. Coupling and deprotection of Fmoc-amino acids (ChemImpex International Inc.) was achieved in the same fashion as described for the Fmoc-PAL linker, with the Fmoc-PAL linker replaced by 2.5 mL of 0.2 M Fmoc-L- or D-amino acid (5 equiv.) in DMF.

Following the deprotection of the N-terminal Gly residue, the resin was transferred to a fritted reaction vessel (Torviq). The peptide was then cleaved from the resin by addition of 4 mL of a cleavage cocktail consisting of 2.5 % H₂O, 2.5% triisopropylsilane (Sigma-Aldrich) and 95%

trifluoroacetic acid (TFA, Sigma-Aldrich). The solution was then rocked for 2 hours and subsequently filtered through the fritted syringe. The resin was washed 2 more times with TFA, and the filtered TFA solutions were combined. The TFA in the combined solution was evaporated under a stream of nitrogen. To the resulting brown oil was added chilled diethyl ether, which caused precipitation. Crude peptide was collected as a white powder.

Crude products of the L- and D-macrolittin 70 syntheses were purified via RP-HPLC using a Waters Prep LC150 System. The purification was conducted with an XSelect® CSH™ Prep C18 column (5 µm particle size, 19 mm x 250 mm) using an optimized purification method consisting of a gradient elution of 48-54% B solvent over 15 minutes. Solvent A was 0.1% TFA in filtered H₂O, and solvent B was 0.1% TFA in HPLC-grade acetonitrile (Sigma-Aldrich). The purities of the peptides were analyzed by RP-HPLC, using a Hewlett Packard 1090 Series II Liquid Chromatography System equipped with a Kinetex (Phenomenex, C18, 5 µm, 100 Å, 4.6 x 250 mm) column. Ultrapure water with 0.1% TFA, and a 1:9 water to acetonitrile solution with 0.095% TFA were selected as mobile phases [A] and [B], respectively. The flow rate was set at 1.0 mL/min. The analysis portion of the gradient lasted for 20 minutes with 1% per minute increase of [B] from 50% to 70% [B] (Supplementary Figures 10 and 11). The masses of the peptides were confirmed via MALDI-TOF mass spectrometry; data were acquired using a Bruker ULTRAFLEX™ III equipped with a SmartBeam™ laser (Supplementary Figures 12 and 13).

Crystallization Trials for X-ray Diffraction

Proper conditions for the crystallization of racemic (5 mg/mL, a 50:50 mixture of 5 mg/mL L- and D-macrolittin 70 solutions) and L-homochiral macrolittin 70 (5 mg/mL) samples were determined by conducting trials with Crystal and Index HT crystal screens (Hampton). Several crystal hits were found for both samples. Four distinct crystal morphologies were observed in the racemic macrolittin 70 samples, while only one crystal morphology was observed in the

homochiral system (Supplementary Figure 1). These hits were used as a guide for optimization of crystal conditions.

Further optimization was carried out in 24-well trays (Table S4). 2 μ L of crystallization reagent and 2 μ L of peptide solution were added either mixed or unmixed onto siliconized or siliconized glass. All wells had the same condition of 25% iPrOH, 0.1 M Tris pH 8.5, 0.2 NH₄OAc. The initial concentration of the peptide solution was either 10 mg/mL or 5 mg/mL. This value denotes the concentration of each enantiomer separately not in total. The temperature dependent crystal formation was discovered while working with this particular tray (Supplementary Figure 3).

X-ray Diffraction Data Collection

Diffraction data were collected from the crystals from two conditions: 25% isopropanol, 0.2M NH₄OAc, 0.1M Tris-HCl, pH 8.5 and 30% PEG 3350, 0.2M NH₄OAc, 0.1M HEPES, pH 7.5. Crystals were vitrified in liquid nitrogen after looping. Diffraction data were collected at Life Sciences Collaborative Access Team (LS-CAT) beamline 21-ID-D using a Dectris Eiger 9M detector at the Advanced Photon Source (APS), Argonne National Laboratory. Diffraction was observed for all samples but suffered from poor resolution and spot definition. The data were not of high enough quality to allow for structural elucidation.

Batch Crystallization

Two 20 mg/mL peptide solution were made: one for L-macrolittin 70 and one for D-macrolittin 70. This was done by weighing out 0.8 mg of each enantiomer into separate 1-mL Eppendorf tube and adding 40 μ L of mammalian cell culture grade water to the weighed powders. Then, 20 μ L of each enantiomerically pure solution was added to a 1-mL tube with an o-ring sealed cap to minimize evaporation over time. This racemic peptide solution (10 mg/mL of each enantiomer) was cooled to 4°C along with a tube of crystallization solution (25% isopropanol, 0.2M NH₄OAc, 0.1M Tris-HCl, pH 8.5). After temperature was equilibrated, 40 μ L of crystallization solution was added to the peptide solution. Crystals immediately formed. This was repeated with

the solution temperatures equilibrated to 15.5°C (the average temperature of the X-ray instrumentation room at UCLA-DOE). Again, crystals immediately formed.

After the isopropanol crystallization condition was determined to be inadequate. An alternative condition, 0.2 M NH₄OAc, 0.1 M Tris Buffer, pH 8.5, and 35% PEG 3350, was selected. A 15-mL solution of the crystallization condition was made. A 1 mL aliquot was taken from this stock and subjected to a 2-fold dilution, giving a final concentration of 0.1 M NH₄OAc, 0.05 M Tris Buffer, pH 8.5, and 17.5% PEG 3350. The former solution will be used for crystallization while the later solution will be used for manipulations after the formation of the crystals, such as dilutions on the EM grid during freezing.

For the first batch crystallization, 60 and 53 µL of mammalian cell culture grade water was added to the 0.60 mg of L-peptide and 0.53 mg of D-peptide in 1-mL Eppendorf tubes, respectively. For the first batch, 25 µL from each tube was then mixed together in an o-ring sealed tube, yielding a 50 µL racemic peptide solution with a 5 mg/mL peptide concentration for each enantiomer. From this point, all solutions are kept in an ice bucket to keep temperatures for all solutions as close to 0°C. After enough time had been giving for equilibration, 50 µL of crystallization solution was added to the peptide solution, yielding a 100 µL solution. Amorphous aggregation was immediately observed upon mixing of the solution. The batch crystallization was kept on ice overnight and small crystals were observed to form (Supplementary Figure 4). The batch was transferred to a 4°C fridge for storage. Not only did more crystals form over time, there was also an increase in average crystal size.

Collection of MicroED data by continuous rotation

1.8 microliters of aqueous crystallization solution (0.1 M NH₄OAc, 0.05 M Tris Buffer, pH 8.5, and 17.5% PEG 3350) was added to the copper-exposed side of a glow-discharged holey carbon grid (C-flat™, 2/2 300 mesh Cu, Electron Microscopy Sciences) while 1.8 microliters of crystal slurry was applied to carbon-layered side. The grid was plunged into liquid ethane using

an FEI Vitrobot Mark IV set to a blot time of 24 and a blot force of 20. Diffraction patterns and crystal images were collected under cryogenic conditions using a FEI Tecnai F30 operated at 300keV in diffraction mode. Diffraction patterns were recorded while continuously rotating at 0.188 degrees per second using a bottom mount TemCam-F416 CMOS camera (TVIPS). Individual image frames were acquired with 5 second exposures per image using selected area apertures corresponding to approximately 5 μm at the sample plane. A virtual detector distance of 3520 mm was used to prevent overlapping reflections for such a large unit cell.

MicroED data processing

The collected TVIPS movies were converted to individual images in SMV format, which are compatible with x-ray data processing software. The diffraction images were indexed and integrated with XDS. Raster size and scan pattern for indexing and integration in XDS were optimized to minimize contributions by background. Due to the longer than usual detector distance, beam shifts was more significant and accumulated over time during data collection. XDS does not account for beam shift between frames, so each movie had to be divided into two datasets for indexing and integration. Thorough, beam center optimization was essential. The reflection outputs from XDS were sorted and merged in XSCALE. 224 diffraction images from 4 unique crystals (6 partial datasets) were merged to produce a final dataset with acceptable completeness up to 2.3 \AA . To determine the phases, a developer version of fragment-based phasing program was recoded by Isabel Usón Finkenzeller (Catlan Instiution for Research and Advanced Studies – Institut de Biologia Molecular de Barcelona) to allow the space group assignment of C2/c, a non-Sohncke space group, for fragment-based phasing.

Refinement

Refinement was executed with PHENIX refine, using their built-in electron form factors. Initially, refinement was carried out with C2/c as the space group assignment. With C2/c, the best R_{work} and R_{free} achieved was 27.85% and 34.83%, respectively. The atomic content of the

asymmetric unit was expanded from C2/c to C2 using PDBSET as pseudo-symmetry with the glide plane was suspected. Care was taken to insure the R_{free} flags of the original C2/c reflection file was transferred to C2 reflection file used for further refinement. This insured that no artificial R_{work} to R_{free} convergence occurred. Individual ADPs and group ADPs were not refined: refinement of these parameters caused a R_{work} , R_{free} split of 10% or more. Non-crystallographic symmetry and refinement of individual sites were enabled during refinement. Waters, ions, and rotamers were added and refined as normal. The final refinement resulted in a R_{work} of 22.09% and R_{free} of 28.28%.

Buried Surface Area and Energy Analysis

PDBePISA was used to calculate buried surface area (BSA) and $\Delta G_{\text{formation}}$. In order to clearly identify all lattice interfaces that a helical bundle participates in, the final refined structure was stripped of all spheres (water and ions). Then the terminal amides had to be mutated back to carboxylate acids because PDBePISA considers the terminal amides as a separate entity from the peptide chain, complicating calculations. Using PyMOL, a large lattice of the modified structure was generated. A bundle centered in the lattice, composed of 4 L-chains was selected, and labeled Chain A, B, C, and D (core, helical bundle). Any chains too distal and obviously not interacting with the core, helical bundle was deleted from the lattice. The unit cell dimensions were increased by an order of magnitude (484.20 306.90 604.90) and the space group assignment changed to P1, preventing PDBePISA from considering interactions of the generated lattice with symmetry generated chain. Finally, the calculations were executed. Any chain that does not interact with the core, helical bundle was pruned from the lattice (Supplementary Figure 14). Interactions between chains core, helical bundle were summed and labeled “bundle” in Table S3. Interactions between the core, helical bundle chains and any L-chain were summed and labeled as “L-”, denoting interactions with other L-bundles in the L-enantiomer bundle sheet. Interactions between the core helical bundle and any D-chain were summed and labeled as “D-”,

denoting interactions with D-bundles in the adjacent D-enantiomer bundle sheets. The above procedures were executed with the structure of racemic melittin (6O4M).

Hydrogen Bonding Analysis

To obtain a full list of hydrogen bonds, PDBSET was used to generate the full unit cell content of racemic melittin (6O4M), racemic macrolittin 70, PE/PPE (2G38), and proteinase K (4ZAR). As an outlier example, a multilayer structure of TMEM106B amyloid fiber (7SAQ) was also generated. Then HBPLUS v.3.06 was used to generate a complete list of possible hydrogen bonds in the unit cell. Michael R. Sawaya wrote an in-house algorithm that used the list generated by HBPLUS and the previously generated .pdb file containing the filled unit cell to: 1) identify atoms participating in hydrogen bonding; 2) calculate the 2 vectors of each hydrogen bond using the .pdb coordinates of the atoms identified in 1) and display them (Supplementary Figure 7 and 8); 3) compile and plot the vectors as unit vectors on a sphere (Figure 2); and 4) quantify the directionality of the unit cell by summing the hydrogen bond vector components.

AlphaFold and Molecular Replacement Phasing

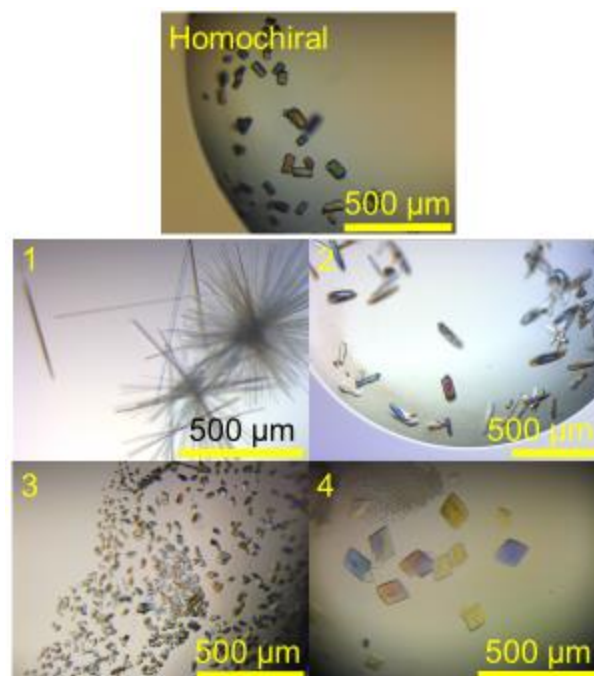
A publicly available online version of AlphaFold2 was used predicted monomeric and tetrameric²⁴. PyMOL was used to compare the predicted tetramer and an all L-chain helical bundle built from the final refined structure. Phaser from the PHENIX suite was used to attempt molecular replacement with the predicted monomer. Three versions of the predicted monomer was used as the search model: 1) an unmodified version of the monomer with all side chains present; 2) the monomer model stripped to a polyalanine; and 3) the monomer model stripped to a polyglycine. The space group assignment was C2/c for the duration of the experiment. The polyglycine model yielded an acceptable initial solution and the electron density was evaluated in COOT. Side chain mutations and head to tail real space refinements were also executed in COOT, followed by refinement with PHENIX refine.

References

1. Gallagher-Jones, M.; Bastillo, 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* (3), 490–499.
2. Andrusenko, I.; Hamilton, V.; Mugnaioli, E.; Lanza, A.; Hall, C.; Potticary, J.; Hall, S. R.; Gemmi, M. The Crystal Structure of Orthocetamol Solved by 3D Electron Diffraction. *Angewandte Chemie International Edition* **2019**, *58* (32), 10919–10922.
3. Bücker, R.; Hogan-Lamarre, P.; Mehrabi, P.; Schulz, E. C.; Bultema, L. A.; Gevorkov, Y.; Brehm, W.; Yefanov, O.; Oberthür, D.; Kassier, G. H.; Dwayne Miller, R. J. Serial Protein Crystallography in an Electron Microscope. *Nat Commun* **2020**, *11* (1), 996.
4. 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.
5. Krauson A. J.; He J.; Wimley W.C. Gain-of-Function Analogues of the Pore-Forming Peptide Melittin Selected by Orthogonal High-Throughput Screening. *J. Am. Chem. Soc.* **2012**, *134*, 12732–12741.
6. Wiedman, G.; Fuselier, T.; He, J.; Searson, P. C.; Hristova, K.; Wimley, M. C. Highly Efficient Macromolecule-Sized Poration of Lipid Bilayers by a Synthetically Evolved Peptide. *J. Am. Chem. Soc.* **2014**, *136*, 12, 4724-4731.
7. Wiedman, G.; Kim, S. Y.; Zapata-Mercado, E.; Wimley, M. C.; Hristova, K. pH-Triggered Macromolecule-Sized Poration of Lipid Bilayers by Synthetically Evolved Peptides. *J. Am. Chem. Soc.* **2017**, *139*, 2, 937-945.
8. Li, S.; Kim, S. Y.; Pittman, A. E.; King, G. M.; Wimley, W. C.; Hristova, K. Potent Macromolecule-Sized Poration of Lipid Bilayers by the Macrolittins, A Synthetically Evolved Family of Pore-Forming Peptides. *J. Am. Chem Soc.* **2018**, *140*, 6441-6447.
9. Lam, K. S.; Lehman, A. L.; Song, A.; Doan, N.; Enstrom, A. M.; Maxwell, J.; Liu, R. Synthesis and Screening of “One-bead One-compound” Combinatorial Peptide Libraries. *Methods Enzymol.* **2003**, *369*, 298-322.
10. Kurgan, K. W.; Kleman, A. F.; Bingman, C. A.; Kreidler, D. F.; Weisblum, B.; Forest, K. T.; Gellman, S. H. Retention of Native Quaternary Structure in Racemic Melittin Crystals. *J. Am. Chem. Soc.* **2019**, *141*(19), 7704-7708.
11. Yeates, T. O.; Kent, S. B. H. Racemic Protein Crystallography. *Annual Review of Biophysics* **2012**, *41* (1), 41–61.

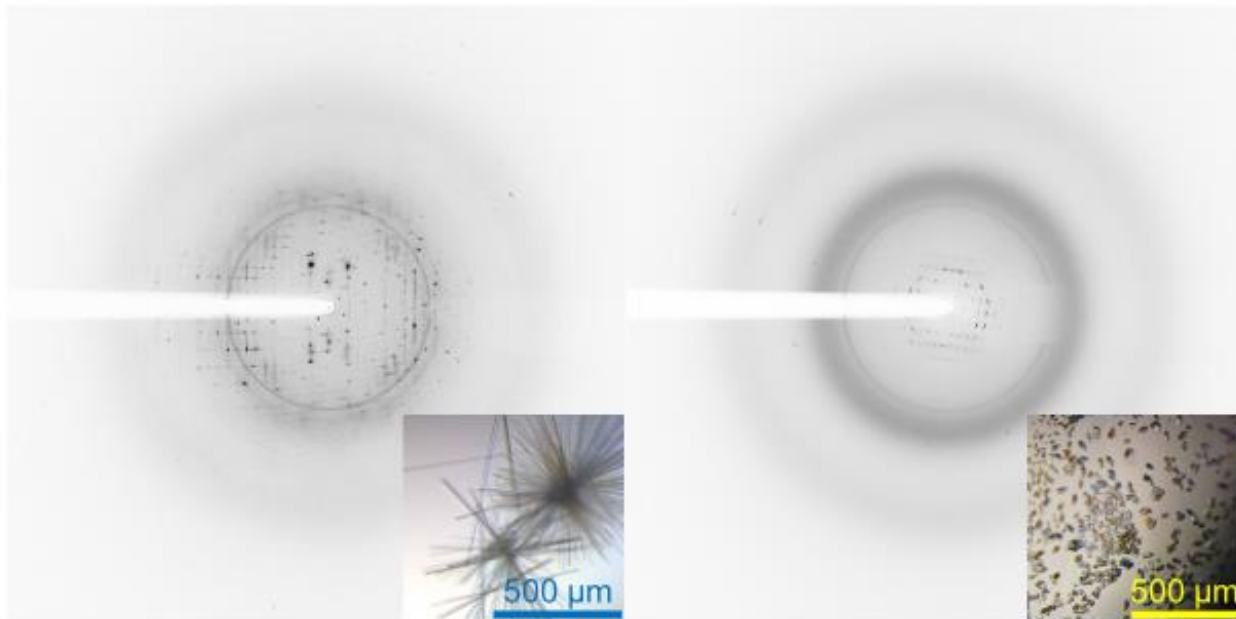
12. Wang, C. K.; King, G. J.; Conibear, A. C.; Ramos, M. C.; Chaousis, S.; Henriques, S. T.; Craik, D. J. Mirror Images of Antimicrobial Peptides Provide Reflections on Their Function and Amyloidogenic Properties. *J. Am. Chem. Soc.* 2016, *148*(17), 5706-5713.
13. Hayouka, Z.; Mortenson, D. E.; Kreitler, D. F.; Weisblum, B.; Forest, K. T.; Gellman, S. H. Evidence for Phenylalanine Zipper-Mediated Dimerization in the X-ray Crystal Structure of a Magainin 2 Analogue. *J. Am. Chem. Soc.* 2013, *135*(42), 15738-15741.
14. Hayouka, Z.; Thomas, N. C.; Mortenson, D. E.; Satyshur, K. A.; Weisblum, B.; Forest, K. T.; Gellman, S. H. Quasiracemate Crystal Structures of Magainin 2 Derivatives Support the Functional Significance of the Phenylalanine Zipper Motif. *J. Am. Chem. Soc.* 2015, *137*(37), 11884-11887.
15. Guenther, E. L.; Ge, P.; Trinh, H.; Sawaya, M. R.; Cascio, D.; Boyer, D. R.; Gonen, T.; Zhou, Z. H.; Eisenberg, D. S. Atomic-Level Evidence for Packing and Positional Amyloid Polymorphism by Segment from TDP-43 RRM2. *Nat Struct Mol Biol* **2018**, *25* (4), 311–319.
16. 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. *Proceedings of the National Academy of Sciences* **2016**, *113* (40), 11232–11236.
17. Sancho-Vaello, E.; Gil-Carton, D.; François, P.; Bonetti, E.; Kreir, M.; Pothula, K. R.; Kleinekathöfer, U.; Zeth, K. The structure of the antimicrobial human cathelicidin LL-37 shows oligomerization and channel formation in the presence of membrane mimics. *Sci. Rep.* **2020**, *10*, 17356.
18. Eisenberg, D.; Schwarz, E.; Komarony, M.; Wall, R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol.* **1984**, *179*(1), 125-142.
19. Jiang, Y. X.; Cao, Q.; Sawaya, M. R.; Abskharon, R.; Ge, P.; DeTure, M.; Dickson, D. W.; Fu, J. Y.; Ogorzalek Loo, R. R.; Loo, J. A.; Eisenberg, D. S. Amyloid Fibrils in FTLTDP Are Composed of TMEM106B and Not TDP-43. *Nature* **2022**, *605* (7909), 304–309.
20. Strong, M.; Sawaya, M. R.; Wang, S.; Phillips, M.; Cascio, D.; Eisenberg, D. Toward the Structural Genomics of Complexes: Crystal Structure of a PE/PPE Protein Complex from Mycobacterium Tuberculosis. *Proceedings of the National Academy of Sciences* **2006**, *103* (21), 8060–8065.
21. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohli, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589.

22. Tunyasuvunakool, K.; Adler, J.; Wu, Z.; Green, T.; Zielinski, M.; Žídek, A.; Bridgland, A.; Cowie, A.; Meyer, C.; Laydon, A.; Velankar, S.; Kleywegt, G. J.; Bateman, A.; Evans, R.; Pritzel, A.; Figurnov, M.; Ronneberger, O.; Bates, R.; Kohl, S. A. A.; Potapenko, A.; Ballard, A. J.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Clancy, E.; Reiman, D.; Petersen, S.; Senior, A. W.; Kavukcuoglu, K.; Birney, E.; Kohli, P.; Jumper, J.; Hassabis, D. Highly Accurate Protein Structure Prediction for the Human Proteome. *Nature* **2021**, *596* (7873), 590–596.
23. Evans, R.; O'Neill, M.; Pritzel, A.; Antropova, N.; Senior, A.; Green, T.; Žídek, A.; Bates, R.; Blackwell, S.; Yim, J.; Ronneberger, O.; Bodenstein, S.; Zielinski, M.; Bridgland, A.; Potapenko, A.; Cowie, A.; Tunyasuvunakool, K.; Jain, R.; Clancy, E.; Kohli, P.; Jumper, J.; Hassabis, D. Protein Complex Prediction with AlphaFold-Multimer. bioRxiv March 10, 2022, p 2021.10.04.463034.
24. Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M.; ColabFold: making protein folding accessible to all. *Nat Methods* **2022**, *19*, 679-682.

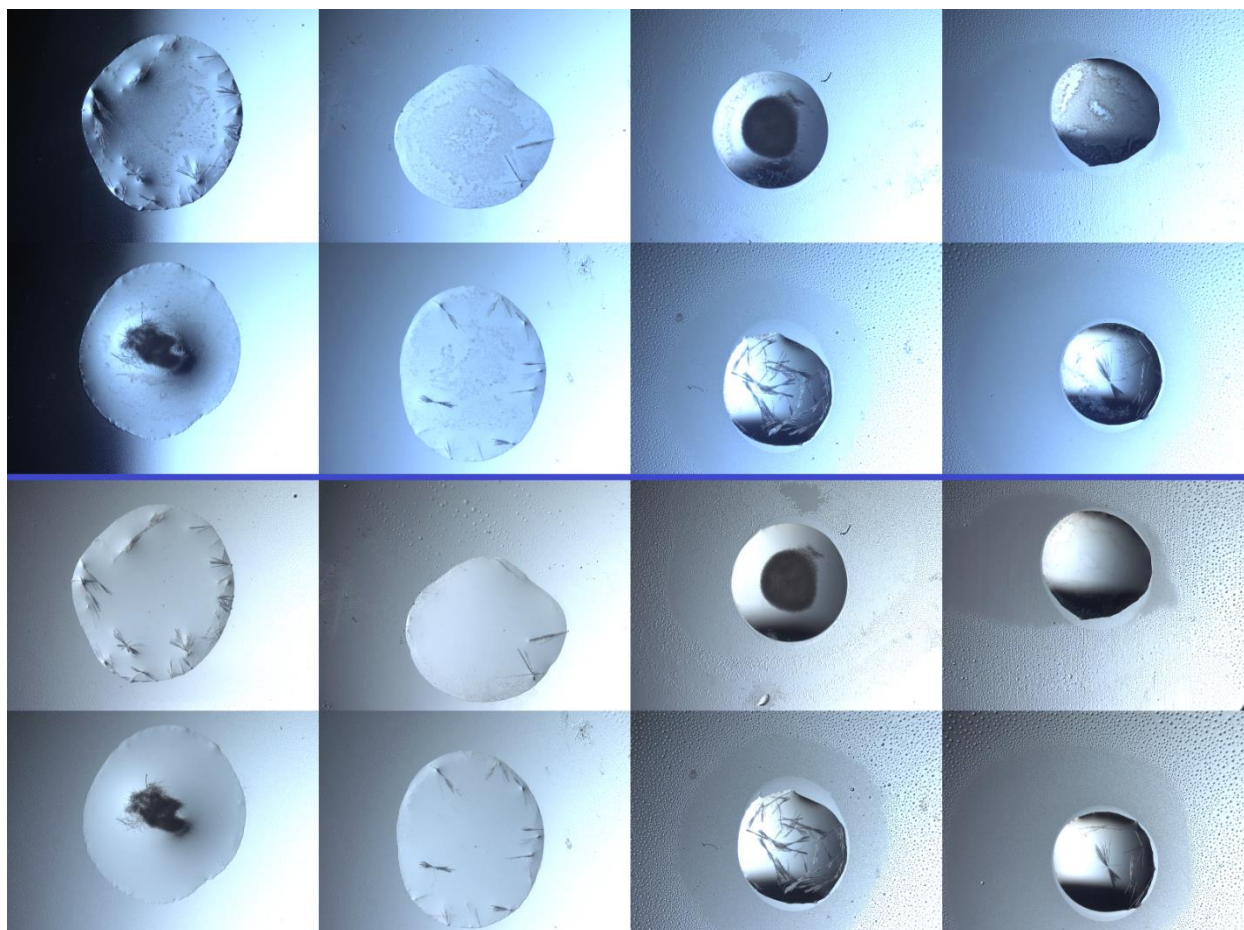


	Crystallization Reagent	Buffer	Salt
Homochiral	30% w/v polyethylene glycol monomethyl ether 550 25% w/v PEG 3350 30% w/v PEG 4000 20% w/v EtOH	0.1 M HEPES pH 7.5 0.1 M Bis-Tris pH 6.5 0.1 M Tris-HCl pH 8.5 0.1 M Tris-HCl pH 8.5	0.2 M MgCl ₂ • 6 H ₂ O 0.2 M MgCl ₂ • 6 H ₂ O 0.2 M MgCl ₂ • 6 H ₂ O
Racemic			
Morphology 1	20% v/v 2-PrOH 10% w/v PEG 6000 30% w/v polyethylene glycol monomethyl ether 2000 20% w/v EtOH 2% w/v 1,4-dioxane, 10% w/v PEG 20,000	0.1 M HEPES pH 7.5 0.1 M NaOAc • 3 H ₂ O pH 4.6 0.1 M Tris pH 8.5 0.1 M Bicine pH 9	0.2 M trisodium citrate • 2 H ₂ O 2.0 NaCl 0.2 M (NH ₄) ₂ SO ₄
	30% v/v 2-PrOH	0.1 M Tris-HCl pH 8.5	0.2 M NH ₄ OAc
Morphology 2	28% w/v polyethylene glycol monomethyl ether 2000	0.1 M Bis-Tris pH 6.5	0.01 M Zn ₂ SO ₄ • 7 H ₂ O
Morphology 3	25% polyethylene glycol monomethyl ether 550 30% w/v PEG 4000 30% w/v PEG 4000 25% w/v PEG 3350	0.1 M MES • H ₂ O pH 6.5 0.1 M NaOAc • 3 H ₂ O pH 4.6 0.1 M Tris-HCl pH 8.5 0.1 M HEPES pH 7.5	0.2 M NH ₄ OAc 0.2 M NaOAc • 3 H ₂ O 0.2 M NaCl
	25% w/v PEG 3350	0.1 M HEPES pH 7.5	0.2 M NH ₄ OAc
	25% w/v PEG 3350	0.1 M Tris-HCl pH 8.5	0.2 M NH ₄ OAc
Morphology 4	25% w/v PEG 3350	0.1 M Tris pH 8.5	0.2 M MgCl ₂ • 6 H ₂ O

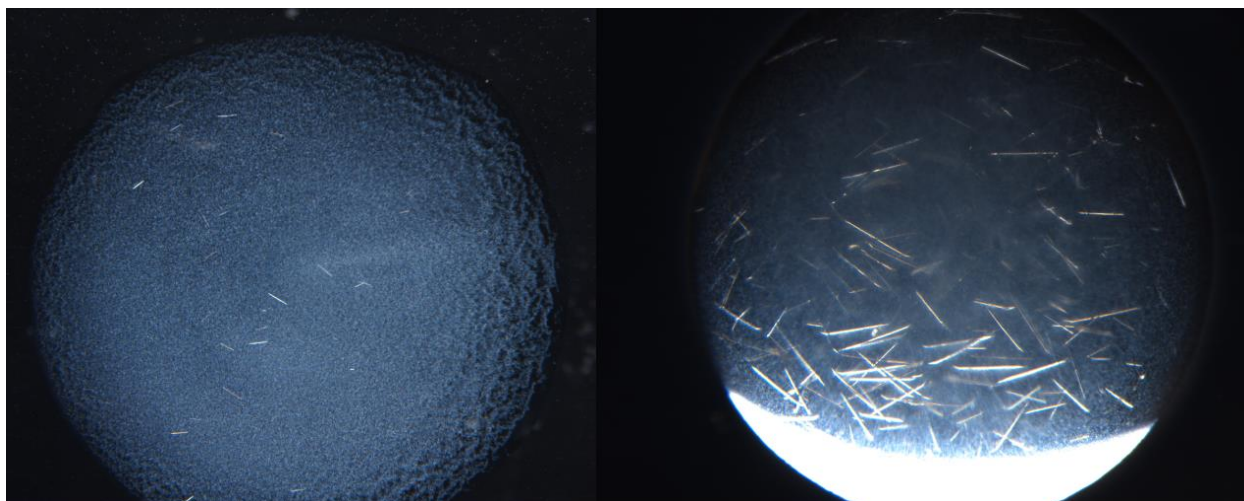
Supplementary Figure 1: Summary of successful crystallization trials for homochiral (L-) and racemic macrolittin 70. Crystal images at the top show the variety of morphologies observed during crystallization trials. All scale bars are set at 500 μm. Chart below lists all conditions that produced the pictured morphologies. The conditions highlighted in blue and yellow were selected for diffraction (Supplementary Figure 2).



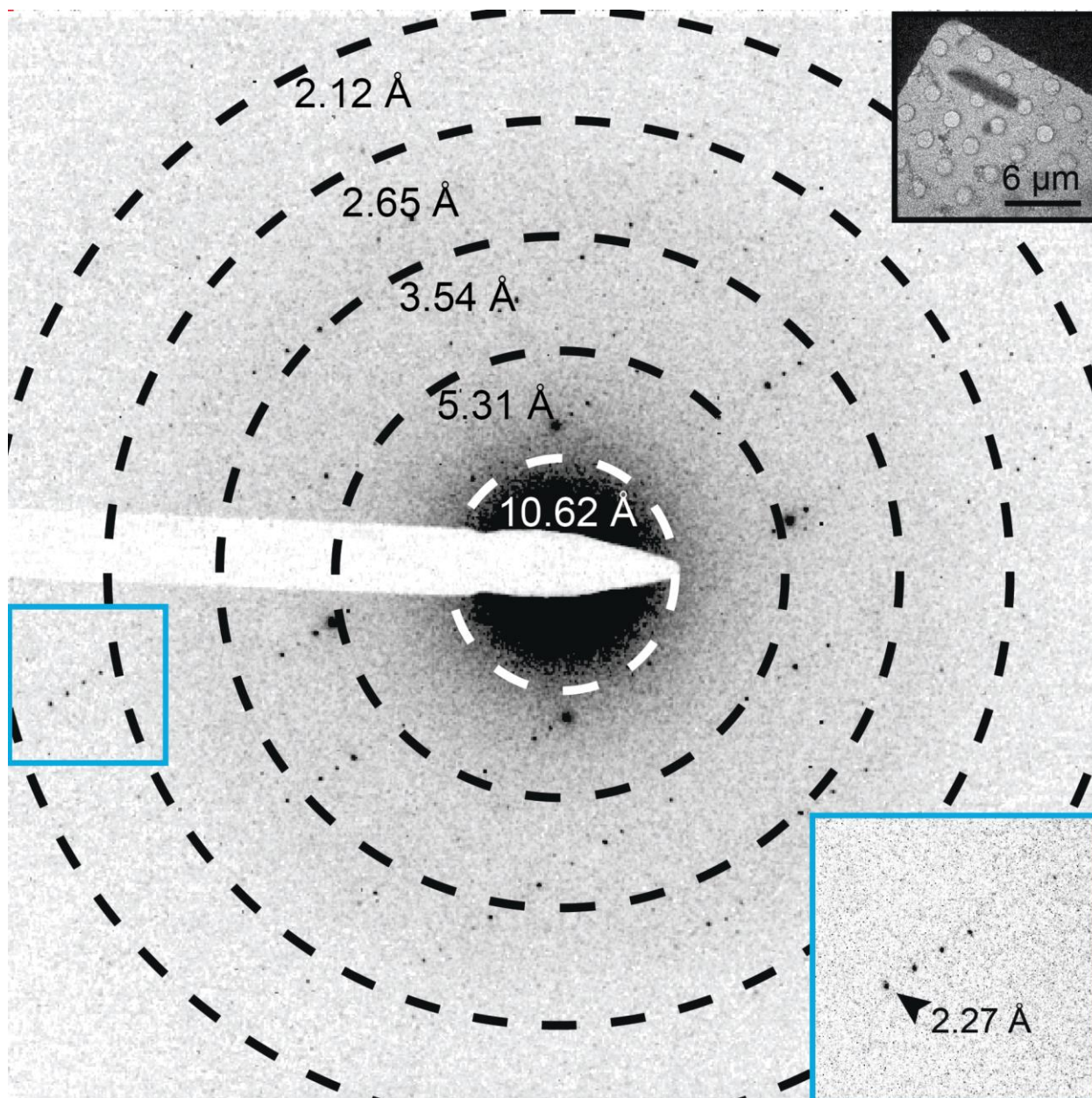
Supplementary Figure 2: Attempts were made to diffract crystals from two conditions: 25% isopropanol, 0.2M NH_4OAc , 0.1M Tris-HCl, pH 8.5, on the left, and 30% PEG 3350, 0.2M NH_4OAc , 0.1M HEPES, pH 7.5, on the right. Diffraction was observed but suffered from poor resolution and spot definition. The data were not of high enough quality to allow for structural elucidation.



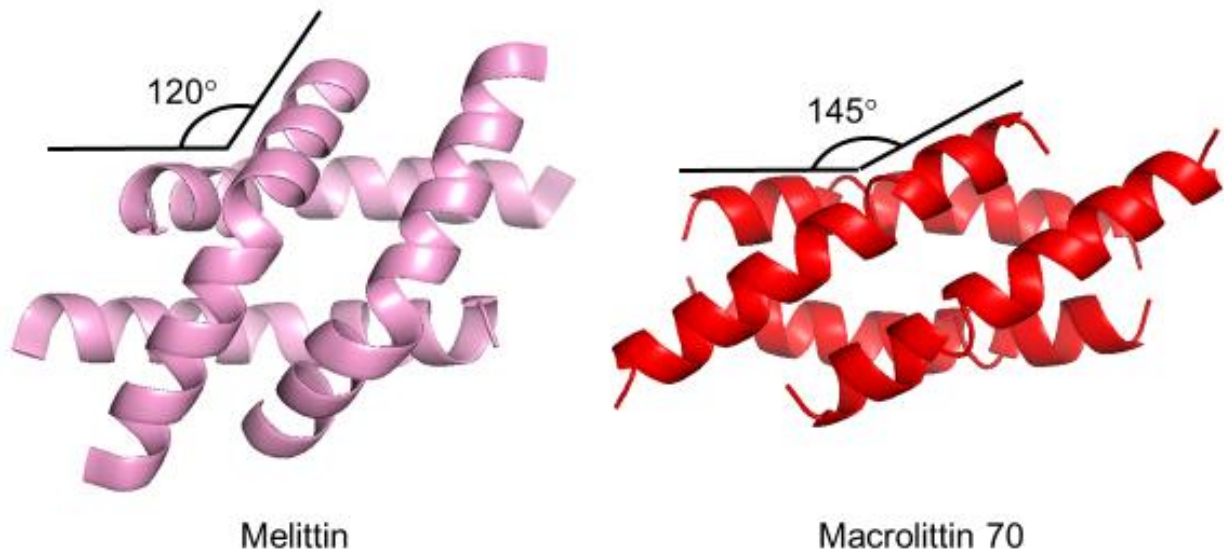
Supplementary Figure 3: Images of 4- μ L hanging drop in 24-well optimization tray. The top two rows are 8 hanging drops that have equilibrated to 16°C. Notice the formation of small needle crystals in almost every well. The bottom two row are the same drops but equilibrated to 22°C after the top images of been taken. The formation of the crystals is not only temperature dependent but also reversible.



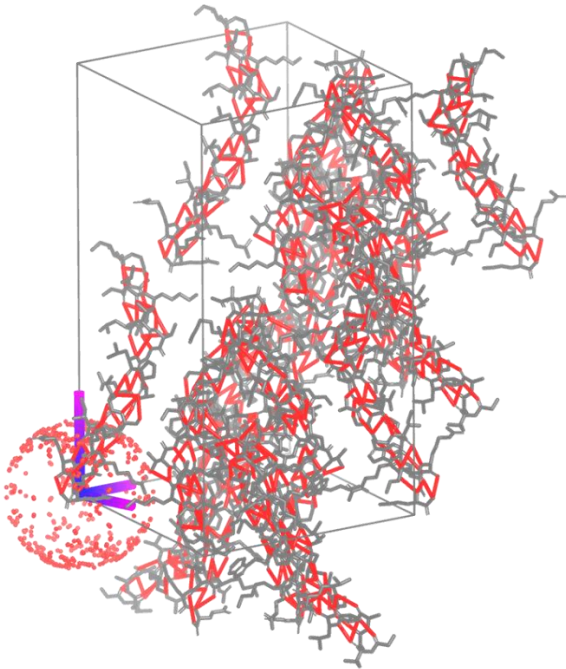
Supplementary Figure 4: Batch crystallization of macrolittin 70 in PEG3350-based crystallization condition. The image on the left is a 0.5- μ L drop aliquoted from the batch crystallization after being left in an ice bath overnight; small needle crystals started to form. The image on the right is a 0.5- μ L drop aliquoted from the same batch crystallization after a week in storage at 4°C.



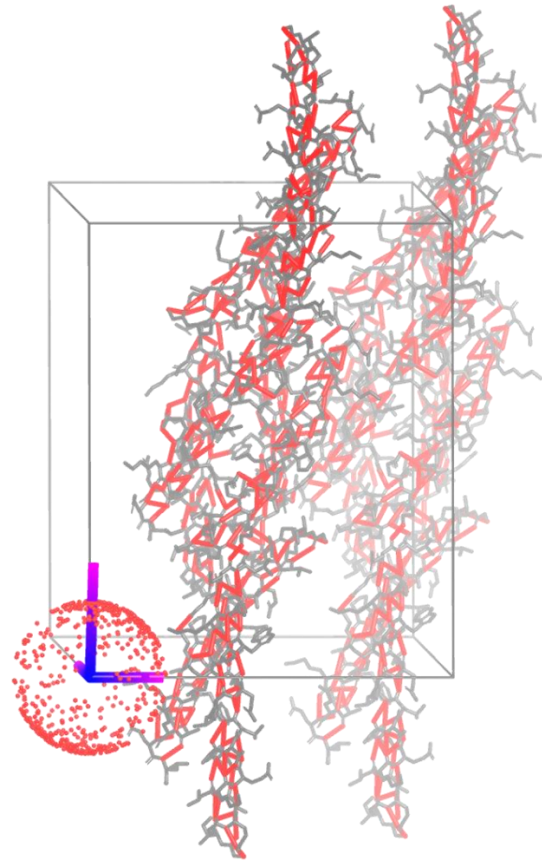
Supplementary Figure 5: Single diffraction pattern of racemic macrolittin 70 measured during continuous-rotation MicroED data collection. Each pattern corresponds to a 0.94° wedge of reciprocal space. The Black inset shows an overfocused diffraction image of the crystal used for diffraction. The blue square correspond to the magnified region (blue inset) of the pattern that show diffraction at 2.27 \AA resolution (black arrow). Resolution circles are indicated by rings; scale bar is 6 \mu m in length.



Supplementary Figure 6: Tetrameric helical bundle of racemic melittin and macrolittin 70. For the racemic melittin crystal structure, the peptide pairs overlap with each other at a much steeper angle than in the macrolittin 70 crystal.



Melittin (6o4m)

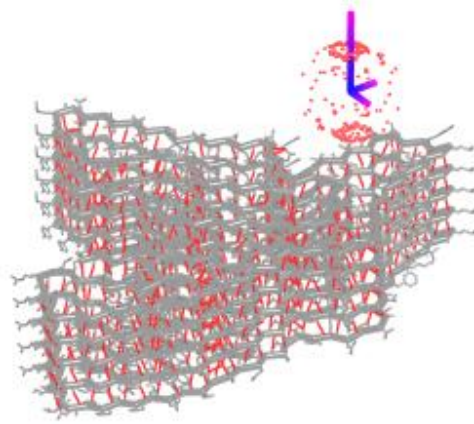


Macrolittin 70

Supplementary Figure 7: Visualization of hydrogen bond vector in the racemic melittin unit cell and the racemic macrolittin 70 unit cell with each hydrogen bond represented as a red lines. The directionality of each hydrogen bond is accounted for as a point on the sphere at the unit cells origin.

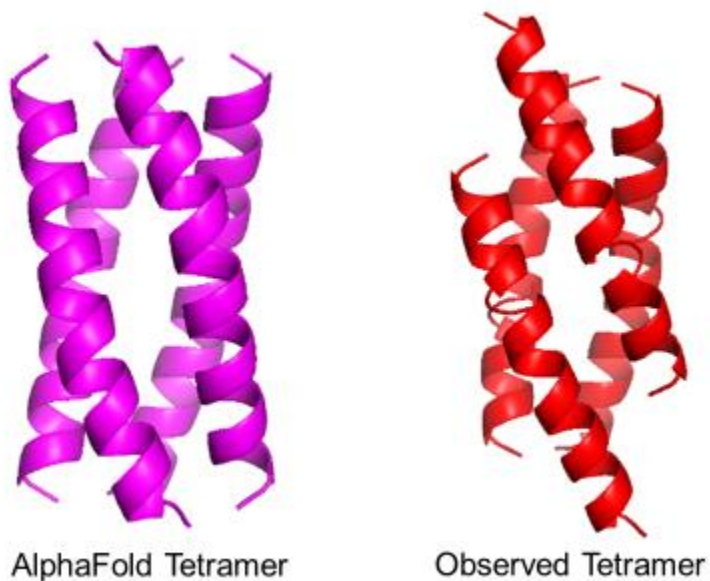


Proteinase K (4zar)

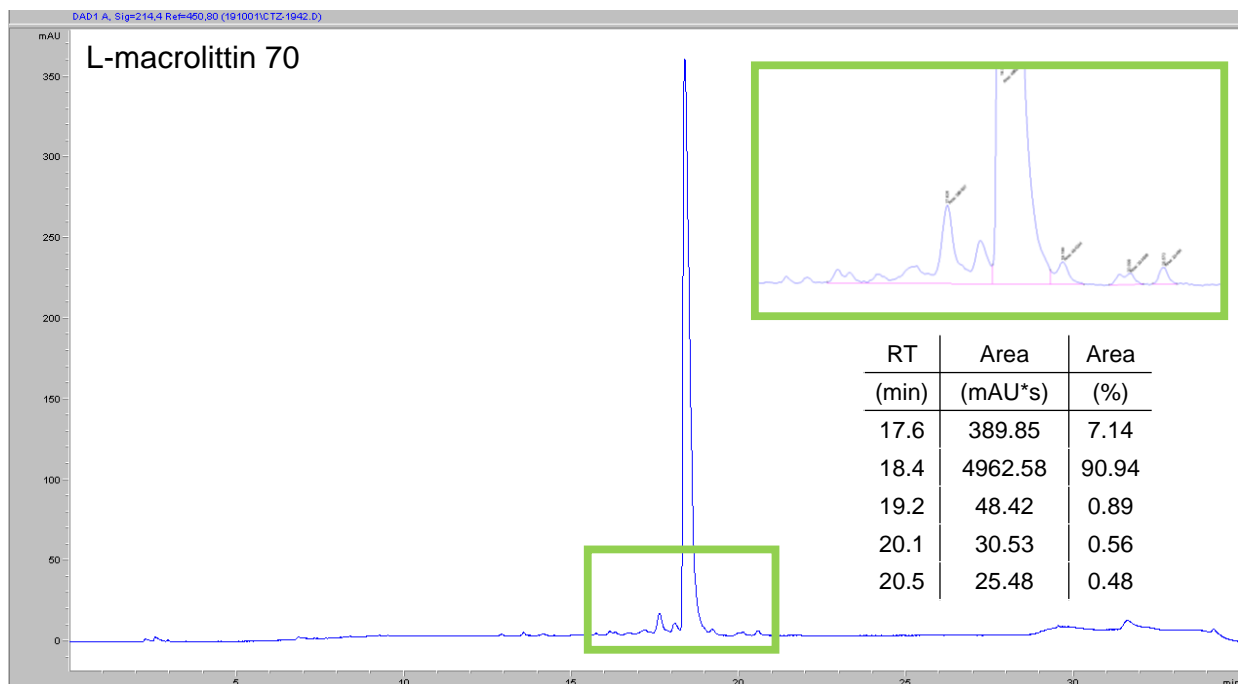


TMEM106B amyloid fibril (7saq)

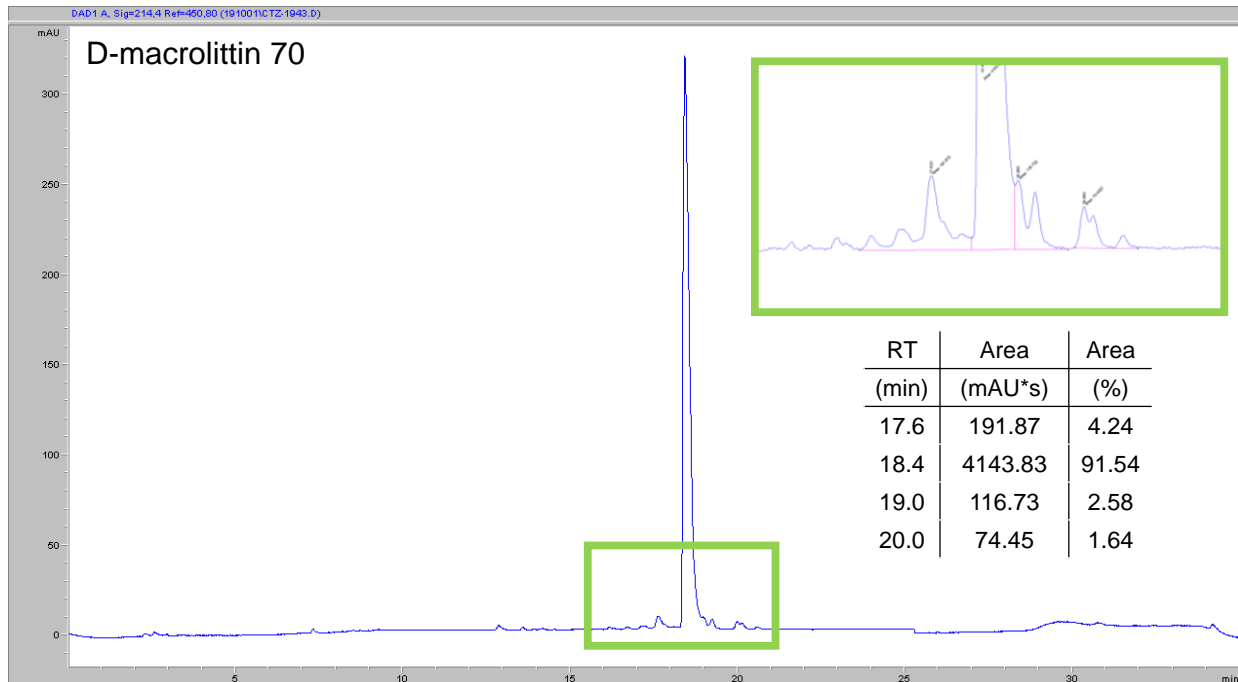
Supplementary Figure 8: Visualization of hydrogen bond vector in the Proteinase K unit cell and in the TMEM106B amyloid fiber structure with each hydrogen bond represented as a red dot. The directionality of each hydrogen bond is accounted for as a point on the sphere at the corner of the unit cell or above the fiber.



Supplementary Figure 9: Side by side comparison of the AlphaFold's predicted tetramer (magenta) with the observed L-macrolittin 70 helical bundle (red) extracted from the crystal lattice. Notice that the helices in the AlphaFold model are in plane at the top and bottom while the heads and tails of each intertwining helix in the macrolittin 70 bundle are offset from one another along the helical rise.



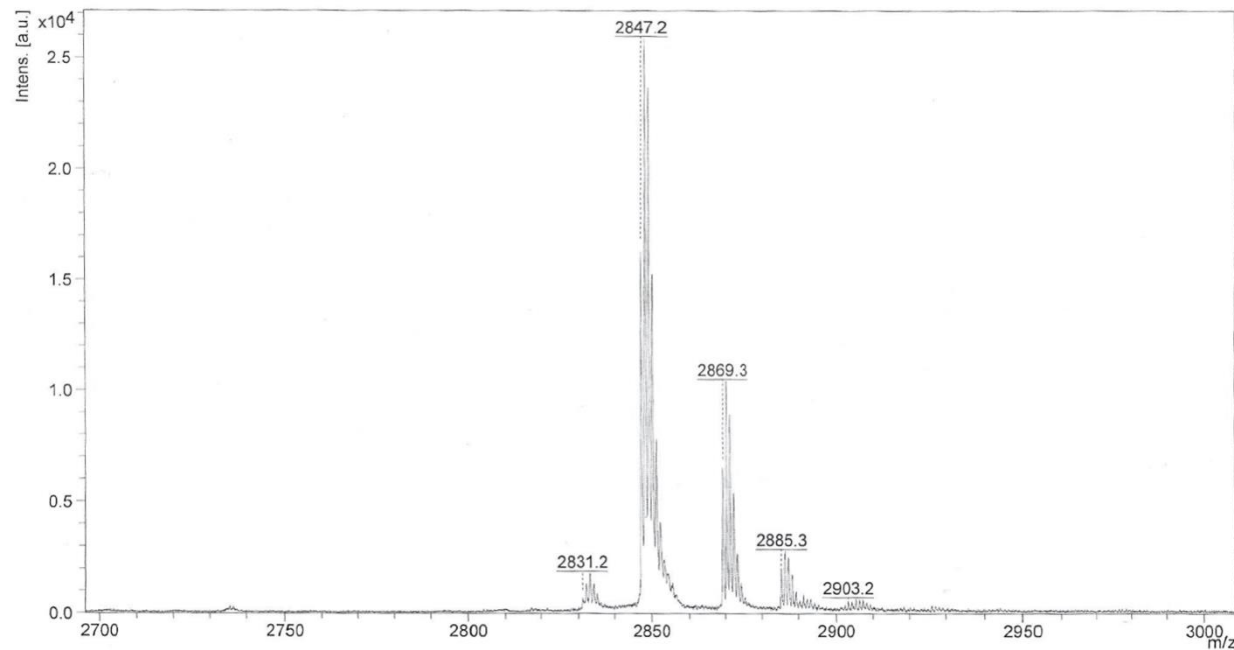
Supplementary Figure 10: Analytical HPLC trace of L-macrolittin 70. The analyte is detected by their absorbance (y-axis, mAU) at 214 nm. The gradient is 35 minutes (x-axis) long and the desired product elutes at 18.4 minutes.



Supplementary Figure 11: Analytical HPLC trace of D-macrolittin 70. The analyte is detected by their absorbance (y-axis, mAU) at 214 nm. The gradient is 35 minutes (x-axis) long and the desired product elutes at 18.4 minutes.

Comment 1 L_Macroli70_V33\CHCA\RP_PepMix.par
Comment 2 400, 25%, 0-5000

printed: 1/25/2021 3:34:30 PM

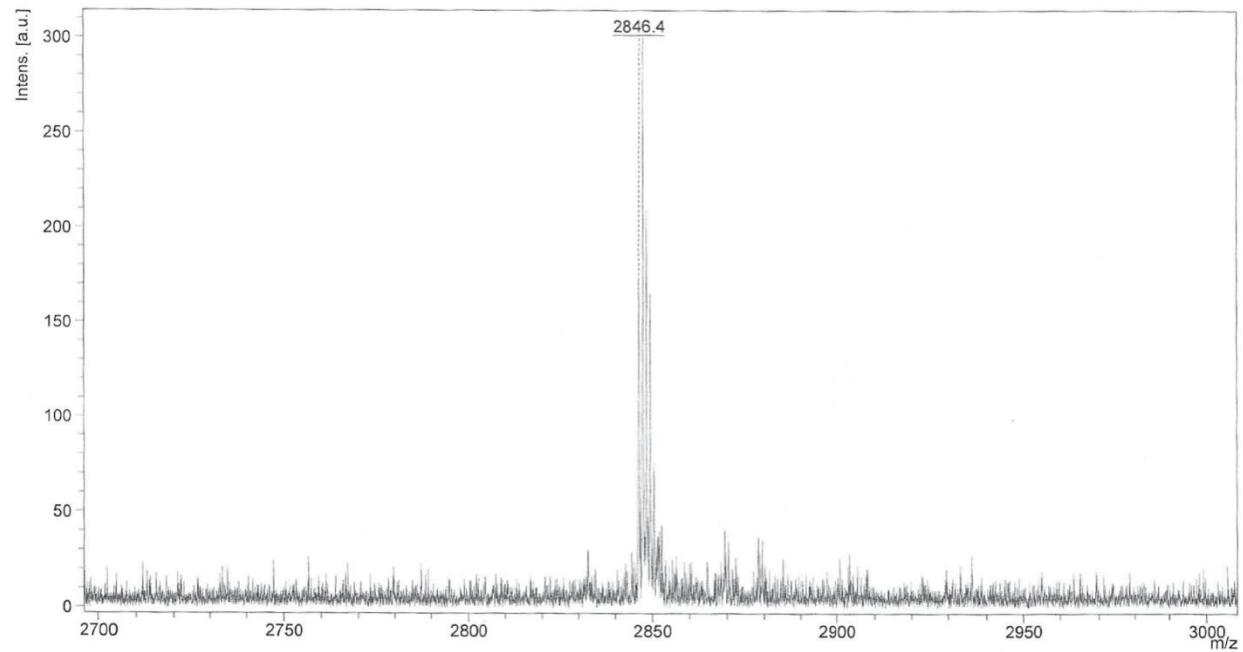


Bruker Ultraflex III, Chemistry Instrumentation Center, University of Wisconsin-Madison, Cite NIH 1S10RR024601-01

Supplementary Figure 12: MALDI-TOF MS analysis of purified L-macroli70. Monoisotopic [M+H]⁺ peaks observed at 2847.2; calculated m/z is 2847.6.

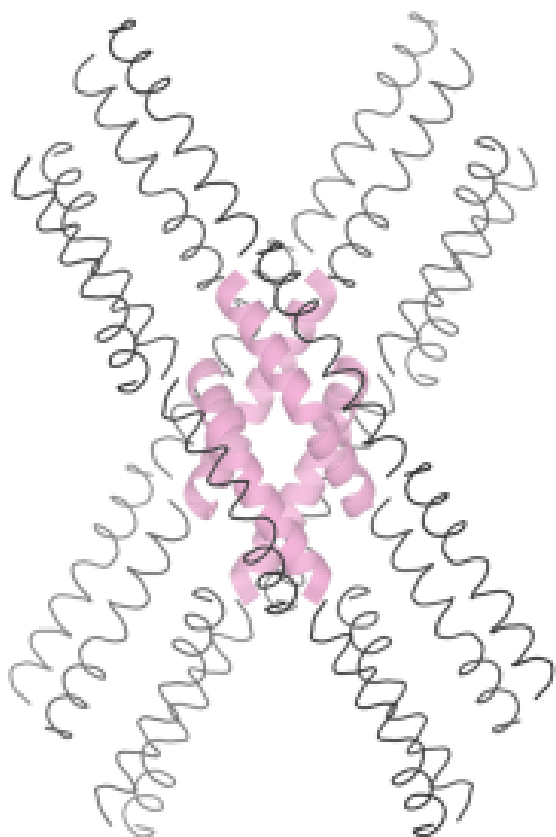
Comment 1 D_Macrolit70_R5/CHCA//RP_PepMix.par
Comment 2 500, 95%, 0-5000

printed: 1/25/2021 3:35:12 PM

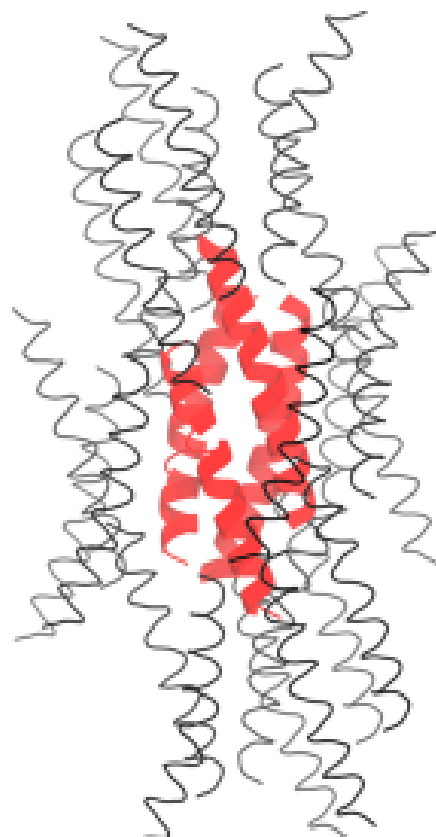


Bruker Ultraflex III, Chemistry Instrumentation Center, University of Wisconsin-Madison, Cite NIH 1S10RR024601-01

Supplementary Figure 13: MALDI-TOF MS analysis of purified D-macrolittin 70. Monoisotopic [M+H]⁺ peaks observed at 2846.4; calculated m/z is 2847.6.



Melittin PISA Input



Macrolittin 70 PISA Input

Supplementary Figure 14: The final expanded lattice that includes all adjacent chains that engages with Chain A, B, C, and D (the helical bundles in the core, colored pink for melittin and colored red for macrolittin 70). The calculated BSA and ΔG values from the two lattice above were sorted, summed, and tabulated in Table S3.

Racemic Macrolittin 70

Data collection		
Beam Type	Electron	
Space group	C2	C2/c
Cell dimensions		
a, b, c (Å)	48.42	30.69 60.49
α , β , γ (°)	90.000	91.879 90.000
Resolution limit (Å)	2.3 (2.48-2.30)	
Wavelength (Å)	0.019687	
No. of crystals merged	4 (6 Integrations)	
R _{merge} (%)	15.0 (47.1)	14.7 (46.4)
I/ σ I	5.26 (2.43)	5.53 (2.51)
CC1/2 (%)	98.9 (88.7)	98.7 (88.9)
Completeness (%)	82.8 (80.8)	82.1 (80.4)
No. reflections	14355 (2570)	13733 (2487)
No. unique reflections	3390 (654)	3184 (624)
Multiplicity	4.23 (3.93)	4.31 (3.99)
Refinement		
Resolution range (Å)	22.72-2.30 (2.48-2.30)	
No. of Reflections	3378	
work	3039 (583)	
free	339 (65)	
R-work	22.09 (33.32)	
R-free	28.28 (42.16)	

Table S1: Values in parentheses are for the highest-resolution shell unless otherwise specified.

Comparison of R_{work} between C2 and C2/c Refinement

Resolution (Å)	C2 R_{work}	Differential	C2 R_{merge}	Differential	C2/c R_{work}	Differential	C2/c R_{merge}	Differential
22.73 to 3.93	17.30%		10.1%		20.50%		9.9%	
3.93 to 3.12	18.30%	-1.00%	13.0%	-2.90%	24.87%	-4.37%	12.8%	-2.90%
3.12 to 2.73	25.64%	-7.34%	22.9%	-9.90%	35.45%	-10.58%	22.5%	-9.70%
2.73 to 2.48	28.35%	-2.71%	36.9%	-14.00%	34.50%	0.95%	36.2%	-13.70%
2.48 to 2.30	33.32%	-4.97%	47.1%	-10.20%	41.82%	-7.32%	46.4%	-10.20%
Overall	22.09%		15.0%		27.85%		14.70%	

Table S2: R_{work} and R_{merge} values, binned by resolution, for refinements executed with the structure assigned as space group C2 or C2/c.

Melittin	Total BSA (Å ²)	Total ΔG _{form} (kcal/M)	Macrolittin 70	Total BSA (Å ²)	Total ΔG _{form} (kcal/M)
L-	1151.1	13	L-	2025.8	-33
D-	1447.6	-10.1	D-	2224.3	-17.4
Bundle	2221.6	-56.9	Bundle	1668.1	-38.9

Table S3: Calculated buried surface area and $\Delta G_{\text{formation}}$ for a 4 chain helical bundle. The values in the row labeled bundle denotes calculated values for the interaction between the chains in the helical bundle. The L- row are the calculated values for the interaction of a core L-helical bundle with adjacent L- helical bundles within the enantiomerically pure sheet. The D- row are the calculated values for the interaction of a core helical bundle with adjacent D- helical bundles.

A1 – 10 mg/mL Unsiliconized Glass Mixed Drop	A2 – 5 mg/mL Unsiliconized Glass Mixed Drop	A3 – 10 mg/mL Siliconized Glass Mixed Drop	A4 – 5 mg/mL Siliconized Glass Mixed Drop
B1 – 10 mg/mL Unsiliconized Glass Unmixed Drop	B2 – 5 mg/mL Unsiliconized Glass Unmixed Drop	B3 – 10 mg/mL Siliconized Glass Unmixed Drop	B4 – 5 mg/mL Siliconized Glass Unmixed Drop

Table S4: Additional optimization executed after X-ray diffraction data collection shown in supplementary figure 2. Position in chart correspond to their relative positions on a 24-well tray (Supplementary Figure 3). All drops are composed of 2 μ L of crystallization reagent and 2 μ L of peptide solution. The concentration denotes the concentration of each enantiomer in the starting peptide solution.

Conclusion

In this work, we executed the concurrent development and usage of racemic crystallography for micro-electron diffraction. We demonstrated its utility when investigating the structural mechanism by which ice-nucleating proteins may nucleate ice for *Pseudomonas syringae*. Also, by elucidating the structure of racemic macrolittin 70, we are one of the first to use micro-electron diffraction to solve a structure with a unit cell size and suboptimal packing similar to that of globular proteins. In 2016, direct methods and molecular replacement were the only published methods for phasing electron diffraction data, but our work has validated other techniques and advanced micro-electron diffraction towards routine usage as a tool for structure determination.

APPENDIX TO THE DISSERTATION

INTRODUCTION

Methods are developed for application, broad or specific. This appendix includes several projects that I have contributed to that highlight the value of crystallography and its associated methods in advancing science for other fields.

Appendix A Preface

Some peptides, derived from amyloidogenic proteins, exhibit such facile nucleation that it precludes the possibility of growing large crystals due to the rapid conversion of the solvated species to their aggregated counterpart. Not only are the crystals small, but these rapidly nucleating species also tend to experience secondary nucleation that result in the observation of secondary lattice(s) (Figure 1).

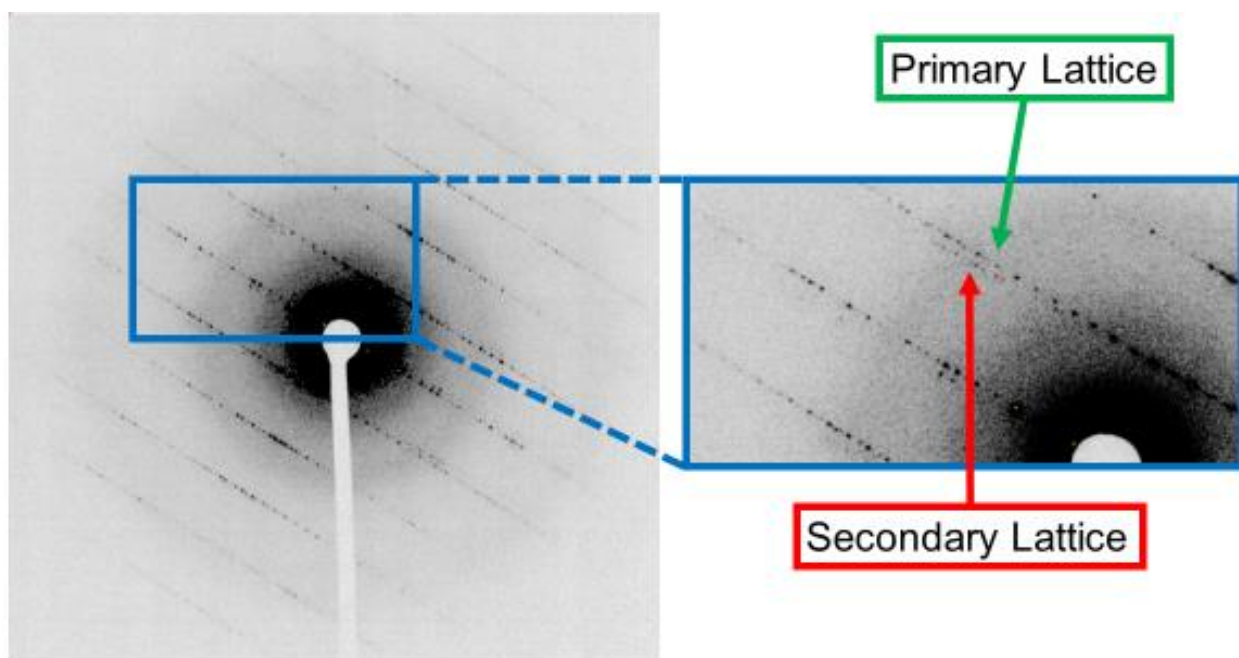


Figure 1: Integrated image of 5 separate diffraction exposures collected via electron diffraction ($\Delta 0.9^\circ$ oscillation per exposure; $\Delta 4.5^\circ$). A secondary lattice was visible in images that most likely resulted from secondary nucleation on the primary crystal needle.

In this work, data from amyloid- β (20-34) and amyloid- β (20-34, isoAsp23) crystals were collected by electron diffraction, and the structures were solved using direct methods. Presently, these structures are not only some of the longest post-aggregation structures solved for amyloidogenic peptides but also some of the few peptide structures, solved by crystallography, that have high agreement with fiber structures that have been solved in recent years¹ (Figure 2).

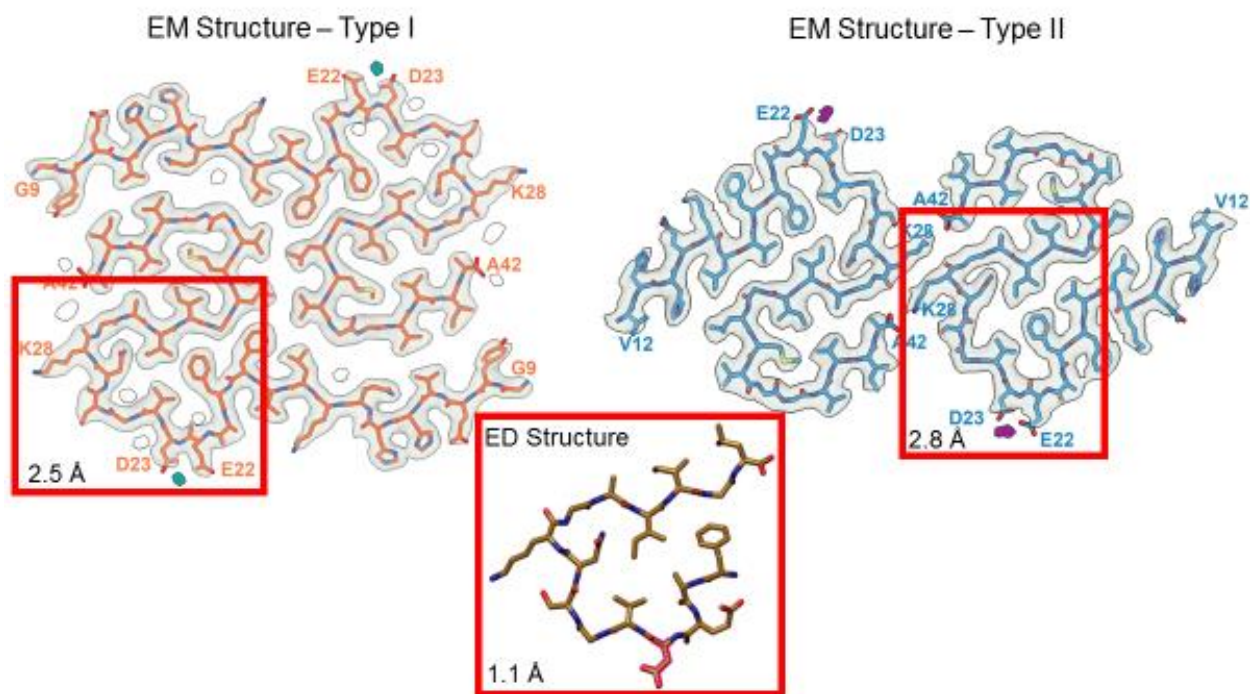


Figure 2: Cryo-EM structures of amyloid- β 42 filaments from human brains (left and right, with resolutions of 2.5 Å and 2.8 Å, respectively). Crystal structure of wild-type structure of amyloid- β (20-34) is shown on the bottom (1.1 Å resolution). The regions corresponding to amino acids 20-34 for the cryo-EM structures are boxed in red for comparison with electron diffraction structure.

References

Yang, Y. *et al.* Cryo-EM structures of amyloid- β 42 filaments from human brains. *Science* **375**, 167-172 (2022).

Appendix A

Structure of amyloid- β (20-34) with Alzheimer's-associated isomerization at Asp23 reveals a distinct protofilament interface

The following is a reprint of a research article from

Nature Communications

10, 3357 (2019)

DOI: 10.1038/s41467-019-11183-z

Reprinted by permission from Springer Nature.

This is an open access article under CC-BY License






(<https://creativecommons.org/licenses/by/4.0/>)

ARTICLE

<https://doi.org/10.1038/s41467-019-11183-z>

OPEN

Structure of amyloid- β (20-34) with Alzheimer's-associated isomerization at Asp23 reveals a distinct protofilament interface

Rebecca A. Warmack ¹, David R. Boyer¹, Chih-Te Zee¹, Logan S. Richards¹, Michael R. Sawaya ^{1,2,3}, Duilio Cascio ^{1,2,3}, Tamir Gonen ^{2,4,5,6}, David S. Eisenberg ^{1,2,3,4,5} & Steven G. Clarke^{1,2}

Amyloid- β (A β) harbors numerous posttranslational modifications (PTMs) that may affect Alzheimer's disease (AD) pathogenesis. Here we present the 1.1 Å resolution MicroED structure of an A β 20-34 fibril with and without the disease-associated PTM, L-isoaspartate, at position 23 (L-isoAsp23). Both wild-type and L-isoAsp23 protofilaments adopt β -helix-like folds with tightly packed cores, resembling the cores of full-length fibrillar A β structures, and both self-associate through two distinct interfaces. One of these is a unique A β interface strengthened by the isoaspartyl modification. Powder diffraction patterns suggest a similar structure may be adopted by protofilaments of an analogous segment containing the heritable lowa mutation, Asp23Asn. Consistent with its early onset phenotype in patients, Asp23Asn accelerates aggregation of A β 20-34, as does the L-isoAsp23 modification. These structures suggest that the enhanced amyloidogenicity of the modified A β segments may also reduce the concentration required to achieve nucleation and therefore help spur the pathogenesis of AD.

¹Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA 90095-1569, USA. ²Molecular Biology Institute, University of California, Los Angeles, Los Angeles, CA 90095-1570, USA. ³UCLA-DOE Institute, University of California, Los Angeles, Los Angeles, CA 90095-1570, USA. ⁴Howard Hughes Medical Institute, University of California, Los Angeles, Los Angeles, CA 90095-1570, USA. ⁵Department of Biological Chemistry, University of California, Los Angeles, Los Angeles, CA 90095-1737, USA. ⁶Department of Physiology, University of California, Los Angeles, Los Angeles, CA 90095-1751, USA. Correspondence and requests for materials should be addressed to S.G.C. (email: clarke@mbi.ucla.edu)

A prevalent theory for the biochemical basis of Alzheimer's disease (AD) is the amyloid cascade hypothesis, which describes the aggregation of the A β peptide into oligomeric or fibrous structures that then trigger the formation of neurotoxic tau neurofibrillary tangles^{1–3}. The A β peptide is subject to a number of posttranslational modifications (PTMs) that may affect its aggregation *in vivo*⁴. Specifically, A β phosphorylation (Ser8, Ser26), pyroglutamylation (Glu3, Glu11), nitration (Tyr10), and racemization/isomerization (Asp1, Asp7, Asp23, Ser26) have been shown *in vitro* to increase the aggregation propensity or neurotoxicity of the A β 1–42 peptide^{5–11}, while other modifications, such as dityrosine crosslinking (Tyr10), have been shown to increase the stability of the A β aggregates¹².

Isomerized products of aspartic acid residues perturb protein structure by rerouting the peptide backbone through the side chain β -carbonyl. This age-dependent modification introduces a methylene group within the polypeptide backbone and thus may have a significant effect on the structure of A β oligomers or fibrils^{13–15}. In addition, the isopeptide bond is resistant to degradation, potentially increasing the concentration of the isomerized A β form with respect to the native peptide. Despite the presence of a repair enzyme in the brain, the L-isospartate (D-aspartate) O-methyltransferase (PCMT1) for L-isospartate, the isomerization of A β Asp1, Asp7, and Asp23 has been identified within AD brain parenchyma^{16,17}. In the cases of the heritable early-onset AD Iowa mutation (Asp23Asn), 25–65% of Asn23 residues have been shown to be isomerized in frontal lobe tissues¹⁸, consistent with the increased rates of spontaneous deamidation/isomerization of asparagine relative to aspartate¹⁹. *In vitro* studies demonstrate that L-isospartate at Asp23 (L-isoAsp23) significantly accelerates A β 1–42 fibril formation, while L-isoAsp7 alone does not^{11,20}. Subsequent studies using peptides with multiple sites of isomerization showed only minor accelerated aggregation of the tri-isomerized species (1, 7, and 23), over the di-isomerized species (7 and 23)¹⁸. Taken together, these results suggest that among the known sites of Asp isomerization in A β , L-isoAsp23 is primarily responsible for the increase in aggregation propensity *in vitro*.

Given the relevance of the isomerization of Asp23 to both sporadic and hereditary Iowa mutant forms of AD, we sought to discover the structural basis for its acceleration of fibril formation^{10,17,18}. As a platform for evaluating this modification, we chose synthetically generated 15-mer peptides encoding residues 20–34 of the A β peptide (A β ^{20–34}) with and without an L-isoAsp modification at position 23 and spanning the core of known A β fibril structures^{21–29}. Challenged by the small size of crystals formed by this segment, we employed the cryo-electron microscopic (cryo-EM) method microcrystal electron diffraction (MicroED) to determine the structures. The structures of A β ^{20–34} and A β ^{20–34, isoAsp23}, determined to 1.1 Å resolution by direct methods, reveal with atomic detail conserved kinked β -helix-like turns with complex features similar to those observed previously at lower resolution in the cores of fibrillar A β 1–42, as well as a distinct pair of protofilament interfaces. Our results suggest that the L-isoAsp23 residue facilitates the formation of a more stable form of this unique interface, promoting enhanced fiber formation and stability. The length of these peptide segments, four residues longer than any other crystallographically determined amyloid structures^{30–33}, is key in facilitating their complex fold—a conformation more representative of the full-length A β fibrils.

Results

Fibril formation and characterization of A β ^{20–34} peptides. Six early-onset hereditary Alzheimer's mutations and two PTMs, including the isomerized Asp23, are localized in the A β 1–42

peptide to a region spanning six residues from Ala21 to Ser26 near the center of the peptide (Fig. 1a, b)^{17,34,35}. The amyloid-forming propensity of segments in this region of A β was assessed using a computational method of predicting steric zippers by a threading protocol (ZipperDB³⁶). This method highlights a region of A β from Asn27 to Gly37 with high aggregation propensity near the site of Asp23 isomerization (Fig. 1b). To characterize segments containing an isomerized Asp residue at position 23, we utilized synthetic 15 residue peptides spanning the A β residues 20–34 (A β ^{20–34}) in which Asp23 was substituted with either an L-Asn residue (Iowa mutant; A β ^{20–34, Asp23Asn}) or an L-isoAsp residue (A β ^{20–34, isoAsp23}).

To evaluate the effect of these variations on this 15-residue segment of A β , we assayed its capacity to form fibrils as measured by light scattering at 340 nm (Fig. 1c). Both the peptide based on the Iowa mutant (A β ^{20–34, Asp23Asn}) and the peptide based on L-isoAsp23 (A β ^{20–34, isoAsp23}) demonstrated significantly enhanced fibril formation over that of A β ^{20–34}, with the Iowa mutant peptide displaying the fastest initial rate of fibril formation (Fig. 1c). Fibers of the native peptide at this concentration (1.6 mM) were not observed by light scattering or EM. We further discovered that only 34% of these A β ^{20–34, isoAsp23} aggregates could be methylated by the L-isoAsp repair protein carboxyl methyltransferase (PCMT1) *in vitro* (Supplementary Fig. 1). These data suggest that a majority of the L-isoAsp sites are occluded from the normal repair pathway once in this aggregate form.

To determine the ability of these modified forms to accelerate the aggregation of native peptide, seeding of 3.2 mM A β ^{20–34} was performed using 10 μ M final concentrations of pre-aggregated seeds of A β ^{20–34}, A β ^{20–34, Asp23Asn}, and A β ^{20–34, isoAsp23} (Fig. 1d). The addition of each of the preformed aggregates caused significant acceleration in the onset of fiber formation. The largest shift occurred with the native A β ^{20–34} seed, followed by isomerized A β ^{20–34, isoAsp23} and A β ^{20–34, Asp23Asn}. Powder diffraction performed on the final aggregates revealed nearly identical sets of reflections, suggesting that the three seeds have similar enough structures to template wild-type (WT) A β ^{20–34} aggregates whose diffraction resembles unseeded fibrils (Supplementary Fig. 2a). Fibrillization experiments of full-length A β 1–40 with and without the L-isoAsp modification at residue 23 reveal that the isomerized species displays a shorter lag time, consistent with the results obtained with the corresponding A β ^{20–34} peptides (Supplementary Fig. 2b; Fig. 1c). Thus, while the isomerized form may be only a minor component of the *in vivo* A β population, it aggregates at a faster rate and can cross-seed the native form efficiently *in vitro*.

In contrast to the results obtained with 1.6 mM A β ^{20–34}, increasing the concentration to 3.2 mM A β ^{20–34} did yield aggregates ~77 nm in width (Fig. 1c, d). Importantly, light scattering under these conditions for this native peptide is not detected until 3.5 h at the earliest, while shifts in light scattering for the 1.6 mM isomerized and mutated peptides were detected by 1.5 and 0.5 h, respectively (Fig. 1d). Direct comparisons of formation rates were complicated by the insolubility of the A β ^{20–34, Asp23Asn} peptide at high concentrations, but the delayed onset of even the 3.2 mM A β ^{20–34} incubation compared to the 1.6 mM A β ^{20–34, Asp23Asn} and A β ^{20–34, isoAsp23} incubations also support the increased rates of aggregation of the mutated and isomerized peptides (Fig. 1c, d).

Fibrils of each segment were also investigated for their resistance to dissociation by dilution into increasing concentrations of sodium dodecyl sulfate (SDS) at 70 °C as measured by light scattering at 340 nm (Fig. 2). Fibrils of the native 15-residue A β segment appeared to partially dissolve upon dilution into the SDS-free buffer, although remaining aggregates were found by

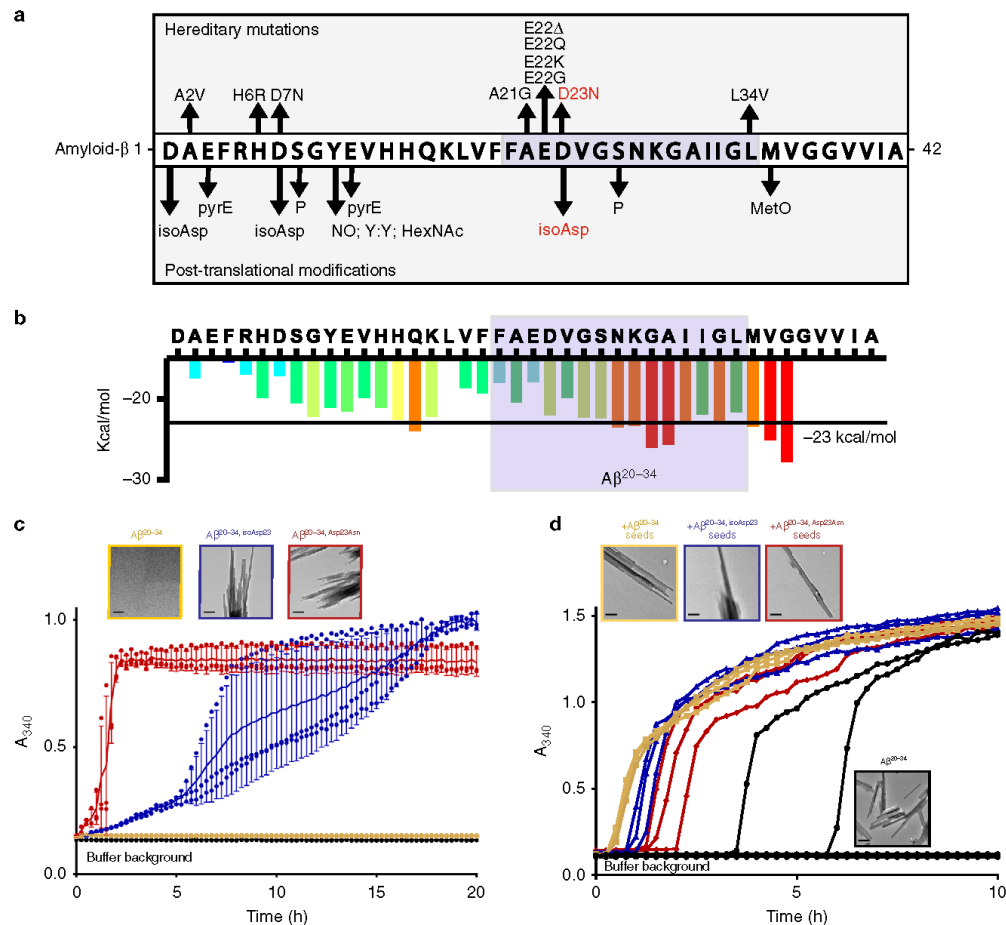


Fig. 1 L-isoAsp in A β ²⁰⁻³⁴ accelerates fiber formation and can seed native segment. **a** Sequence of human A β including known early-onset hereditary mutations and posttranslational modifications (pyrE pyroglutamate, P phosphorylation, NO nitration, Y:Y dityrosine crosslink, HexNAc glycosylation, MetO oxidation). **b** ZipperDB²² amyloid propensity profile for the human A β sequence with the A β ²⁰⁻³⁴ sequence highlighted in light blue. **c** 1.6 mM of A β ²⁰⁻³⁴, A β ²⁰⁻³⁴, Asp23Asn, and A β ²⁰⁻³⁴, isoAsp23 peptide aggregation was monitored by turbidity at 340 nm. Each data point is shown as a round symbol, the solid line represents the mean value, and error bars represent SD of three replicates. Transmission electron micrographs of aggregates are shown at the top left of the graph, scale bars represent 0.5 μ m in each image. **d** Aggregation of 3.2 mM A β ²⁰⁻³⁴ in 50 mM Tris, pH 7.5, 150 mM NaCl, and 1% dimethyl sulfoxide was monitored by turbidity at 340 nm alone (black lines) or with 10 μ M pre-aggregated seeds of A β ²⁰⁻³⁴ (yellow), A β ²⁰⁻³⁴, isoAsp23 (blue), and A β ²⁰⁻³⁴, Asp23Asn (red). Each line represents a replicate well. Transmission electron micrographs of aggregates are shown at the top left of the graph; scale bars shown at the lower left represent 0.5 μ m in each image. Source data are provided as a Source Data file

EM. However, these were completely dissolved upon incubation with 1% SDS and higher concentrations (Fig. 2). In contrast, the isomerized peptide showed increased resistance to dissolution compared to the native peptide and still showed light scattering at a concentration of 2% SDS, though no more aggregates were seen at 5% SDS (Fig. 2b). The fibrils of the Iowa mutant appeared to be largely unaffected by dilution even at the highest concentrations of SDS, with no significant changes observed in the levels of light scattering. However, the aggregates in 5% SDS seen by EM appeared to be less bundled than at lower concentrations (Fig. 2b). These results show that alterations of the structure at Asp23 strongly contribute to fibril formation and stability.

Crystallization and data collection of the A β ²⁰⁻³⁴ segments. To understand the atomic structural basis for changes in the properties of the isomerized peptide, we sought to crystallize it in the amyloid state. Vapor diffusion screening yielded no crystals large enough for analysis by conventional X-ray crystallography for either the A β ²⁰⁻³⁴ or the A β ²⁰⁻³⁴, isoAsp23 segment. Instead ordered nanocrystals of the native segment were obtained with continuous shaking at 1200 rpm, and ordered nanocrystals of the isomerized segment were generated with constant mixing using an acoustic resonant shaker^{37,38} for analysis by MicroED^{39,40} as described in the “Methods” section. Nanocrystals obtained in varying buffer conditions were evaluated by morphology and diffraction via light and EM, respectively. Those formed under

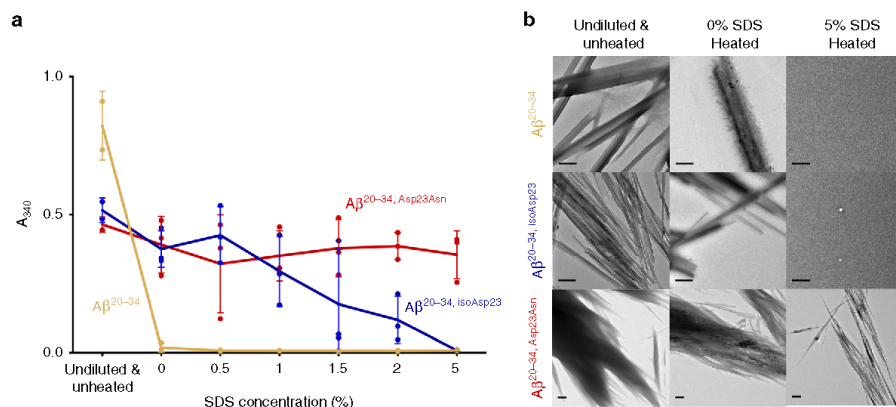


Fig. 2 Modified fibers have increased resistance to sodium dodecyl sulfate (SDS) disaggregation. **a** Fiber stocks (Undiluted & Unheated initial points are two readings of the fiber stocks) were mixed 1:1 in buffer (0% SDS final) and increasing concentrations of SDS (1, 1.5, 2, 5% final) as described in the “Methods” section. Each data point is shown as a round symbol, the solid line represents the mean value, and error bars represent the SD of three technical replicates. **b** Transmission electron micrographs of disaggregated fibers, scale bars in the lower left represent 0.5 μm . Source data are provided as a Source Data file

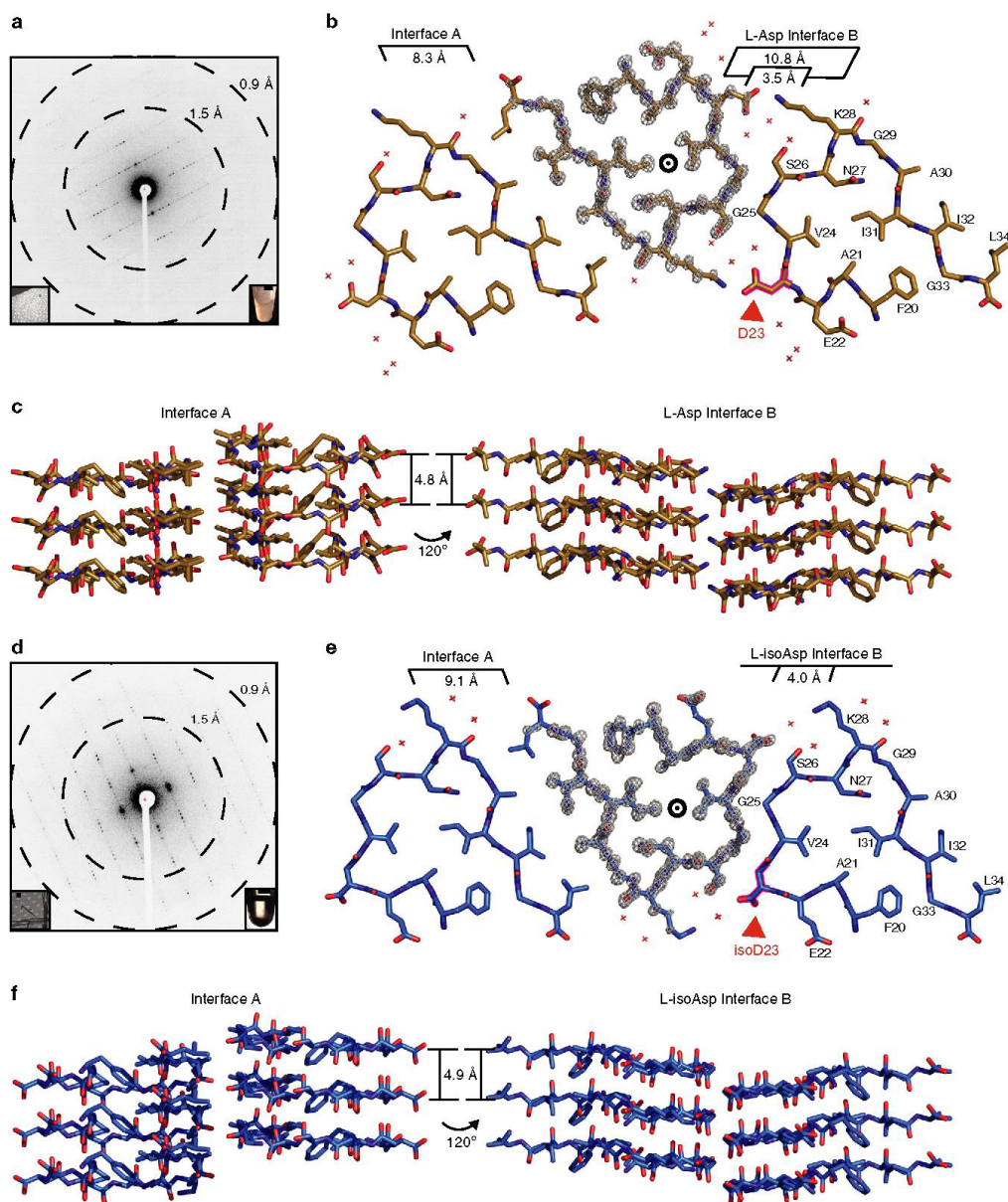
the most promising conditions were used as seeds for additional rounds of batch crystal formation. The optimal crystallization condition for the isomerized segment was 50 mM Tris, pH 7.6, 150 mM NaCl, and 1% dimethyl sulfoxide (DMSO) for 48 h with 2% seeds. Crystals of the native segment grew in 50 mM Tris, pH 7.5, 150 mM NaCl, and 1% DMSO for 30 h without seeding. Isomerized crystal trials produced densely bundled nanocrystals that could not be disaggregated by sonication and freeze–thawing. However, washing crystal solutions with a 0.75% (*w/v*) solution of β -octyl glucoside in TBS, pH 7.6 yielded a higher number of single crystals for subsequent data collection. Dilution one to one in buffer yielded sufficient single crystals of the native segment for data collection (Fig. 3a, d). Data were collected on a Thermo Fisher TALOS Arctica microscope operating at 200 kV using a bottom mount CetaD CMOS detector. Each $\text{A}\beta^{20-34}$ nanocrystal could be rotated continuously up to 140 degrees during data collection. A 1.1- \AA -resolution structure was obtained by direct methods for each segment as described in the “Methods” section; refinement statistics for the structures are shown in Table 1.

MicroED structures of $\text{A}\beta^{20-34}$ and $\text{A}\beta^{20-34, \text{isoAsp}23}$ segments. The structures of both the $\text{A}\beta^{20-34}$ and the $\text{A}\beta^{20-34, \text{isoAsp}23}$ protofilaments reveal parallel, in-register architectures in which individual peptide chains stack through backbone hydrogen bonds every 4.8 and 4.9 \AA along the protofilament axis, respectively (Fig. 3). In cross-section, both protofilaments appear triangular owing to sharp turns (β -arches) at Gly25 and Gly29, which divide each chain into three short, straight segments (Fig. 3b, e and Supplementary Fig. 3a). When compared with the structures in the protein databank, the three-sided $\text{A}\beta^{20-34, \text{isoAsp}23}$ structure aligns best with a β -helical antifreeze protein from *Marinomonas primoryensis* but lacks linker regions between each stacked chain. We thus designate this amyloid motif as a β -helix-like turn (Supplementary Fig. 4)^{41,42}. At the central core of both $\text{A}\beta^{20-34}$ and the $\text{A}\beta^{20-34, \text{isoAsp}23}$ protofilaments are the buried side chains of Phe20, Ala21, Val24, Asn27, and Ile31 in a zipper-like “intraface” that is completely dry. The side chain of Asn27 further stabilizes the assembly by forming a ladder of hydrogen bonds (polar zipper) along the length of the protofilament⁴³ (Supplementary Fig. 3b).

Each protofilament self-associates with neighboring protofilaments in the crystals through two distinct interfaces. Interface A in both structures resembles a canonical steric zipper—with intersheet distances of 8.3 and 9.1 \AA for the native and isomerized, respectively (Fig. 3b, e). Both are lined by the hydrophobic side chains of Ala30, Ile32, and Leu34 that are related by 2_1 screw symmetry (steric zipper symmetry class 1⁴⁴). Interface A is completely dry owing to a high S_c of 0.73 in the native and 0.62 in the isomerized. This interface buries approximately 130 \AA^2 per chain in the native form and 131 \AA^2 in the isomerized form.

Unlike the dry steric zipper interface A, six water molecules line the second $\text{A}\beta^{20-34}$ interface, which we designate the “L-Asp Interface B” (Fig. 3b, e). Here the protofilaments are also related by a two-fold screw symmetry axis. Nearest this central axis, Gly25 and Ser26 contact their symmetry partners across the interface, separated by only 3.5 \AA . Furthest from the axis, Asp23 and Lys28 from opposing protofilaments form charged pairs. In between each of these two regions is a solvent channel with the three ordered waters, yielding low shape complementarity ($S_c = 0.43$) to this interface overall. In contrast, in the $\text{A}\beta^{20-34, \text{isoAsp}23}$ “L-isoAsp interface B” the truncated side chain of the L-isoAsp23 residue no longer forms a charged pair with Lys28 and instead the isomerized protofilaments form a completely dry interface containing the methylene group of L-isoAsp23, Val24, Gly25, and Ser26 with high surface complementarity ($S_c = 0.81$; Fig. 3e). This interface is tightly mated over its entire surface with an average distance of 4.0 \AA between the backbones. Interface B buries approximately 139 and 122 \AA^2 per chain for the native and isomerized forms, respectively. The exclusion of water molecules from the L-isoAsp interface B likely results in a favorable gain in entropy for the structure, and there are attractive van der Waals forces along the tightly mated residues L-isoAsp23–Ser26.

Powder diffraction studies of $\text{A}\beta$ peptides. X-ray powder diffraction (XRD) patterns revealed that the fibrils of $\text{A}\beta^{20-34}$ segments appear largely isomorphous, sharing major reflections at $\sim 4.7, 10, 12.2, 14,$ and $29\text{--}31$ \AA (Fig. 4a, b). The similarity among the powder diffraction patterns of $\text{A}\beta^{20-34, \text{isoAsp}23}, \text{A}\beta^{20-34, \text{Asp}23\text{Asn}}$, and $\text{A}\beta^{20-34}$ indicates that $\text{A}\beta^{20-34, \text{Asp}23\text{Asn}}$ mimics the



structures of the native and isomerized segments. We modeled an L-Asn residue at position 23 of the $\text{A}\beta^{20-34}$, isoAsp^{23} structure to see if the native L-amino acid could be accommodated in the dry L-isoAsp interface B (Fig. 4c, right panel). The L-Asn residue was integrated into the $\text{A}\beta^{20-34}$, isoAsp^{23} interface B scaffold without significant clashes. However, this Asn model lacks a backbone hydrogen bond extending between the isoAsp23 amide carboxyl to the Val24 amide nitrogen of the adjacent protofilament that is

present in our $\text{A}\beta^{20-34}$, isoAsp^{23} structure (Fig. 4c). The residue at site 23 has to adopt an allowed, but unusual left-handed helical conformation to form the L-isoAsp interface B. Both the methylene of the isoAsp residue and the isoAsp23 to Val24 main chain hydrogen bond may help stabilize this structure. This backbone hydrogen bond is present in the native $\text{A}\beta^{20-34}$ structure (Fig. 4c). In this native structure, the Asp main chain adopts a more canonical β -sheet conformation, but the side chain

Fig. 3 $A\beta^{20-34}$, $isoAsp^{23}$ structure contains an altered protofilament interface. **a** Representative single crystal electron diffraction pattern of $A\beta^{20-34}$ with resolution rings obtained during microcrystal electron diffraction (MicroED) data collection. Left inset shows the diffracting crystal (lower left scale bar represents 1 μm). Right inset shows light microscopic image of microcrystal sediment in a 1.6-mL microfuge tube (scale bar represents 3 mm). **b** One layer of the $A\beta^{20-34}$ crystal structure viewed down the fibril axis highlighting two distinct steric zipper interfaces. Interface distances are labeled (center circle indicates fibril axis). Waters are represented by red crosses. The $2F_o - F_c$ density is shown as a gray mesh at 2σ on the center protofilament. The Asp23 residue is outlined in magenta and shown by the red arrow. **c** Three layers of the $A\beta^{20-34}$ structure viewed perpendicular to the fibril axis (indicated by arrows; Left panel—interface A, right panel—L-Asp interface B). **d** Representative single crystal electron diffraction pattern of $A\beta^{20-34}$, $isoAsp^{23}$ with resolution rings obtained during MicroED data collection. Left inset shows the diffracting crystal (lower left scale bar represents 1 μm). Right inset shows a light microscopic image of microcrystal sediment in a 1.6-mL microfuge tube (scale bar represents 3 mm). **e** One layer of the $A\beta^{20-34}$, $isoAsp^{23}$ crystal structure viewed down the fibril axis highlighting two distinct steric zipper interfaces. Interface distances are labeled (center circle indicates fibril axis). Waters are represented by red crosses. The $2F_o - F_c$ density is shown as a gray mesh at 2σ on the center protofilament. The L- $isoAsp^{23}$ residue is outlined in magenta and shown by the red arrow. **f**, Three layers of the $A\beta^{20-34}$, $isoAsp^{23}$ structure viewed perpendicular to the fibril axis (indicated by arrows; Left panel—interface A, right panel—L- $isoAsp$ interface B)

Table 1 Data collection and refinement statistics		
	$A\beta^{20-34}$	$A\beta^{20-34}$, $isoAsp^{23}$
Data collection		
Space group	P2 ₁	P2 ₁
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	33.17, 4.78, 30.33	29.20, 4.87, 32.44
α , β , γ (°)	90.00, 111.10, 90.00	90.00, 101.90, 90.00
Resolution (Å)	1.10 (1.13–1.10) ^a	1.05 (1.20–1.05) ^{b,c}
<i>R</i> _{sym} or <i>R</i> _{merge} (%)	18.9	19.7
<i>I</i> / <i>σI</i>	5.41 (3.28)	3.76 (1.38)
Completeness (%)	85.2	82.7 (53.0)
Redundancy	6.67 (6.14)	4.19 (3.10)
Refinement		
Resolution (Å)	7.74–1.10 (1.26–1.10)	5.96–1.05 (1.20–1.05)
No. of reflections	3544 (1141)	3943 (1167)
<i>R</i> _{work} / <i>R</i> _{free} (%)	19.4/21.3 (21.3/26.9)	19.7/24.6 (27.0/32.4)
No. of atoms		
Protein	210	204
Ligand/ion	0	0
Water	7	4
B-factors		
Protein	6.50	8.29
Ligand/ion	—	—
Water	20.78	27.70
R.m.s. deviations		
Bond lengths (Å)	0.56	1.04
Bond angles (°)	0.68	0.90

^aTen crystals were used in determining the $A\beta^{20-34}$ structure
^bFive crystals were used in determining the $A\beta^{20-34}$, $isoAsp^{23}$ structure
^cValues in parentheses are for the highest-resolution shell

protrudes toward the opposite protofilament, prohibiting a tight, dry interface along residues Asp23–Ser26 as in the L- $isoAsp$ interface B.

The L-Asn side chain in the L- $isoAsp$ interface B model may be able to compensate for the loss of this dry interface by forming another ladder of hydrogen bonds along the protofilament axis (Fig. 4c, d right panel). Thus this second interface packing may be achievable for a $A\beta^{20-34}$, $Asp^{23}Asn$ structure as shown in the L- $isoAsp$ interface B model; however, the XRD patterns reveal that the native $A\beta^{20-34}$ and mutated $A\beta^{20-34}$, $Asp^{23}Asn$ peptides share more similarities than the isomerized $A\beta^{20-34}$, $isoAsp^{23}$ and the $A\beta^{20-34}$, $Asp^{23}Asn$ peptide. Both the native and heritable Iowa mutant forms lack more defined peaks at 22.9, 24.7, 29.4, and 32.5, while both have more broad peaks at 30.9 Å (Fig. 4a, b). These similarities between the $A\beta^{20-34}$ and $A\beta^{20-34}$, $Asp^{23}Asn$ fiber diffractions patterns, and the lack of a methylene group in the normal L-residues, may suggest that the Iowa mutant

$A\beta^{20-34}$, $Asp^{23}Asn$ peptide will assume a structure more similar to the native $A\beta^{20-34}$ structure, as modeled in Fig. 4c, d (left panels). This model maintains the backbone hydrogen bond between Asn23 and Val24, the ordered core of the $A\beta^{20-34}$ structure, and allows for the additional polar zipper between stacked Asn23 residues. The added network of hydrogen bonds along the asparagine side chain may explain in part the increased fiber formation rates and stability of $A\beta^{20-34}$, $Asp^{23}Asn$ against SDS and heat denaturation. While the isomorphous powder diffraction patterns seen between $A\beta^{20-34}$, $isoAsp^{23}$, $A\beta^{20-34}$, $Asp^{23}Asn$, and $A\beta^{20-34}$ do support the models in which $A\beta^{20-34}$, $Asp^{23}Asn$ mimics the native and isomerized structures, it cannot be ruled out that $A\beta^{20-34}$, $Asp^{23}Asn$ forms a distinct structure, perhaps lacking either the L- Asp or the L- $isoAsp$ novel interface B, with the ordered core simply stabilized further by the Asn polar ladder.

Importantly, the powder diffraction of full-length $A\beta$ and the shorter peptide segments all display cross- β patterns with strong reflections at ~ 4.7 and 9–10 Å (Fig. 4a), and the crystal structures of $A\beta^{20-34}$ and $A\beta^{20-34}$, $isoAsp^{23}$ form parallel, in-register beta-sheets similar to other full-length $A\beta$ structures. Thus we hypothesized that the $A\beta^{20-34}$, $isoAsp^{23}$ structure could form the core of a distinct isomerized $A\beta$ polymorph. To visualize a potential full-length fiber with the $A\beta^{20-34}$, $isoAsp^{23}$ structure as its core, we added the remaining residues of $A\beta$ 1–42 onto the ends of the $A\beta^{20-34}$, $isoAsp^{23}$ protofilaments and energy minimized the entire model as described in the “Methods” section. The resulting model demonstrates that the remainder of the residues of $A\beta$ 1–42 can be accommodated in a favorable conformation with the isomerized segment as a core with interface A or B as the primary interface (Fig. 5).

Comparison of segment structures to known $A\beta$ structures.

The structures presented here are the longest segments of an amyloid peptide determined by crystallography—four residues longer than the previous amyloid spines determined by MicroED^{30–33}. This extension is significant due to the fact that, as the number of residues in a segment grows, the packing of idealized β -strands in a lattice becomes more difficult owing to the strain created by the natural twist of the β -sheet/strand. This strain hypothesis is consistent with observations that, as the number of residues in an amyloid segment grows, the crystals that can be grown are correspondingly smaller⁴⁵. In the literature to date, the crystal structures of shorter segments of amyloid proteins have revealed that the dominant forces stabilizing protofilaments occur between different peptide chains⁴⁶. In the native and modified $A\beta^{20-34}$ structures, we are not only able to see interactions between protofilaments, such as the interfaces A and B, but we also see folding of the peptide to produce a β -helix-like

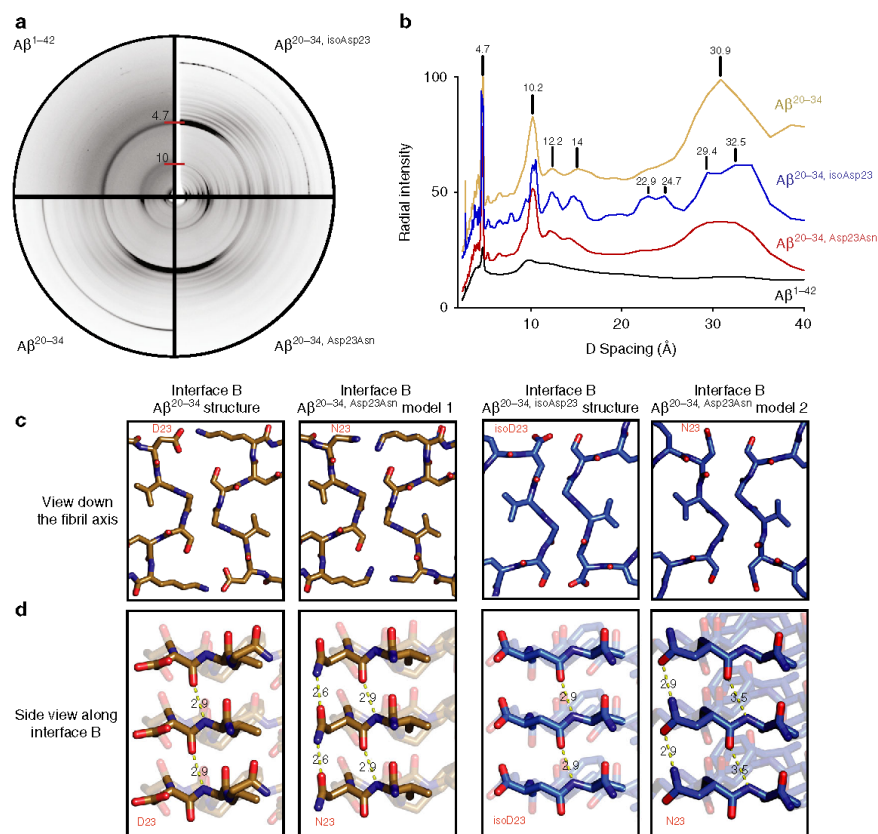


Fig. 4 A putative model of the heritable Iowa mutation in interface B. **a** Fiber diffraction patterns of A β 1–42, A β 20–34, A β 20–34, isoAsp23, and A β 20–34, Asp23Asn. All fibers including A β 1–42 were prepared in 50 mM Tris, pH 7.6, 150 mM NaCl, and 1% dimethyl sulfoxide (DMSO), except A β 20–34, Asp23Asn, in which the DMSO concentration was raised to 5%. **b** Intensities of reflections from fiber diffraction of the segments were plotted against *D* spacing. Radial intensity values are vertically staggered for visibility of peaks. **c** From left to right, interface B down the fibril axis of A β 20–34 structure, a model A β 20–34, Asp23Asn on the backbone of the A β 20–34 structure, A β 20–34, isoAsp23 structure, and a model A β 20–34, Asp23Asn on the backbone of the A β 20–34, isoAsp23 structure. **d** A view perpendicular to the fibril axis of residues 23–24 of each structure. Yellow dashed lines represent measured distances in Å between the amide carboxyl of residue 23 and the amide nitrogen of Val24 on the adjacent strand. Source data are provided as a Source Data file

turn with a hydrophobic core of interacting residues within the same chain.

While not all full-length native structures contain β -arches, such as the peptide dimer structure shown in Schmidt et al.⁴⁷ (PDB code: 5AEF), all do include ordered cores involving steric zippers similar to those found in shorter amyloid peptide structures, and a majority of the known A β structures do display β -helix-like turns as seen in the segment structures (Fig. 6 and Supplementary Fig. 5). The native A β 20–34 structure aligns well with a number of these full-length A β structures, and both the native and isomerized structures presented here have the lowest total atom root-mean-square deviation (RMSD) with a structure of the A β Osaka mutant²⁹, E22A, at 2.741 and 2.963 Å, respectively. A tree representing the structural relationships between residues 20 and 34 of eight full-length A β structures and our A β 20–34 structure based on total atom RMSD values shows that 6 of the 8 structures contain turns about the Gly25 and Gly29 residues^{21,23,24,26,28,29}, creating interfaces which align well with

interface B of our L-Asp A β 20–34 structure. Four^{21,23,26,29} of these structures correspond to both the A β 20–34 segment structures with regard to the placement of charged residues Glu22, Asp23, and Lys28 outside the hydrophobic core and yield total atom RMSD values of ≤ 4 Å with A β 20–34 (Fig. 6 and Supplementary Fig. 5). These strong overlaps between our segment structure and other full-length A β structures support the validity of this segment as an atomic resolution structure of an A β core. Importantly, in each of the full-length structures shown here, the putative interface B is accessible as a possible secondary nucleation site (Fig. 6). This interface is stabilized within our structures by the L-isoAsp modification, which mates more tightly between protofilaments than the L-Asp interface B and excludes waters. Thus a full-length structural polymorph with this interface may be isolated more readily with the modification.

The increased structural complexity afforded by extending from 11 to 15 residues is appreciated best in comparing the crystal structures of A β 20–34 to the shorter A β 24–34 crystal

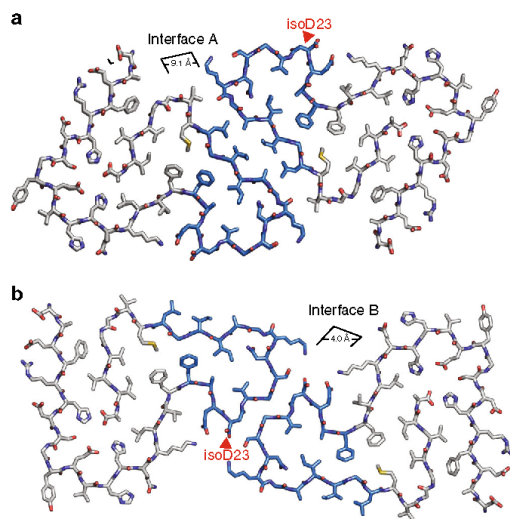


Fig. 5 Model of $A\beta^{20-34}$, $isoAsp23$ as the core of an $A\beta$ 1–42 modified polymorph. **a** Model of $A\beta^{1-42}$, $isoAsp23$ centered on interface A. **b** Centered on interface B. Blue residues correspond to the crystal structure core ($A\beta^{20-34}$, $isoAsp23$), gray sticks correspond to the modeled extension (1–19 and 35–42)

structure, 5VOS³² (Fig. 7). The four extra N-terminal residues in both native and modified $A\beta^{20-34}$ facilitate formation of kinks at Gly25 and Gly29, creating an internal core, whereas the $A\beta$ 24–34 peptide assumes a linear β -strand. Despite $A\beta$ 24–34 lacking these kinks, there is remarkable alignment between residues Gly29 to Leu 34 and interface A of the $A\beta^{20-34}$ crystals, yielding a total atom RMSD of 0.70 and 0.68 Å with the native and isomerized forms, respectively (Fig. 7). An inhibitor was previously developed to the human islet amyloid polypeptide (hIAPP) steric zipper interface analogous to this interface of the 5VOS $A\beta$ 24–34 segment and was shown to be effective against fibril formation of both hIAPP and full-length $A\beta^{32}$. Given the striking alignment between our $A\beta^{20-34}$ interface A and the 5VOS Gly29–Leu34 segment, as well as the distinct lack of modifications and mutations in the region of Asn27–Gly33, this interface may be an ideal scaffold for $A\beta$ inhibitor design in both its homotypic steric zipper form as shown here or in the heterotypic zippers displayed in many of the full-length $A\beta$ structures (Fig. 6).

Discussion

The typical age of onset for sporadic AD is after 65 years, suggesting that slow spontaneous processes such as the accumulation of age-dependent PTMs in $A\beta$ may be contributing factors to aggregation and toxicity⁴. The spontaneous isomerization of aspartate ($isoAsp$) has been identified at all three aspartate residues within the $A\beta$ 1–42 peptide—1, 7, and 23. However, immunohistochemical studies have shown that, while native $A\beta$ and $isoAsp7$ $A\beta$ are present in senile plaques from four non-disease patient controls, $isoAsp23$ $A\beta$ was identified only in one of the four non-disease patient controls, as well as in the senile plaques from all AD patient samples, indicating that the $isoAsp23$ may be more specifically associated with AD pathology than native $A\beta$ and the L- $isoAsp7$ form¹⁰. This implied pathogenicity of $isoAsp23$ correlates with *in vitro* studies, which have

demonstrated accelerated amyloid formation of the $isoAsp23$ $A\beta$ 1–40 and 1–42 peptides compared to native $A\beta^{10,11,17,18,20}$. These results suggest that the change in the structure of $A\beta$ accompanying isomerization at Asp23 may represent a route to the pathogenesis of AD.

In this work we present the 1.1 Å structures of segments spanning residues 20–34 of the $A\beta$ peptide containing either an Asp or an $isoAsp$ residue at site 23. These 15-residue segments, crystallized at physiological pH, maintain a topology seen in the core of $A\beta$ fibrils, a β -helix-like turn (Fig. 6). The length of these peptides facilitates their similar overall fold to previous WT $A\beta$ fibril structures and demonstrates that amyloid cores are rigid and ordered enough to form crystals. These structures reveal a previously unseen protofilament interface (B) involving residues Asp23–Lys28 in the native structure, and residues L- $isoAsp23$ –Ser26 in the isomerized structure. The native interface (L-Asp interface B) has low surface complementarity and contains six water molecules encased between charged residue pairs Asp23 and Lys28 on opposing sheets. In contrast, the isomerized interface (L- $isoAsp$ interface B) is a dry tightly mated sheet with high surface complementarity. Our data suggest that the changes in the structure along this interface, namely, the exclusion of water molecules and van der Waals attractive forces associated with the high S_D , are likely responsible in part for the increases in fiber formation rate and stability of the aggregate observed for the isomerized peptide. The modified interface may provide a better site for secondary nucleation of amyloid formation resulting in the observed enhancements in aggregation. However, it cannot be ruled out from the data presented here that the flexibility imparted by the methylene group of the L- $isoAsp$ residue promotes amyloid formation by allowing an ordered nucleus for primary nucleation to form at a faster rate than the native peptide.

Our models of Asn23 in the $A\beta^{20-34}$, $isoAsp23$ “L- $isoAsp$ interface B” indicate that the completely dry interface may be possible for native residues (Fig. 4). However, the native $A\beta^{20-34}$ structure did not preferentially adopt this interface and instead forms a hydrated L-Asp Interface B. Similar to our $A\beta^{20-34}$ peptide structure, alignments of previous $A\beta$ structures onto the $A\beta^{20-34}$ and $A\beta^{20-34}$, $isoAsp23$ protofilaments show the native Asp23 side chain carboxyl group protruding into the putative interface B region (Supplementary Fig. 5). The hereditary Iowa mutant nuclear magnetic resonance (NMR) structure (Fig. 6 and Supplementary Fig. 5 (PDB: 2MPZ²²)) kinks at Gly25 and Asn27, rather than at Gly25 and Gly29, and thus there is no equivalent interface A. Yet, our preparations of crystals in TBS of the $A\beta^{20-34}$, $A\beta^{20-34}$, $isoAsp23$, and $A\beta^{20-34}$, Asp23Asn constructs appear nearly identical by XRD, suggesting that the structure of an Iowa mutant protofilament would resemble the native and isomerized structures presented here (Fig. 4), barring minor differences due to packing polymorphisms or different environmental conditions.

It is clear that both the isomerization and Iowa mutation at residue 23 accelerate aggregation and increase stability of $A\beta$ fibrils. Our structures of $A\beta^{20-34}$ and $A\beta^{20-34}$, $isoAsp23$ reveal a potential mechanism for the increases in fiber formation rate and fiber stability within the $isoAsp23$ form: the addition of a completely dry interface with high surface complementarity. This analysis leads to the hypothesis that the Asp23 isomerization *in vivo* could lead to the accelerated formation of $A\beta$ fibrils, thereby contributing to the aggregation of $A\beta$ and AD pathology. The hereditary Iowa mutation Asp23Asn may work in a similar manner either by forming the same fold as the isomerized Asp23 or, since Asn undergoes isomerization more rapidly relative to Asp, may also produce an isomerized $A\beta$ with accelerated aggregation and increased stability. The isomerized structure may

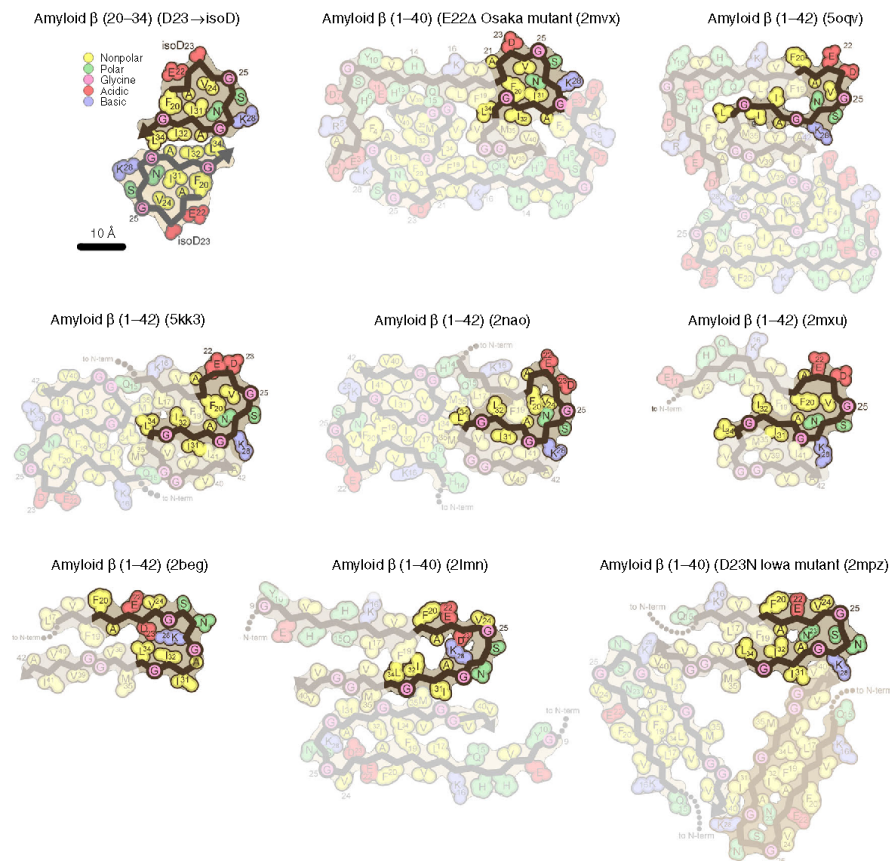


Fig. 6 $A\beta^{20-34}$, $isoAsp^{23}$ core assumes a similar fold to full-length native $A\beta$ structures. Schematic diagrams of residues 20–34 of previously solved $A\beta$ structures^{21–26,28,29}. The structures most divergent from $A\beta^{20-34}$, $isoAsp^{23}$ are shown in the bottom row. Residues are colored according to general chemical properties (legend—top left)

also provide insight into the mechanisms behind the A21G, E22G, and E22A hereditary mutations that introduce flexibility into the same region of the backbone. Importantly, we have also found that the only known repair pathway for L-isoAsp, the enzyme PCMT1, is unable to fully methylate and repair aggregates of $A\beta^{20-34}$, $isoAsp^{23}$ in vitro, thus once the modified aggregates have formed in vivo they may be difficult to repair and clear (Supplementary Fig. 1).

Recent structures of tau isolated from AD patients have revealed distinct structural polymorphs⁴⁸. Both the paired helical filaments and the straight filaments of tau display β -arches in their sheets, which is a feature also shared by the native and isomerized $A\beta^{20-34}$ structures (Fig. 6 and Supplementary Fig. 4). This similarity not only suggests that our structure's β -helix-like turn may be a common amyloid motif but also identifies a potential cross-seeding site between $A\beta$ and the tau protein of AD. This discovery emphasizes the need for atomic-resolution structures of disease-associated amyloid, as these core segments are critical for structure-based drug design and protein prediction efforts^{49–52}. These crystal structures can be used in conjunction with full-length cryo-EM structures to obtain a high-resolution view of the interactions mediating amyloid fiber formation⁵³. High-resolution

structures are also valuable when looking at the effect PTMs may have on amyloid structure as seen here and elsewhere⁵⁴. Therefore, the combination of increasing peptide length and high resolution makes the $A\beta^{20-34}$ and $A\beta^{20-34}$, $isoAsp^{23}$ structures an important step forward for the structural characterization of amyloid proteins and their role in disease.

Methods

Materials. $A\beta^{20-34}$ peptides corresponding to the human sequence were purchased from and validated by Genscript at a purity of $\geq 98\%$ as the trifluoroacetic acid salt and were stored at -20°C . Peptides were validated by electrospray ionization-mass spectrometry (ESI-MS) performed by Genscript. $A\beta$ 1–42 was purchased from Bachem Americas, Inc. (Catalog #, H-1368).

Aggregation of $A\beta^{20-34}$ peptides for fibril-formation rates. Peptides were dissolved at 1.6 mM in 50 mM Tris-HCl, pH 7.6, 150 mM NaCl (TBS) with 2.5% DMSO unless otherwise designated in the figure legend. Peptide solutions were filtered through 0.22- μm cellulose acetate Costar Spin-X centrifuge tube filters (Corning Inc., product #8161). Filtered peptide solutions in a final volume of 100 μL /well in a 96-well plate (Fisherbrand, 12565501) were read at 340 nm in a Varioskan plate reader at 37°C with continuous shaking at 1200 rpm. Readings were recorded every 15 min.

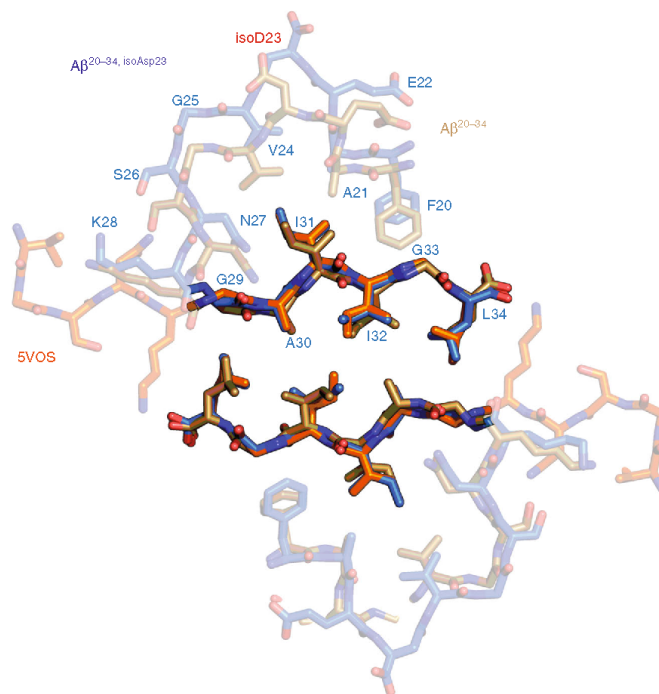


Fig. 7 A β 24–34 peptide structure shares similarities with the A β ^{20–34} structures. The A β steric zipper structure with the lowest total atom root-mean-square deviation of all the short A β segment structures, PDB: 5VOS³² (orange), is shown aligned with interface A of A β ^{20–34} (gold) and A β ^{20–34, isoAsp23} (blue)

Seeding of A β ^{20–34} segment. Seeds were formed shaking continuously on an acoustic resonant shaker at 37 °C at a frequency setting of 37^{33,34}. Seeds of A β ^{20–34} were formed at 5 mg/mL in 50 mM Tris-HCl, pH 7.6, 150 mM NaCl (TBS) with 1% DMSO; seeds of A β ^{20–34, isoAsp23} were formed at 2.5 mg/mL in 50 mM Tris-HCl, pH 7.6, 150 mM NaCl (TBS) with 1% DMSO; and seeds of A β ^{20–34, Asp23Asn} were formed at 2.5 mg/mL in 100 mM Tris-HCl, pH 7.5, 10% isopropanol, and 200 mM sodium acetate. All seeds were diluted to 200 μ M stocks and 5 μ L were added to 3.2 mM A β ^{20–34} in a final volume of 100 μ L. Not all wells of the unseeded 3.2 mM A β ^{20–34} condition aggregated within the time course of this assay (Fig. 1d). Solutions were read in a 96-well plate at 340 nm in a Varioskan plate reader at 37 °C with continuous shaking at 1200 rpm. Readings were recorded every 15 min.

Synthesis and purification of native A β 1–40. The syntheses of A β (1–40) WT and A β (1–40) IsoAsp23 were completed in a CEM Liberty Blue™ Microwave Peptide Synthesizer. The crude peptides were purified using an Interchim puri-Flash® 4125 Preparative Liquid Chromatography System. Details of the syntheses and purifications are available in the Supplementary Methods section (Supplementary Tables 1–3, Supplementary Figs. 6–22).

The purified A β (1–40) WT has an estimated purity of 93% by high-performance liquid chromatography (HPLC; Supplementary Figs. 14 and 15) and was characterized by ESI-MS via direct injection into a Q-Exactive™ Plus Hybrid Quadrupole-Orbitrap™ Mass Spectrometer (Supplementary Figs. 16 and 17). The calculated average mass for C₁₉₄H₂₉₅N₅₃O₅₈S: 4327.148 g/mol, *m/z* calculated: [M+3H]³⁺ = 1443.39; [M+4H]⁴⁺ = 1082.79; [M+5H]⁵⁺ = 866.44; [M+6H]⁶⁺ = 722.20. Observed: 1443.3913; 1082.7955; 866.4374; 722.1991. The purified A β (1–40) IsoAsp23 has an estimated purity of 97% by HPLC (Supplementary Figs. 19 and 20) and was characterized by ESI-MS via direct injection into a Q-Exactive™ Plus Hybrid Quadrupole-Orbitrap™ Mass Spectrometer (Supplementary Figs. 21 and 22). The calculated average mass for C₁₉₄H₂₉₅N₅₃O₅₈S: 4327.148 g/mol, *m/z* calculated: [M+3H]³⁺ = 1443.39; [M+4H]⁴⁺ = 1082.79; [M+5H]⁵⁺ = 866.44. Observed: 1443.3912; 1082.7959; 866.4373.

Crystallization of the segments. A β ^{20–34} was resuspended at a concentration of 3.2 mM in 50 mM Tris-HCl, pH 7.5, 150 mM NaCl (TBS) with 1% DMSO in a final volume of 100 μ L. The peptide solution was then shaken continuously for 30 h at

1200 rpm at 37 °C. A β ^{20–34, isoAsp23} was resuspended at a concentration of 1.6 mM in 50 mM Tris-HCl, pH 7.6, 150 mM NaCl (TBS) with 1% DMSO in a final volume of 200 μ L. The filtered peptide solution was then shaken for 2 days on an acoustic resonant shaker at 37 °C at a frequency setting of 37^{34,35}. Four microliters of this suspension was then used to seed 196 μ L of a second peptide solution (1.6 mM) as a 2% seed on the acoustic resonant shaker at 37 °C. Crystals were obtained within 48 h. The presence of crystals was verified by EM, using a standard holder, with no negative stain. Crystals of the native and isomerized segments were on average ~77 and ~71 nm in width, respectively, and were typically >2 μ m in length.

MicroED sample preparation. Quantifoil R1.2/1.3 cryo-EM grids (Electron Microscopy Sciences, product # Q325CR1.3) were glow discharged for 30 s on either side, and 1.5 μ L of a 1:1 dilution of A β ^{20–34} crystals in 50 mM Tris-HCl, pH 7.5, 150 mM NaCl (TBS) with 1% DMSO was pipetted on both sides. Twenty microliters of A β ^{20–34, isoAsp23} crystal suspensions were spun down at 5000 \times g for 5 min, the supernatant was removed, and pelleted crystals were resuspended in 50 μ L TBS + 0.75% (*w/v*) β -octyl-glucoside (VWR, P-1110), and rotated at 4 °C for 1 h. These detergent-treated crystals were then spun down a second time. Pelleted crystals were resuspended in 50 μ L water. A total of 1.5 μ L of the washed crystal solution was then applied on both sides of a glow discharged Quantifoil R1.2/1.3 cryo-EM grid (Electron Microscopy Sciences, product # Q325CR1.3). All grids were plunge frozen into supercooled ethane using a Vitrobot Mark 4 instrument.

MicroED data collection and processing. MicroED data was collected in a manner similar to previous studies⁴². Briefly, plunge-frozen grids were transferred to an FEI Talos Arctica electron microscope and diffraction data were collected using a bottom-mount CetaD 16M CMOS camera with a sensor size of 4096 \times 4096 pixels, each 14 \times 14 μ m. Diffraction patterns were recorded by operating the detector in continuous mode with 2 \times 2 pixel binning, producing datasets with frames 2048 \times 2048 pixels in size. The exposure rate was set to <0.01 e⁻/Å²/s. The exposure time per frame was set at 3 s while the rotation speed was set to 0.3 deg/s resulting in a final oscillation range of 0.9 deg/exposure for the A β ^{20–34} data collection and to 0.443 deg/s resulting in a final oscillation range of 1.329 deg/exposure for the A β ^{20–34, isoAsp23} data collection. This rotation rate was optimized to allow a maximum amount of reciprocal space to be sampled before crystal decay

was observed while also slow enough to prevent overlapping diffraction spots in the diffraction images. Diffraction movies typically covered a 50–140 deg wedge of reciprocal space and were taken of crystals randomly orientated on the grid with respect to the incident beam. These crystals had a highly preferred orientation on the grid, resulting in a systematic missing cone and hence lower completeness along the c^* axis; however, this did not preclude structure determination, with a high overall completeness of >80% for both structures (see Table 1).

Structure determination. Diffraction datasets were converted to SMV format to be compatible with the X-ray data processing software⁵⁶. Data were indexed and integrated using XDS⁵⁶. The parameters controlling the raster size during indexing and integration were optimized to reduce contributions by background and to exclude intensities that conform poorly to the lattice determined during indexing. The number of diffraction images used per crystal was aggressively pruned to maximize I/σ . The resulting outputs from XDS were sorted and merged in XSCALE. To produce a final merged dataset, partial datasets were selected based on their effects on the Rmerge values. In total, for the $A\beta^{20-34}$ structure, 10 partial datasets, containing 404 diffraction images, were merged to produce a final dataset with high completeness up to 1.1 Å. An ab initio solution was achieved using SHELXD⁵⁷. In total, for the $A\beta^{20-34}$, isoAsp²³ structure, 5 partial datasets, containing 159 diffraction images, were merged to produce a final dataset with high completeness up to 1.1 Å, and an ab initio solution was also achieved using SHELXD. The phases obtained from both $A\beta^{20-34}$ coordinates produced by SHELX were used to generate maps of sufficient quality for subsequent model building in Coot⁵⁸. The resulting models were refined with Phenix⁵⁹, using electron scattering form factors, against the measured data.

Powder diffraction sample preparation and data collection. Designated aggregates of $A\beta$ 1–42 and $A\beta^{20-34}$ peptides were prepared in buffers as described in the figure legends. Aggregates were spun at 20,000 × g for 5 min. The pellet was resuspended in water and re-spun. Pelleted fibrils were resuspended in 5 μ L water and pipetted between two facing glass rods that were 2 mm apart and allowed to dry overnight at room temperature. These glass rods with ordered fibrils were secured to a brass pin and mounted for diffraction at room temperature using 1.54 Å X-rays produced by a Rigaku FRE+ rotating anode generator equipped with an HTC imaging plate. Patterns were collected at a distance of 200 mm and analyzed using the ADXV software package⁶⁰.

SDS dissolution of aggregates. Aggregates of $A\beta^{20-34}$, $A\beta^{20-34}$, isoAsp²³, and $A\beta^{20-34}$, Asp²³Asn were all prepared in TBS, with 1%, 2.5%, and 2.5% DMSO, respectively. Both $A\beta^{20-34}$, isoAsp²³ and $A\beta^{20-34}$, Asp²³Asn were prepared at a peptide concentration of 2.5 mg/mL, while $A\beta^{20-34}$ was prepared at 5 mg/mL, shaking at 1200 rpm at 25 °C. The $A\beta^{20-34}$ was diluted to 2.5 mg/mL prior to the denaturation assay. Suspensions of $A\beta^{20-34}$ aggregates were diluted 1:1 in 2, 3, 4, and 10% SDS stocks in TBS and heated for 15 min at 70 °C in a PTC-100 Peltier thermal cycler as described by Guenther et al.⁵⁴. Measurements at 340 nm were recorded on a Nanodrop 2000 instrument. Two microliters of each solution was analyzed by EM for remaining aggregates on glow discharged Formvar/Carbon 400 mesh, Copper grids (Ted Pella, Catalog # 01754-F).

Analysis of S_b and surface S_c in $A\beta^{20-34}$ structures. The structures of $A\beta^{20-34}$ and $A\beta^{20-34}$, isoAsp²³ were used to measure buried surface area (S_b) and (S_c) from an assembly consisting of two sheets generated by translational symmetry each consisting of ten stacked β -strands. S_b was calculated as the average of the buried surface area per chain and the difference between the sum of the solvent accessible surface area of the two sheets and the solvent accessible surface area of the entire complex, divided by the total number of strands in both sheets using the CCP4 suite.

Modeling modified and full-length $A\beta$ and RMSD calculations. Residues 1–42 of $A\beta$ were modeled onto the N- and C-termini of the $A\beta^{20-34}$, isoAsp²³ structure using Coot, and the resulting structures were energy minimized using the Crystallography & NMR System (CNS)⁶¹ suite of programs.

Distance matrices for RMSD relationships between $A\beta^{20-34}$, isoAsp²³ and residues 20–34 from native structures were generated in the LSQKAB program of CCP4, and resulting matrices were used to generate the tree shown in Fig. 4.

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

Data availability

Atomic coordinates and structure factors for the $A\beta^{20-34}$ structure have been deposited in the Protein Data Bank under accession code 6OIZ. The map for this structure has been deposited in the EMDB with accession code EMD-20082. Atomic coordinates and structure factors for the $A\beta^{20-34}$, isoAsp²³ structure have been deposited in the Protein Data Bank under accession code 6NB9. The map for this structure has been deposited in the EMDB with accession code EMD-0405. The source data underlying Figs. 1c, d, 2, and

4b and Supplementary Figs. 1 and 2 are provided as a Source Data file. Other data are available from the corresponding author upon reasonable request.

Received: 6 February 2019 Accepted: 24 June 2019

Published online: 26 July 2019

References

- Selkoe, D. J. The molecular pathology of Alzheimer's disease. *Neuron* **6**, 487–498 (1991).
- Hardy, J. A. & Higgins, G. A. Alzheimer's disease: the amyloid cascade hypothesis. *Science* **256**, 184–185 (1992).
- Selkoe, D. J. & Hardy, J. The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Mol. Med.* **8**, 595–608 (2016).
- Roher, A. E. et al. APP/A β structural diversity and Alzheimer's disease pathogenesis. *Neurochem. Int.* **110**, 1–13 (2017).
- Kummer, M. P. & Heneka, M. T. Truncated and modified amyloid-beta species. *Alzheimers Res. Ther.* **6**, 28 (2014).
- Rezaei-Ghaleh, N. et al. Phosphorylation modifies the molecular stability of β -amyloid-deposits. *Nat. Commun.* **7**, 11359 (2016).
- Kumar, S. et al. Phosphorylation of the amyloid β -peptide at Ser26 stabilizes oligomeric assembly and increases neurotoxicity. *Acta Neuropathol.* **131**, 525–537 (2016).
- Mandler, M. et al. Pyroglutamylated amyloid- β is associated with hyperphosphorylated tau and severity of Alzheimer's disease. *Acta Neuropathol.* **128**, 67–79 (2014).
- Kummer, M. P. et al. Nitration of tyrosine 10 critically enhances amyloid- β aggregation and plaque formation. *Neuron* **71**, 833–844 (2011).
- Shimizu, T., Fukuda, H., Murayama, S., Izumiya, N. & Shirasawa, T. Isoaspartate formation at position 23 of amyloid beta peptide enhanced fibril formation and deposited onto senile plaques and vascular amyloids in Alzheimer's disease. *J. Neurosci. Res.* **70**, 451–461 (2002).
- Fukuda, H. et al. Synthesis, aggregation, and neurotoxicity of the Alzheimer's Abeta1-42 amyloid peptide and its isoaspartyl isomers. *Bioorg. Med. Chem. Lett.* **9**, 953–956 (1999).
- Al-Hilaly, Y. K. et al. A central role for dityrosine crosslinking of amyloid- β in Alzheimer's disease. *Acta Neuropathol. Commun.* **1**, 1–17 (2013).
- Geiger, T. & Clarke, S. G. Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. Succinimide-linked reactions that contribute to protein degradation. *J. Biol. Chem.* **262**, 785–794 (1987).
- Fujii, N., Takata, T. & Fujii, N. Quantitative analysis of isomeric (l- α , l- β , D- α , D- β) aspartyl residues in proteins from elderly donors. *J. Pharm. Biomed. Anal.* **10**, 25–33 (2015).
- Fujii, N. et al. D-amino acids in protein: the mirror of life as a molecular index of aging. *Biochim. Biophys. Acta* **1866**, 840–847 (2018).
- Roher, A. E. et al. beta-Amyloid-(1-42) is a major component of cerebrovascular amyloid deposits: implications for the pathology of Alzheimer disease. *Proc. Natl Acad. Sci. USA* **90**, 10836–10840 (1993).
- Shimizu, T., Watanabe, A., Ogawara, M., Mori, H. & Shirasawa, T. Isoaspartate formation and neurodegeneration in Alzheimer's disease. *Arch. Biochem. Biophys.* **381**, 225–234 (2000).
- Tomikodora, Y. et al. Iowa variant of familial Alzheimer's disease: accumulation of posttranslationally modified AbetaD23N in parenchymal and cerebrovascular amyloid deposits. *Am. J. Pathol.* **176**, 1841–1854 (2010).
- Rackiewicz, J. L., Zipse, H., Clarke, S. & Houk, K. N. Neighboring side chain effects on asparaginyl and aspartyl degradation: an ab initio study of the relationship between peptide conformation and backbone NH acidity. *J. Am. Chem. Soc.* **123**, 3499–3506 (2001).
- Fossati, S. et al. Differential contribution of isoaspartate post-translational modifications to the fibrillization and toxic properties of amyloid- β and the Asn23 Iowa mutation. *Biochem. J.* **456**, 347–360 (2013).
- Wälti, M. A. et al. Atomic-resolution structure of a disease-relevant $A\beta$ (1-42) amyloid fibril. *Proc. Natl Acad. Sci. USA* **113**, E4976–E4984 (2016).
- Sgourakis, N. G., Yau, W. M. & Qiang, W. Modeling an in-register, parallel “iowa” $a\beta$ fibril structure using solid-state NMR data from labeled samples with rosetta. *Structure* **23**, 216–227 (2015).
- Xiao, Y. et al. $A\beta$ (1-42) fibril structure illuminates self-recognition and replication of amyloid in Alzheimer's disease. *Nat. Struct. Mol. Biol.* **22**, 499–505 (2015).
- Paravastu, A. K., Leapman, R. D., Yau, W. M. & Tycko, R. Molecular structural basis for polymorphism in Alzheimer's beta-amyloid fibrils. *Proc. Natl Acad. Sci. USA* **105**, 18349–18354 (2008).
- Lührs, T. et al. 3D structure of Alzheimer's amyloid-beta(1-42) fibrils. *Proc. Natl Acad. Sci. USA* **102**, 17342–17347 (2005).
- Colvin, M. T. et al. Atomic resolution structure of monomeric $A\beta$ 42 amyloid fibrils. *J. Am. Chem. Soc.* **138**, 9663–9674 (2016).

27. Lu, J. X. et al. Molecular structure of β -amyloid fibrils in Alzheimer's disease brain tissue. *Cell* **154**, 1257–1268 (2013).
28. Gremer, L. et al. Fibril structure of amyloid- β (1–42) by cryo-electron microscopy. *Science* **358**, 116–119 (2017).
29. Schütz, A. K. et al. Atomic-resolution three-dimensional structure of amyloid β fibrils bearing the Osaka mutation. *Angew. Chem. Int. Ed. Engl.* **54**, 331–335 (2015).
30. Guenther, E. L. et al. Atomic-level evidence for packing and positional amyloid polymorphism by segment from TDP-43 RRM2. *Nat. Struct. Mol. Biol.* **25**, 311–319 (2018).
31. Rodriguez, J. A. et al. Structure of the toxic core of α -synuclein from invisible crystals. *Nature* **525**, 486–490 (2015).
32. Krotee, P. et al. Common fibrillar spines of amyloid- β and human islet amyloid polypeptide revealed by microelectron diffraction and structure-based inhibitors. *J. Biol. Chem.* **293**, 2888–2902 (2018).
33. Krotee, P. et al. Atomic structures of fibrillar segments of hIAPP suggest tightly mated β -sheets are important for cytotoxicity. *eLife* **6**, e19273 (2017).
34. Hatami, A., Monjazeb, S., Milton, S. & Glabe, C. G. Familial Alzheimer's disease mutations within the amyloid precursor protein alter the aggregation and conformation of the amyloid- β peptide. *J. Biol. Chem.* **292**, 3172–3185 (2017).
35. Milton, N. G. Phosphorylation of amyloid-beta at the serine 26 residue by human cdc2 kinase. *Neuroreport* **12**, 3839–3844 (2001).
36. Goldschmidt, L., Teng, P. K., Reik, R. & Eisenberg, D. Identifying the amyloids, proteins capable of forming amyloid-like fibrils. *Proc. Natl Acad. Sci. USA* **107**, 3487–3492 (2010).
37. Matta, L. M., Zhu, C., Jagoda, J. & Zinn, T. Mixing by resonant acoustic driving in a closed chamber. *J. Propul. Power* **12**, 366–370 (1996).
38. Leung, D. A new and improved method for the preparation of drug nanosuspension formulations using acoustic mixing technology. *Int. J. Pharm.* **473**, 10–19 (2014).
39. Shi, D., Nannega, B. L., Iadanza, M. G. & Gonen, T. Three-dimensional electron crystallography of protein microcrystals. *eLife* **2**, e01345 (2013).
40. Nannega, B. L. & Gonen, T. Protein structure determination by MicroED. *Curr. Opin. Struct. Biol.* **27**, 24–31 (2014).
41. Holm, L. & Laasko, L. Dall server update. *Nucleic Acids Res.* **44**, W351–W355 (2016).
42. Garnham, C. P., Campbell, R. L. & Davies, P. L. Anchored clathrate waters bind antifreeze proteins to ice. *Proc. Natl Acad. Sci.* **108**, 7363–7367 (2011).
43. Perutz, M. Polar zippers: their role in human disease. *Prot. Sci.* **3**, 1629–1637 (1994).
44. Sawaya, M. R. et al. Atomic structures of amyloid cross- β spines reveal varied steric zippers. *Nature* **447**, 453–457 (2007).
45. Colletier, J. et al. Molecular basis for amyloid- β polymorphism. *Proc. Natl Acad. Sci. USA* **108**, 16938–16943 (2011).
46. Eisenberg, D. S. & Sawaya, M. R. Structural studies of amyloid proteins at the molecular level. *Annu. Rev. Biochem.* **86**, 69–95 (2017).
47. Schmidt, M. et al. Peptide dimer structure in an $\text{A}\beta$ (1–42) fibril visualized with cryoEM. *Proc. Natl Acad. Sci. USA* **112**, 11858–11863 (2015).
48. Fitzpatrick, A. W. P. et al. Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **547**, 185–190 (2017).
49. Sievers, S. A. et al. Structure-based design of non-natural amino-acid inhibitors of amyloid fibril formation. *Nature* **475**, 96–100 (2011).
50. Seidler, P. M. et al. Structure-based inhibitors of tau aggregation. *Nat. Chem.* **10**, 170–176 (2018).
51. Saelices, L. et al. Amyloid seeding of transthyretin by ex vivo cardiac fibrils and its inhibition. *Proc. Natl Acad. Sci. USA* **115**, e6741–e6750 (2018).
52. Saelices, L. et al. Crystal structures of amyloidogenic segments of human transthyretin. *Protein Sci.* **27**, 1295–1303 (2018).
53. Li, B. et al. Cryo-EM of full-length α -synuclein reveals fibril polymorphs with a common structural kernel. *Nat. Commun.* **9**, 3609 (2018).
54. Guenther, E. L. et al. Atomic structures of TDP-43 LCD segments and insights into reversible or pathogenic aggregation. *Nat. Struct. Mol. Biol.* **25**, 463–471 (2018).
55. Hattne, J. et al. MicroED data collection and processing. *Acta Crystallogr. A Found. Adv.* **71**, 353–360 (2015).
56. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
57. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
58. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
59. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012).
60. Porebski, B. T., Ho, B. K. & Buckle, A. M. Interactive visualization tools for the structural biologist. *J. Appl. Cryst.* **46**, 1518–1520 (2013).
61. Brünger, A. T. et al. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921 (1998).

Acknowledgements

We thank Dr. Jose Rodriguez for his advice throughout the course of this work. This work was supported in part by grants from the National Science Foundation (MCB-1714569 to S.G.C. and MCB-1616265 to D.S.E.), the National Institutes of Health (AG 054022 to D.S.E.), and the Howard Hughes Medical Institute (to D.S.E. and T.G.). R.A.W. was supported by a USPHS National Research Service Award GM067185 and a UCLA Dissertation Year Fellowship and Pauley Fellowship Award. D.R.B. was supported by a National Science Foundation Graduate Research Fellowship. L.R. was supported by a USPHS National Research Service Award 5T32GM00849. Work in the S.G.C. laboratory was also supported by funds from the Elizabeth and Thomas Plott Chair in Gerontology of the UCLA Longevity Center, a grant from the Life Extension Foundation, and funds from a UCLA Faculty Research Award. We acknowledge the use of instruments at the Electron Imaging Center for Nanomachines supported by UCLA and by instrumentation grants from NIH (1S10RR23057 and 1U24GM116792) and NSF (DBI-1338135). We thank Michael Collazo at the UCLA-DOE Macromolecular Crystallization Core Technology Center for crystallization support. This work was supported in part by NIGMS grant No. R35GM128867 and the Beckman Young Investigators (BYI) Program awarded to J.A. Rodriguez. The mass analysis of $\text{A}\beta$ (1–40) wild-type and IsoAsp23 was supported by the National Institutes of Health under instrumentation grant 1S10OD016387–01 with the assistance of Yu Chen in the Molecular Instrumentation Center of UCLA.

Author contributions

R.A.W. and S.G.C. designed the project and wrote the manuscript with input from all other authors. R.A.W. conducted fibril growth experiments, stability assays, and crystallization of the $\text{A}\beta$ ^{20–34} peptide. R.A.W. and D.R.B. performed fibril diffraction studies. Samples were prepared for MicroED and data were collected by R.A.W. and D.R.B. with advice from T.G. Data were processed by C.-T.Z. and R.A.W. and refined by R.A.W., L.R., and M.R.S. with significant contributions to data processing by D.R.B., D.C. and M.R.S. R.A.W. and M.R.S. conducted computational analysis including buried and accessible surface area, RMSD, and designed models of full-length $\text{A}\beta$ fibrils. All authors contributed to the analyses of the structures.

Additional information


Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-019-11183-z>.

Competing interests: D.S.E. is an advisor and equity shareholder in ADDRx, Inc. The other authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Peer review information: *Nature Communications* thanks Guillermo Calero and other anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

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

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

Supplementary Information

Structure of amyloid- β (20-34) with Alzheimer's-associated isomerization at Asp23 reveals a distinct protofilament interface

R.A. Warmack *et al.*

Supplementary Methods

Determination of L-isoaspartate levels by the PCMT1 methanol vapor diffusion assay

PCMT1 was used as an analytical reagent to quantify L-isoAsp levels in A β ²⁰⁻³⁴ peptide solutions or aggregates. Aggregates of A β ²⁰⁻³⁴ or A β ^{20-34, isoAsp23} were formed at 3.2 or 1.6 mM, respectively in 50 mM Tris-HCl, pH 7.5, 150 mM NaCl with 1% DMSO. In a final volume of 100 μ L, 130 pmol of either these aggregates or freshly dissolved, filtered peptide solutions were incubated for 2 h at 37 °C with 5 μ g PCMT1 (purified as a His-tagged enzyme from *Escherichia coli* (*E. coli*) containing the expression plasmid #34852 available from Addgene.com as described by Patananan et al., 2014¹ with a specific activity at 37 °C of 5,300 pmol of methyl esters formed on KASA(isoD)LAKY/min/mg of enzyme). Final concentrations in the reactions included 135 mM Bis-Tris-HCl, pH 6.4, and 10 μ M S-adenosyl-L-[methyl³H]methionine ([³H]AdoMet) (prepared by a 1600-fold isotopic dilution of a stock of 72 Ci/mmol [³H]AdoMet (PerkinElmer Life Sciences, NET155H00) with nonisotopically labeled AdoMet (p-toluenesulfonate salt; Sigma-Aldrich A2408)). The reaction was stopped by adding 10 μ L of 2 M sodium hydroxide, and 100 μ L of the 110 μ L mixture was transferred to a 9 by 2.5 cm piece of folded thick filter paper (Bio-Rad; catalog number 1650962) wedged in the neck of a 20-mL scintillation vial above 5 mL scintillation reagent (Safety Solve, Research Products International, catalog number 121000), tightly capped, and incubated at room temperature. After 2 h, the folded filter papers were removed, the caps replaced, and the vials were counted thrice for 5 minutes each in a Beckman LS6500 scintillation counter. Background radioactivity in a reaction containing no substrate was determined by incubating the recombinant human PCMT1, 135 mM Bis-Tris-HCl buffer, and 10 μ M [³H] AdoMet as described above and was subtracted from the value obtained in experimental samples. Samples were analyzed in triplicate.

Fibril formation of full length A β 1-40

Wild-type and L-isoAsp23 A β 1-40 were resuspended at a final concentration of 40 μ M with 10 μ M ThT in 10 mM phosphate, 127 mM NaCl, and 2.7 mM KCl, pH 7.4 (PBS). Fibrils were formed at 37 °C with continuous shaking at 600 rpm in a Varioskan plate reader. Fluorescence of three replicate wells was monitored in a 96 well plate, with readings taken every 5 min (excitation – 440 nm, emission - 482 nm, bottom read). Concentrations of all peptide solutions were verified by absorbance at 280 nm. EM images were recorded on an FEI Tecnai G₂ TF20 TEM.

Synthesis and purification of A β 1-40

Peptide syntheses were carried out at 0.1 mmol scale. A 2-chlorotrityl chloride resin (Advanced Chemtech; SC5055) was selected as the solid support with a nominal loading of 1.0 mmol/g. Each loading of the first amino acid was executed by adding 0.1 mmol (0.34 mg) of Fmoc-Val-OH and 0.4 mmol (70 μ L) of diisopropylethylamine (DIPEA), dissolved in 10 mL of dichloromethane (DCM), to 0.5 grams of resin. This mixture was gently agitated by bubbling with air. After 30 minutes, the supernatant was drained, and the resin was rinsed twice with 15 mL aliquots of capping solution, consisting of 17:2:1 DCM/MeOH/DIPEA. With the first amino acid loaded, the elongation of each polypeptide was completed in a CEM Liberty Blue™ Microwave Peptide Synthesizer.

A 1.0 M solution of N,N'-diisopropylcarbodiimide (DIC; Advanced Chemtech; RC8120/33084) in DMF was used as the primary activator, and a 1.0 M solution of ethyl cyanohydroxyiminoacetate (oxyma; CEM; S001-C/CEM1802117001-052118) in DMF, buffered by 0.1 M of DIPEA was used as a coupling additive. All protected, natural amino acids used during syntheses were purchased from Advanced Chemtech. The following table indicates the batch numbers and the listed HPLC purities provided by the manufacturer. All amino acids from Advanced Chemtech contain less than 0.5% of their corresponding D enantiomer, confirmed by chiral HPLC. For the synthesis of amyloid β -protein (1-40) IsoAsp23, Fmoc-Asp-OtBu (Combi-

Blocks; SS 0525/B15012; 98% purity) was used in place of the Fmoc-Asp(OtBu)-OH for position 23. The microwave synthesizer utilizes 0.2 M solutions of each amino acid.

For the deprotection of N-termini, Fmoc protecting groups, a 9% w/v solution of piperazine in 9:1 N-Methyl-2-Pyrrolidone to EtOH, buffered with 0.1 M of oxyma was used. For 0.1 mmol deprotection reactions, 4 mL of the above deprotection solution was added to the resin. The mixture was then heated to 90°C for 2 minutes while bubbled with nitrogen gas. The solution was drained, and the resin washed with 4 times with 4 mL aliquots of DMF. For 0.1 mmol couplings, 2.5 mL of 0.2 M amino acid solution (0.5 mmol) was added to the resin along with 1 mL of the DIC solution (1.0 mmol) and 0.5 mL of oxyma solution (0.5 mmol). This mixture was agitated by bubbling for 2 minutes at 25°C and then heated to 50°C followed by 8 minutes of bubbling. After the coupling reaction, an Fmoc deprotection reaction without conducting additional washes. Double couplings were used to ensure complete coupling, starting from the attachment of Ala21.

After the last deprotection, the resins were washed with methanol, diethyl ether, dried over vacuum, and introduced to a cleavage cocktail consisting of: 20 mL of trifluoroacetic acid (TFA); 0.330 mL of 1,2-ethanedithiol (EDT); 0.380 mL of water; 0.225 grams of ammonium iodide (NH₄I); 0.300 mL of dimethyl sulfide (DMS); 0.150 mL of trisopropylsilane (TIS). After 2 hours of vigorous stirring, the mixtures were filtered, and the filtrate concentrated in vacuo. The residues were triturated with cold diethyl ether, and 0.412 g and 0.444 g of precipitated, crude peptides were collected by filtration for amyloid β -protein (1-40) WT and Amyloid β -Protein (1-40) IsoAsp23, respectively.

The purity of the crude product, amyloid β -protein (1-40) WT, was analyzed by RP-HPLC, using a Hewlett Packard 1090 Series II Liquid Chromatography System equipped with a Kinetex (Phenomenex, C18, 5 μ m, 100 Å, 4.6 x 250 mm) column. Ultrapure water with 0.1% TFA, and a 1:9 water to acetonitrile solution with 0.095% TFA were selected as mobile phases [A] and [B], respectively. The flow rate was set at 1.0 mL/min and the gradient used is detailed

in **Supplementary Table 2**. The UV absorption at 214 nm was monitored. The resulting chromatogram is shown in **Supplementary Figure 6**.

The crude peptide was then purified by RP-HPLC, using an Interchim puriFlash® 4125 Preparative Liquid Chromatography System equipped with a Luna (Phenomenex, C18(2), 5 µm, 100 Å, 30 x 100 mm) column. For purification, two buffer systems were utilized. Initial purifications and salt exchanges were executed with a 1:99 acetonitrile to water solution, buffered by 15 mM NH₄OH ([A]) and a 2:3 water to acetonitrile solution, buffered by 15 mM of NH₄OH ([B]). For better resolution of diastereomers and other impurities, ultrapure water, buffered by 14 mM of HClO₄, and a 2:3 water to acetonitrile solution, buffered by 5.6 mM of HClO₄, were selected as mobile phases A and B, respectively. Buffer systems with TFA were avoided because large decreases in peak efficiency were observed for the product peak in such systems, precluding compound recovery. The purification gradients used are described in **Supplementary Table 3**. Many purifications were truncated after the peak was eluted in order to prepare the LC instrument immediately for the next injection, minimizing the oxidation of methionine. UV absorptions at 214 nm was monitored. 62 mg of crude peptide was dissolved in 10 mL of an aqueous 1.5 M NH₄OH solution and loaded onto the column via valve injection. **Supplementary Table 3** displays the details of the gradient and **Supplementary Figures 7** and **8** show details of the corresponding HPLC trace.

The LC system has been modified (customization provided by Interchim) to allow for direct loading of solutions onto the column through the pump (see **Supplementary Figure 9**). The fractions collected from Purification 1 were subjected to a two-fold dilution and loaded directly back onto the column for purification. **Supplementary Table 3** displays the details of the gradient and **Supplementary Figures 10** and **11** shows details of the corresponding HPLC trace.

The collected fractions from Purification 2 were pooled and directly loaded back onto the column. The peptide was subjected to a salt exchange, displacing the perchlorate ions with

ammonium ions. The resulting fractions were pooled, flash frozen, and lyophilized. The lyophilized powder was dissolved in 10 mL of an aqueous 1.5 M NH₄OH solution, loaded onto the column via valve injection, and subjected to a third purification. **Supplementary Table 3** displays the details of the gradient and **Supplementary Figures 12** and **13** show details of the corresponding HPLC trace.

The collected fractions from Purification 3 were pooled and directly loaded back onto the column. The peptide was subjected to a salt exchange, displacing the perchlorate ions with ammonium ions. The resulting fractions were flash frozen separately and lyophilized. The purity of these fractions was analyzed by RP-HPLC, using a Hewlett Packard 1090 Series II Liquid Chromatography System equipped with a Kinetex (Phenomenex, C18, 5 μm, 100 Å, 4.6 x 250 mm) column. Ultrapure water, buffered by 14 mM of HClO₄, and a 1:9 water to acetonitrile solution buffered by 1.4 mM of HClO₄, were selected as mobile phases A and B, respectively. The flow rate was set at 1.0 mL/min and the gradient used is detailed in **Supplementary Table 2**. The UV absorption at 214 nm was monitored, and the peaks in the resulting chromatogram were manually integrated.

The purest lyophilized fraction yielded 0.3 mg of material from 62 mg of the crude peptide. To obtain more Aβ (1-40) WT, the purification was repeated with 87 mg of crude material, producing 5.3 mg of material with similar purity. The increase in yield was due to better fraction selection, accounting for solvent delay between the flow cell in the UV detector and the fraction collector. **Supplementary Figures 14** and **15** show the purity analysis via HPLC of the second purification, indicating an estimated purity of 93% for the purest fraction.

The purified Aβ (1-40) WT was also characterized by ESI-MS (**Supplementary Figures 16** and **17**) via direct injection into a Q-Exactive™ Plus Hybrid Quadrupole-Orbitrap™ Mass Spectrometer. The calculated average mass for C₁₉₄H₂₉₅N₅₃O₅₈S: 4327.148 g/mol, m/z calculated: [M+3H]³⁺ = 1443.39; [M+4H]⁴⁺ = 1082.79; [M+5H]⁵⁺ = 866.44; [M+6H]⁶⁺ = 722.20. Observed: 1443.3913; 1082.7955; 866.4374; 722.1991.

The purity of the crude product, amyloid β -protein (1-40) IsoAsp23, was analyzed by RP-HPLC, using a Hewlett Packard 1090 Series II Liquid Chromatography System equipped with a Kinetex (Phenomenex, C18, 5 μ m, 100 \AA , 4.6 x 250 mm) column. Ultrapure water with 0.1% TFA, and a 1:9 water to acetonitrile solution with 0.095% TFA were selected as mobile phases [A] and [B], respectively. The flow rate was set at 1.0 mL/min and the gradient used is detailed in **Supplementary Table 2**. The UV absorption at 214 nm was monitored. The resulting chromatogram is shown in **Supplementary Figure 18**.

The crude IsoAsp23 peptide was purified by RP-HPLC with the same methods and gradients as the WT peptide. The resulting fractions were also analyzed in the same manner. The purest lyophilized fractions of A β (1-40) IsoAsp23 yielded 4.75 mg of material from 84 mg of crude peptide. **Supplementary Figures 19 and 20** shows the purity analysis of one such fraction via HPLC, indicating an estimated purity of 94%.

The purified A β (1-40) IsoAsp23 was also characterized by ESI-MS (**Supplementary Figures 21 and 22**) via direct injection into a Q-ExactiveTM Plus Hybrid Quadrupole-OrbitrapTM Mass Spectrometer. The calculated average mass for C₁₉₄H₂₉₅N₅₃O₅₈S: 4327.148 g/mol, m/z calculated: [M+3H]³⁺ = 1443.39; [M+4H]⁴⁺ = 1082.79; [M+5H]⁵⁺ = 866.44. Observed: 1443.3912; 1082.7959; 866.4373.

Supplementary Tables

Amino Acid	Catalog/Lot Number	Purity by HPLC
Fmoc-L-Ala-OH	FA2100/32786	99.6%
Fmoc-L-Arg(Pbf)-OH	FR2136/32914	99.6%
Fmoc-L-Asn(Trt)-OH	FN2152/32769	98.4%
Fmoc-L-Asp(tBu)-OH	FD2192/32770	99.9%
Fmoc-L-Gln(Trt)-OH	FQ2251/31862	99.6%
Fmoc-L-Glu(tBu)-OH	FE2237/32917	99.4%
Fmoc-L-Gly-OH	FG2275/32787	99.8%
Fmoc-L-His(Trt)-OH	FH2316/32530	99.1%
Fmoc-L-Ile-OH	FI2326/32125	99.9%
Fmoc-L-Leu-OH	FL2350/32771	99.7%
Fmoc-L-Lys(Boc)-OH	FK2390/32531	99.9%
Fmoc-L-Met-OH	FM2400/32127	99.7%
Fmoc-L-Phe-OH	FF2425/32484	99.7%
Fmoc-L-Ser(tBu)-OH	FS2476/32532	99.9%
Fmoc-L-Tyr(tBu)-OH	FY2563/32181	99.7%
Fmoc-L-Val-OH	FV2575/32773	99.8%

Supplementary Table 1: Catalog and lot number of amino acids purchased from Advanced Chemtech

Crude Analyses			QC Analyses		
TFA Buffer System			HClO ₄ Buffer System		
Time	[A]	[B]	Time	[A]	[B]
(min)	(%)	(%)	(min)	(%)	(%)
0	95	5	0	80	20
5	80	20	5	65	35
25	60	40	25	45	55
26	0	100	26	0	100
30	0	100	30	0	100
31	95	5	31	80	20
35	95	5	35	80	20

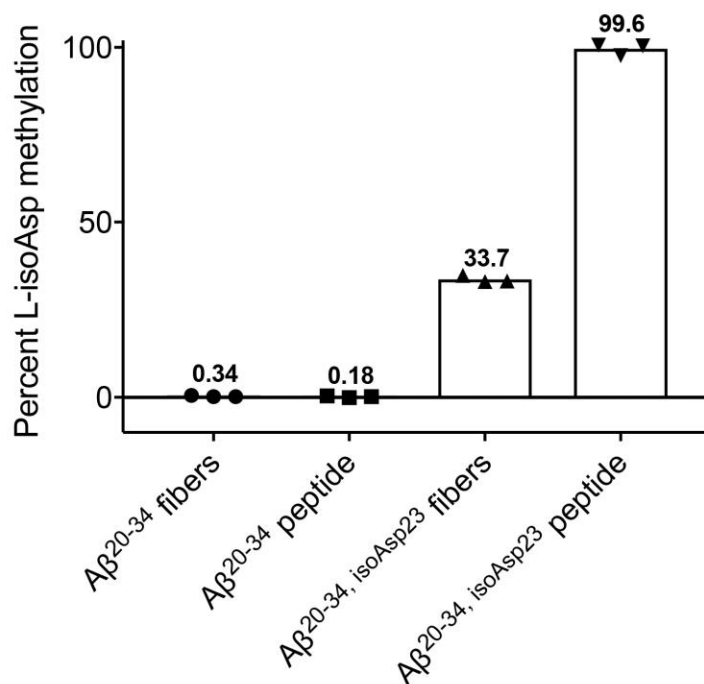
Supplementary Table 2: Gradients utilized for purity analysis. Supplementary Figures 6 and 18 show the analyses of crude peptides, carried out in a TFA buffer system. Supplementary

Figures 14, 15, 19, and 20, show the analyses of purified peptides, carried out in a HClO₄ buffer system.

Purification 1			Purification 2			Purification 3		
NH ₄ OH Buffer System			HClO ₄ Buffer System			HClO ₄ Buffer System		
Time	[A]	[B]	Time	[A]	[B]	Time	[A]	[B]
(min)	(%)	(%)	(min)	(%)	(%)	(min)	(%)	(%)
0	100	0	0	95	5	0	100	0
5	100	0	6	95	5	5	100	0
10	80	20	14	50	50	13	50	50
30	60	40	34	30	70	33	30	70
31	0	100	35	0	100	34	0	100
36	0	100	40	0	100	39	0	100

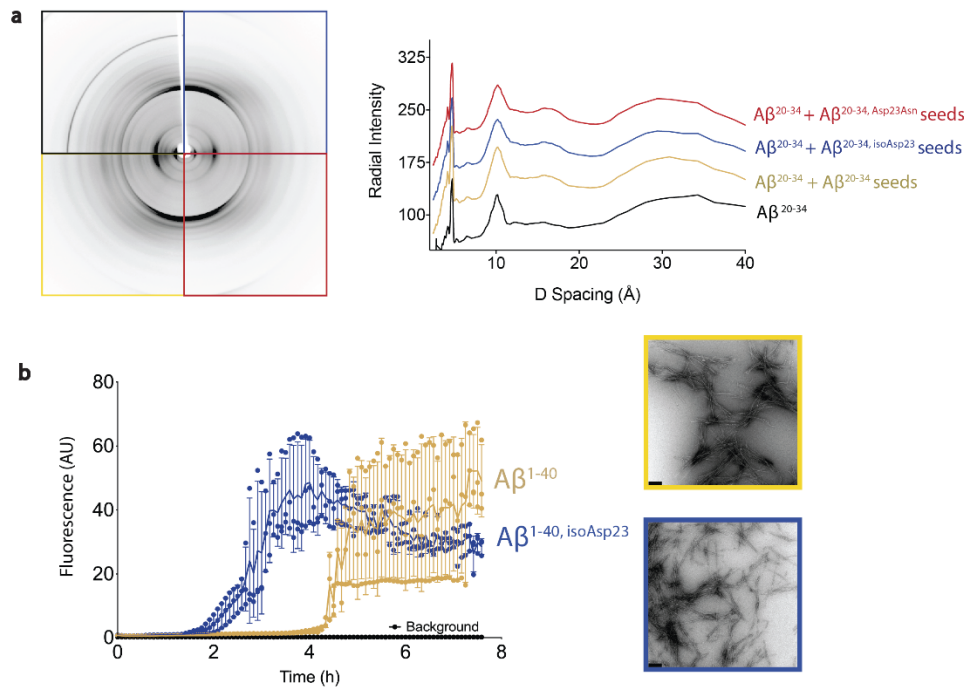
Supplementary Table 3: Gradients utilized for the purification of A β (1-40) WT and A β (1-40) IsoAsp23. Supplementary Figures 7 and 8 correspond to Purification 1. Supplementary Figures 10 and 11 correspond to Purification 2. Supplementary Figures 12 and 13 correspond to Purification 3.

Supplementary Figures

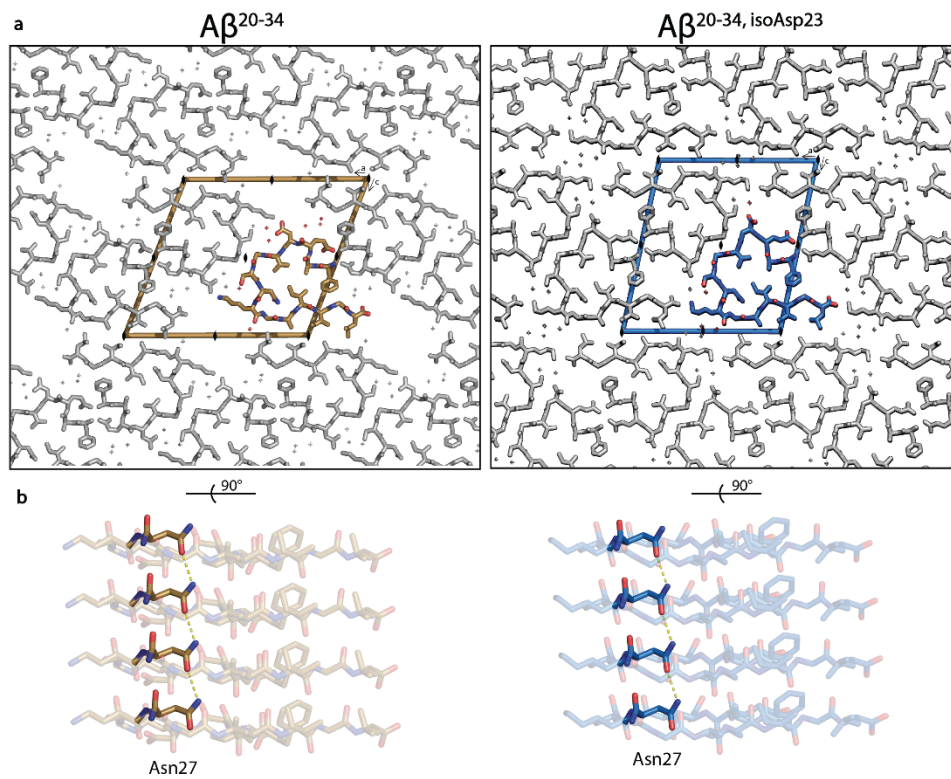


Supplementary Figure 1: PCMT1 is unable to fully methylate aggregated A β ²⁰⁻³⁴, isoAsp23.

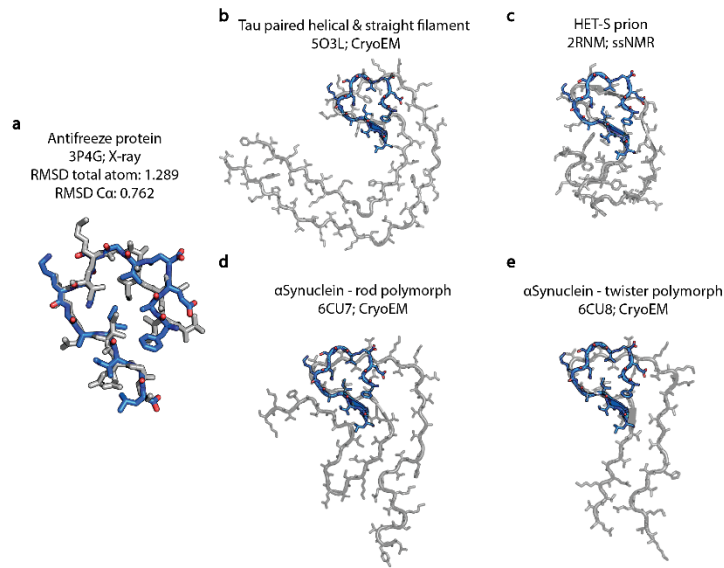
Methylation of free peptide and aggregated fibers of A β ²⁰⁻³⁴ and A β ²⁰⁻³⁴, isoAsp23 was detected as described in the experimental procedures. Levels of detected L-isoAsp were normalized between 0-100% methylation. The normal L-Asp A β ²⁰⁻³⁴ fibers and peptide were included as negative controls and were not methylated by PCMT1, as shown by the averages of 0.34% and 0.18% methylation, respectively. Source data are provided as a Source Data file.



Supplementary Figure 2: Seeded $A\beta^{20-34}$ aggregates display identical fiber diffraction patterns, and the full-length $A\beta^{1-40}$ aggregation matches the $A\beta^{20-34}$ aggregation. **a**, Aggregates from the seeding assay of $A\beta^{20-34}$ shown in the main text Fig. 1d were ordered between glass capillaries and fiber diffraction data was collected as described in the “Methods.” Radial intensity of the reflections was plotted against D Spacing (right). **b**, 20 μM wild-type $A\beta^{1-40}$ (gold lines) or $A\beta^{1-40, \text{isoAsp23}}$ (blue lines) were incubated at 37 $^{\circ}\text{C}$. Fiber formation was monitored by Thioflavin T fluorescence, readings were recorded every 5 min. Each data point is shown as a round symbol, the solid line represents the mean value, and error bars represent SD of three technical replicates. A representative EM image of each condition is shown on the right, scale bars at the lower left represent 200 nm. Source data are provided as a Source Data file.

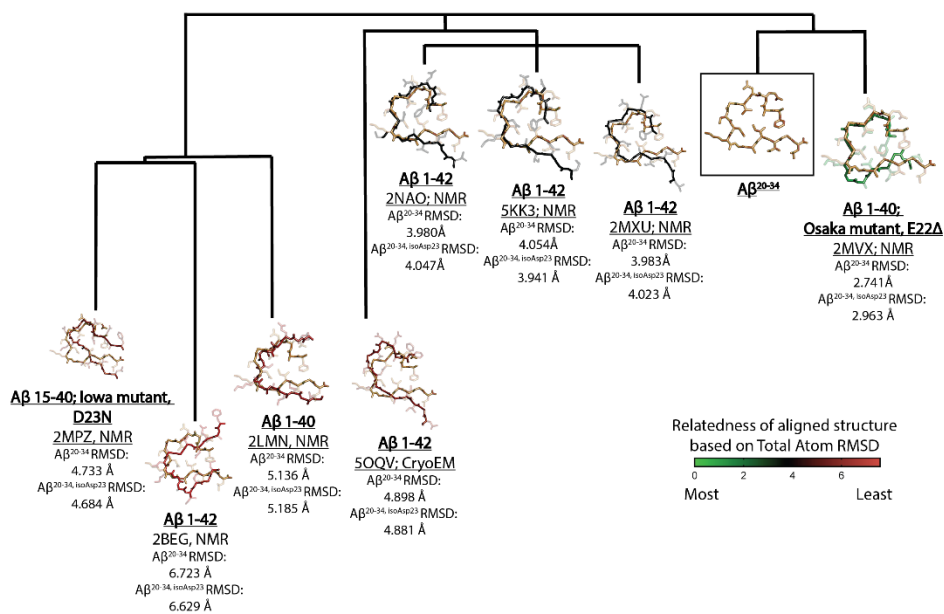


Supplementary Figure 3: $A\beta^{20-34}$ structures reveal two interfaces and a polar zipper. a, The crystal structures are shown here along the 2_1 axis of the unit cell. Black symbols (●) represent the 2_1 axis of symmetry. **b,** The asparagine ladder motif is shown with yellow dashed lines between strands, the structure is shown perpendicular to the protofilament axis, along the face of residues Lys28-Leu34. The dashed lines within the $A\beta^{20-34}$ structure correspond to 2.8 Å, and the dashed lines within the $A\beta^{20-34, \text{isoAsp23}}$ structure correspond to 2.9 Å.

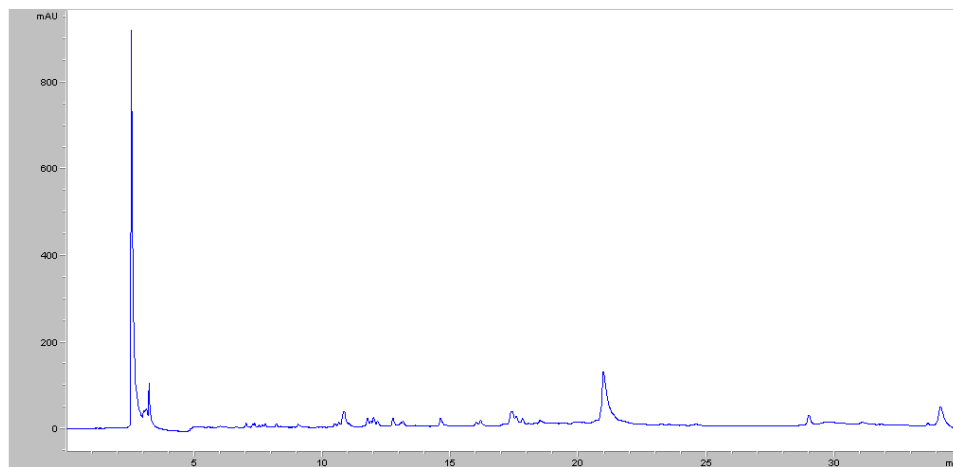


Supplementary Figure 4: A β ^{20-34, isoAsp23} matches β -helical structures and other amyloid

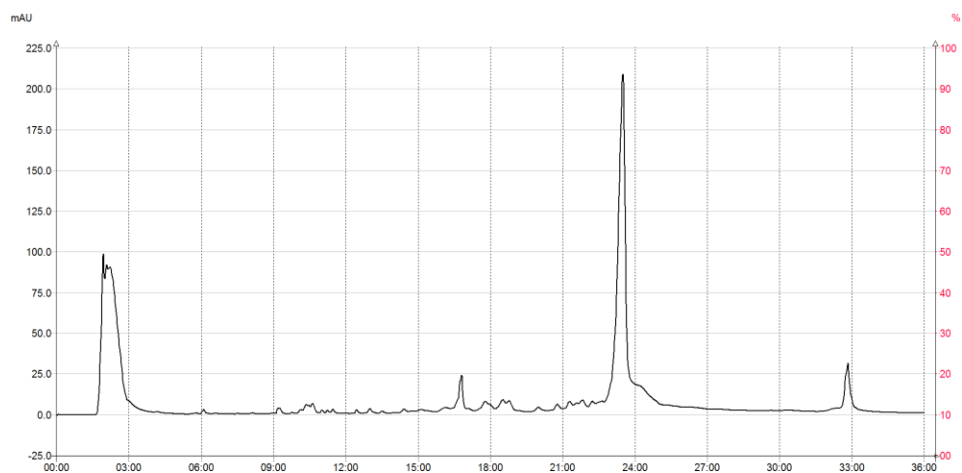
fibrils. **a**, The top match for the A β ^{20-34, isoAsp23} structure from comparison to all PDB structures via a DALI search² was the β -helical antifreeze protein 3P4G³. **b-d**, Other amyloid fibril structures, including tau⁴ and α Synuclein⁵ show similar tight turns (β -arches) between steric zippers.



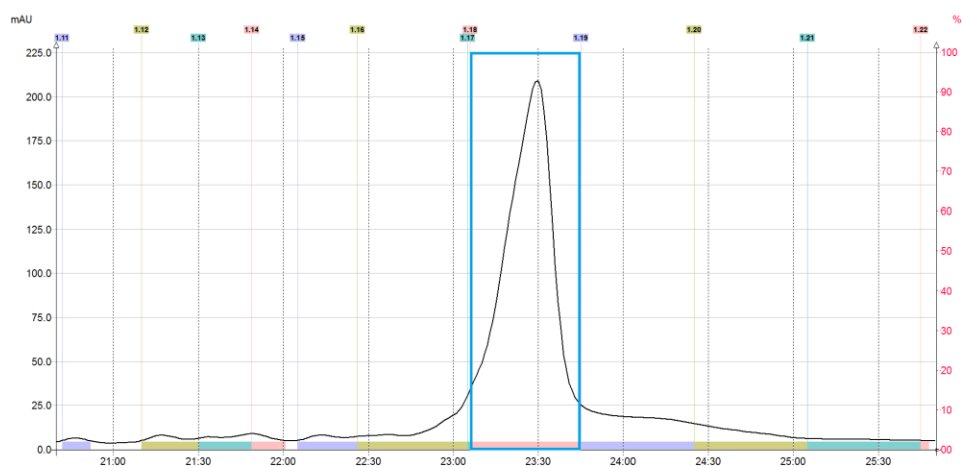
Supplementary Figure 5: Total atom RMSD relationships between Aβ structures. The Aβ^{20-34, isoAsp23} (gold, second from right) was aligned to residues 20-34 of full-length Aβ structures⁶⁻¹³. Backbone and total atom RMSDs were calculated using CCP4. Branches of the evolutionary tree represent total atom RMSD relatedness between the structures and were generated as described in the Methods section. The backbones of the aligned structures are also colored on a scale of red to green according to how closely matched their total RMSD values are to Aβ²⁰⁻³⁴ (legend shown on bottom right).



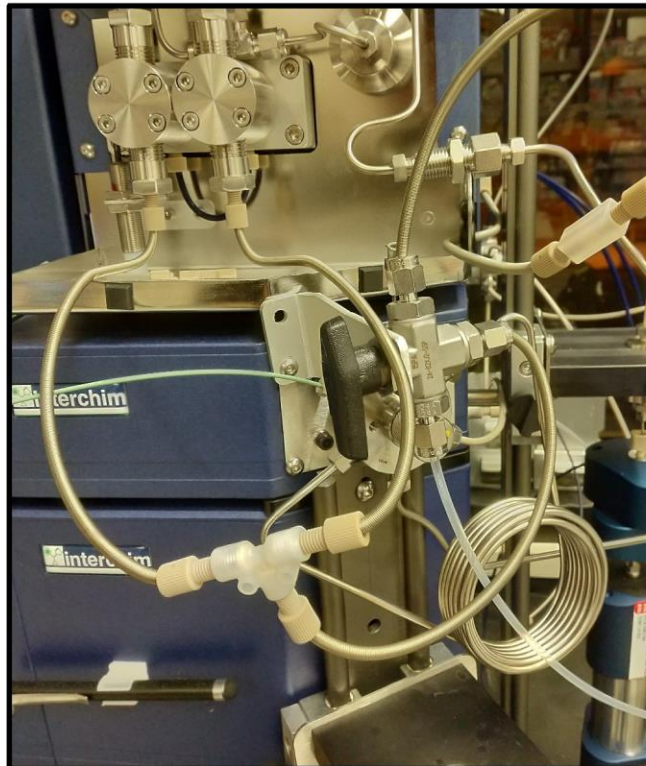
Supplementary Figure 6: Analytical HPLC trace of A β (1-40) WT crude. The analytes are detected by their absorbance (y-axis, mAU) at 214 nm. The gradient is 35 minutes (x-axis) long and the desired product elutes at **20.92** minutes. No significant side products were observed.



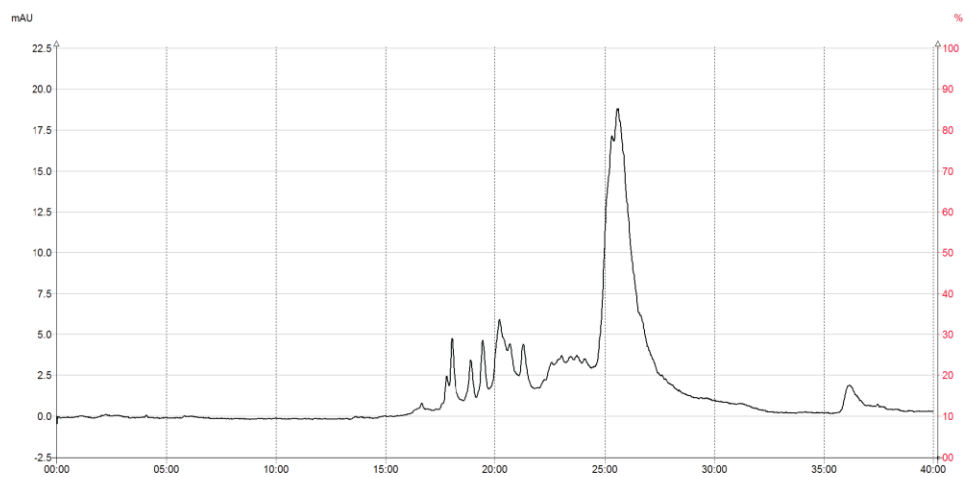
Supplementary Figure 7: HPLC trace of Purification 1 of the A β (1-40) WT crude peptide using an ammonium hydroxide buffer system. Peptides are detected by their absorbance (y-axis, mAU) at 214 nm as they pass through the flow cell over time (x-axis, minutes).



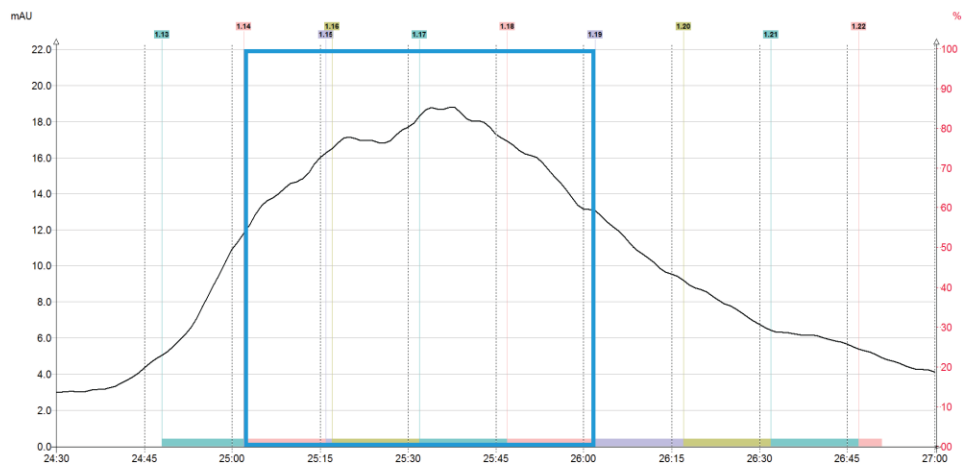
Supplementary Figure 8: Zoomed view of main A β (1-40) WT product peak in Supplementary Figure 7. Collected fraction(s) are indicated by the boxed region. The analytes are detected by their absorbance (y-axis, mAU) at 214 nm as they pass through the flow cell over time (x-axis, minutes).



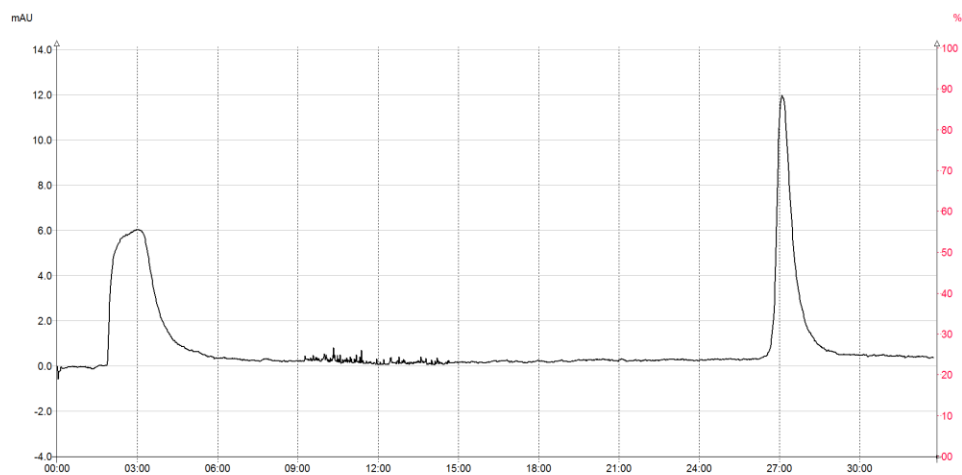
Supplementary Figure 9: Modification of Interchim puriFlash® PrepLC with an additional T-valve allows for direct loading of sample solution to column through pump.



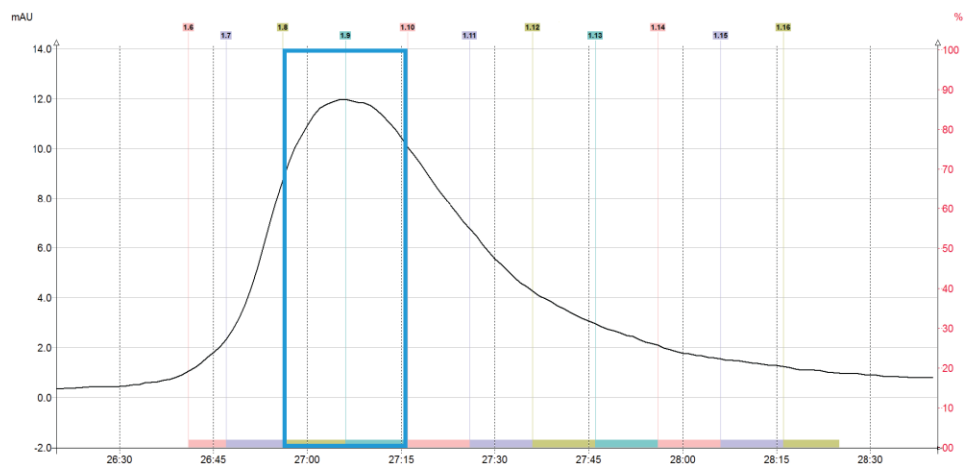
Supplementary Figure 10: HPLC trace of Purification 2 of A β (1-40) WT using a perchloric acid buffer system. The analytes are detected by their absorbance (y-axis, mAU) at 214 nm as they pass through the flow cell over time (x-axis, minutes).



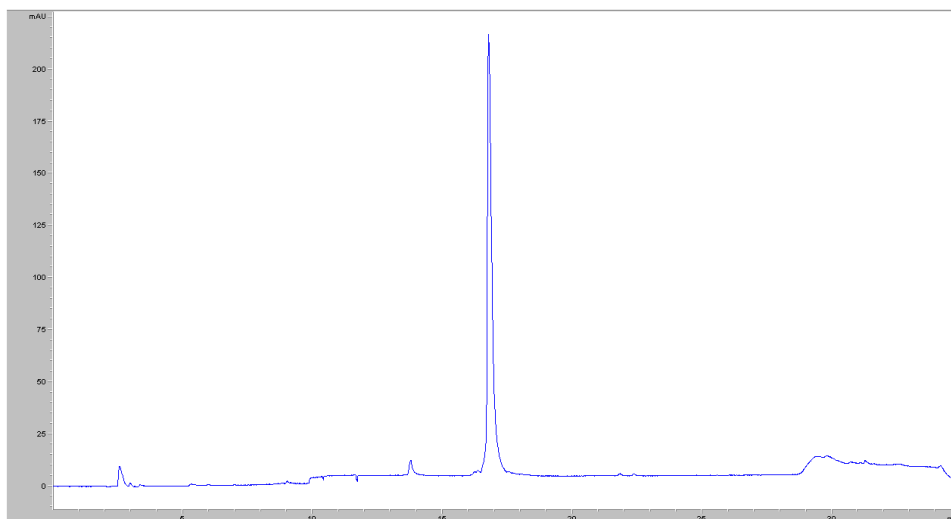
Supplementary Figure 11: Zoomed view of main A β (1-40) WT product peak in Supplementary Figure 10. Collected fraction(s) indicated by boxed region. The analytes are detected by their absorbance (y-axis, mAU) at 214 nm as they pass through the flow cell over time (x-axis, minutes).



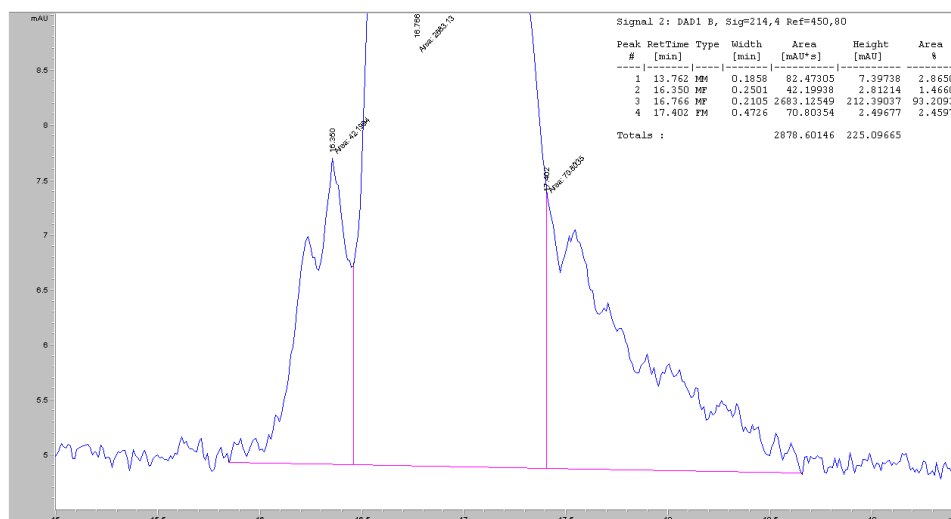
Supplementary Figure 12: HPLC trace of Purification 3 of A β (1-40) WT using a perchloric acid buffer system. The analytes are detected by their absorbance (y-axis, mAU) at 214 nm as they pass through the flow cell over time (x-axis, minutes).



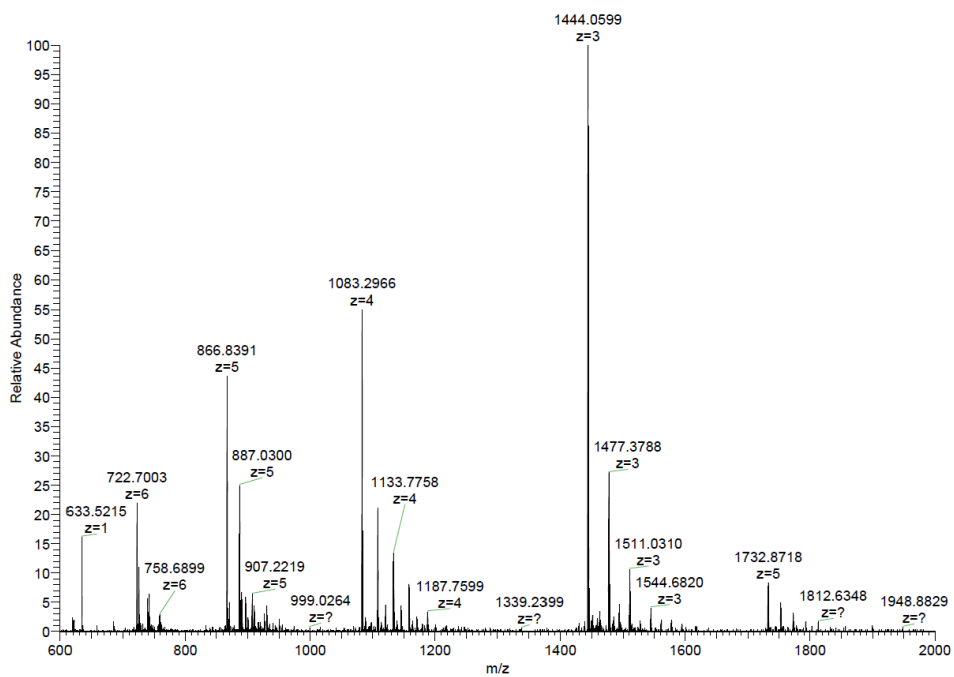
Supplementary Figure 13: Zoomed view of main A β (1-40) WT product peak in Supplementary Figure 12. Collected fraction(s) indicated by boxed region. The analytes are detected by their absorbance (y-axis, mAU) at 214 nm as they pass through the flow cell over time (x-axis, minutes).



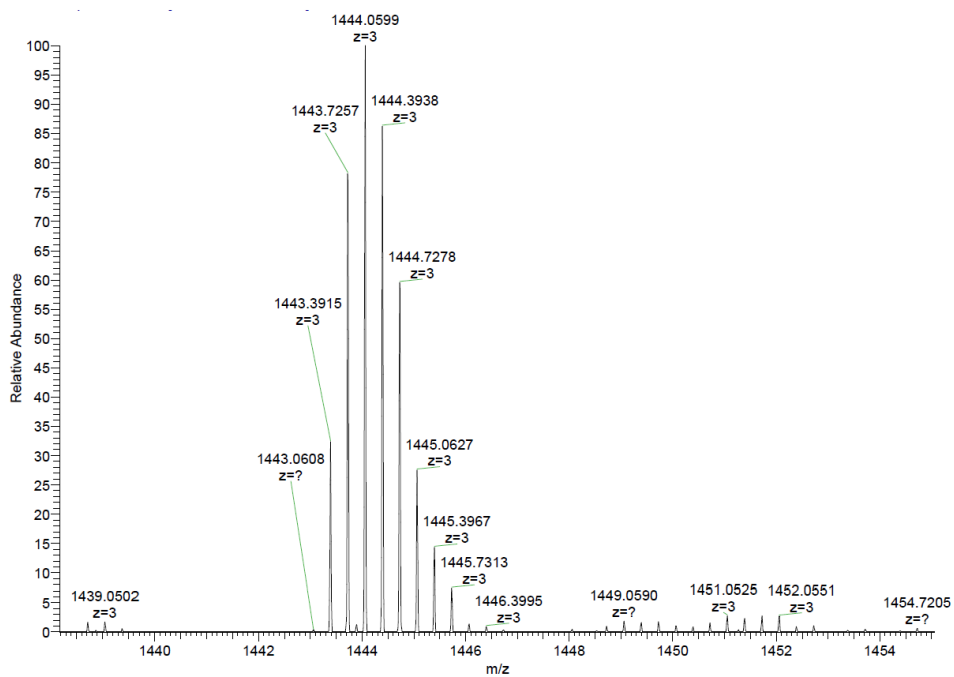
Supplementary Figure 14: Analytical HPLC trace of the purest lyophilized fraction for the second A β (1-40) WT purification. The analytes are detected by their absorbance (y-axis, mAU) at 214 nm as they pass through the flow cell over time (x-axis, minutes).



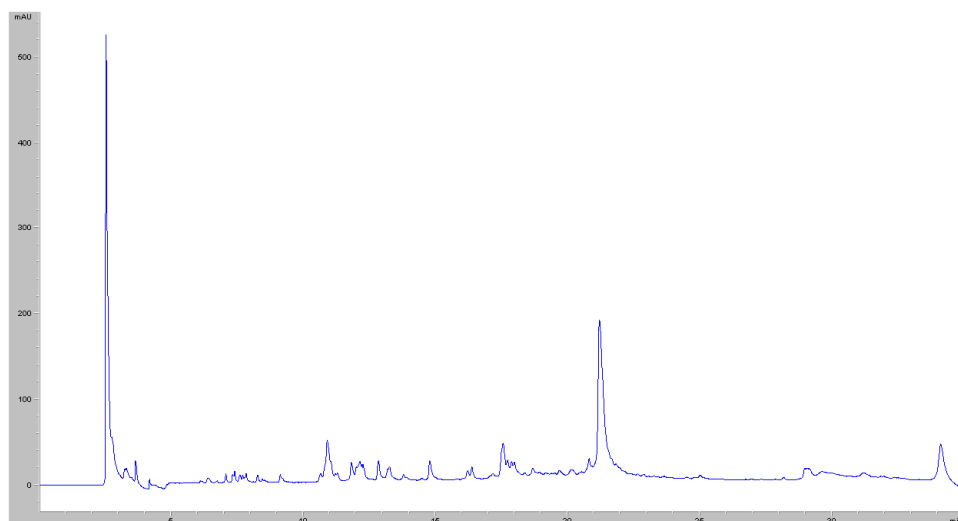
Supplementary Figure 15: Manual integration of the UV trace shown in Supplementary Figure 14 obtained by analytical HPLC. t_R 13.76: 82.47 mAU² (2.87%); t_R 16.35: 42.20 mAU² (1.47%); t_R 16.77: 2683.13 mAU² (93.20%); t_R 17.40: 70.80 mAU² (2.46%).



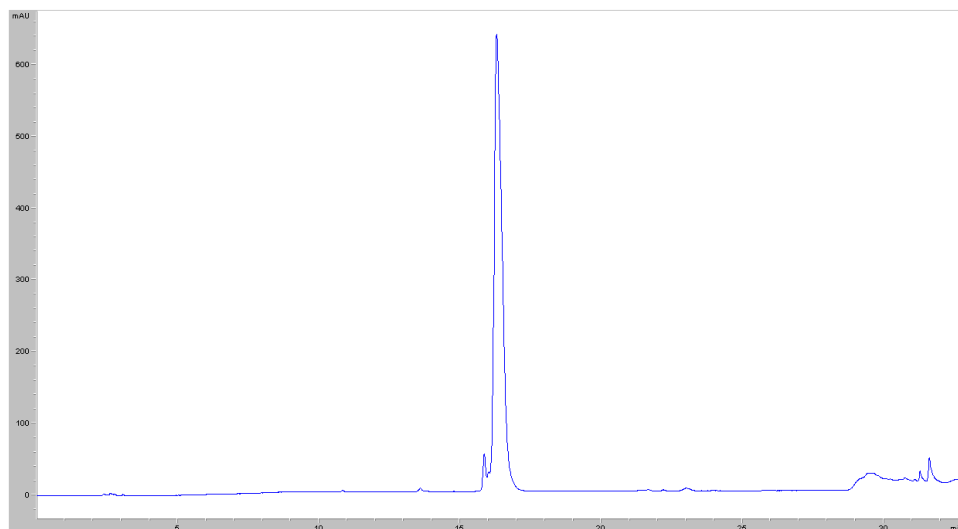
Supplementary Figure 16: Broadband mass spectrum of purified A β (1-40) WT collected by direct injection. The scan range was 600-2000 (m/z), and the population of each ion is represented by relative abundance.



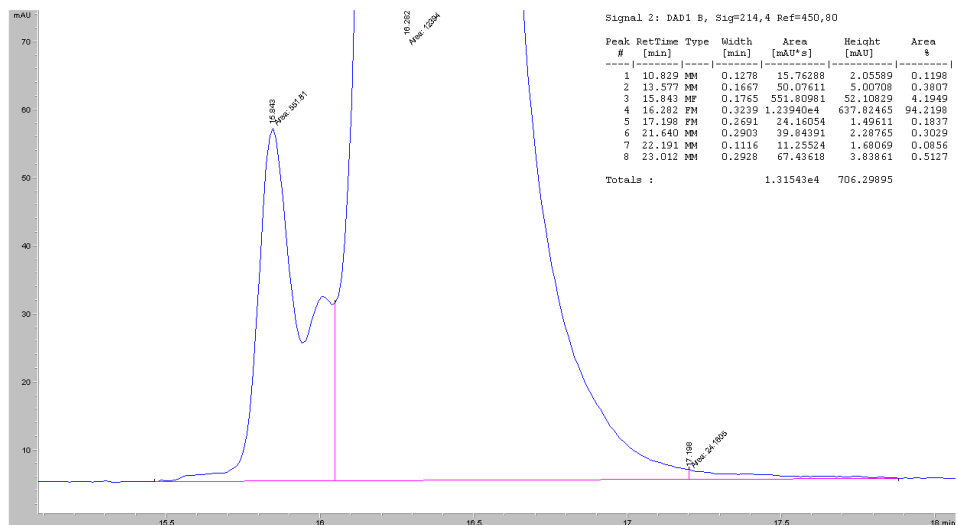
Supplementary Figure 17: Zoomed view of main Aβ (1-40) WT peak in broadband spectrum shown in Supplementary Figure 16. The monoisotopic $[M+3H]^{3+}$ is 1443.3915.



Supplementary Figure 18: Analytical HPLC trace of A β (1-40) IsoAsp23 crude peptide using the TFA solvent system. The analytes are detected by their absorbance (y-axis, mAU) at 214 nm. The gradient is 35 minutes (x-axis) long and the desired product elutes at **21.20** minutes. No significant side products were observed.

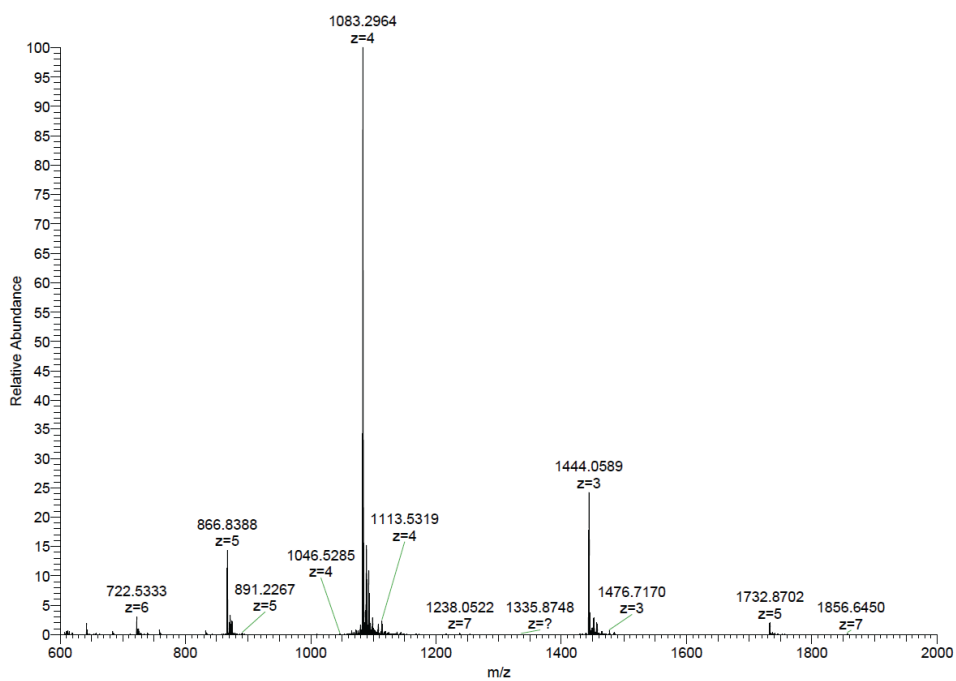


Supplementary Figure 19: Analytical HPLC trace of the purest lyophilized fraction for A β (1-40) IsoAsp23 from material that was subjected to the three preparative HPLC steps described in Supplementary Table 3. The analytes are detected by their absorbance (y-axis, mAU) at 214 nm as they pass through the flow cell over time (x-axis, minutes).

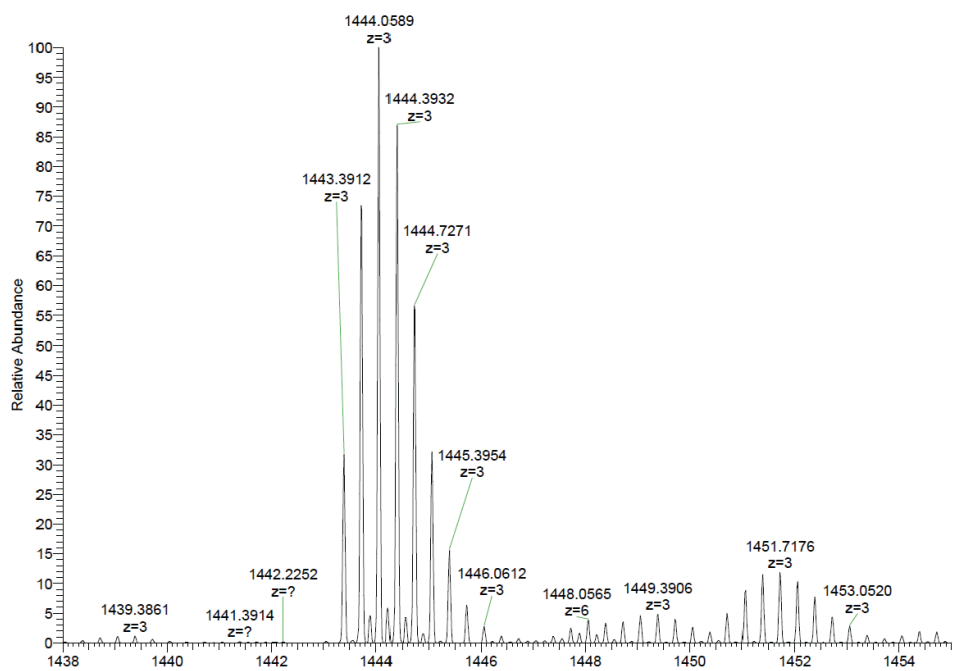


Supplementary Figure 20: Manual integration of the UV trace obtained shown in

Supplementary Figure 19. t_R 10.83: 15.76 mAU² (0.12%); t_R 13.58: 50.08 mAU² (0.38%); t_R 15.84: 551.81 mAU² (4.19%); t_R 16.28: 12394.0 mAU² (94.22%); t_R 17.20: 24.16 mAU² (0.18%); t_R 21.64: 39.84 mAU² (0.30%); t_R 22.19: 11.26 mAU² (0.09%); t_R 23.01: 67.44 mAU² (0.51%).



Supplementary Figure 21: Broadband mass spectrum of purified A β (1-40) IsoAsp23 collected by direct injection. The scan range was 600-2000 (m/z), and the population of each ion is represented by relative abundance.



Supplementary Figure 22: Zoomed view of main Aβ (1-40) IsoAsp23 peak in broadband spectrum shown in Supplementary Figure 21. The monoisotopic $[M+3H]^{3+}$ is 1443.3912.

Supplementary References

1. Patananan, A.N., Capri, J., Whitelegge, J.P., Clarke, S.G. Non-repair Pathways for Minimizing Protein Isoaspartyl Damage in the Yeast *Saccharomyces cerevisiae*. *J. Biol. Chem.* **289**, 16936-16953 (2014).
2. Holm, L., Laasko, L. Dali server update. *Nucleic acids research.* **44**, W351-355 (2016).
3. Garnham, C.P., Campbell, R.L., Davies, P.L. Anchored clathrate waters bind antifreeze proteins to ice. *Proc Natl Acad Sci.* **108**, 7363-7367 (2011).
4. Fitzpatrick, A.W.P et al. Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature.* **547**, 185-190 (2017).
5. Li, B. et al. Cryo-EM of full-length α -synuclein reveals fibril polymorphs with a common structural kernel. *Nat Commun.* **9**, 3609 (2018).
6. Wälti, M.A. et al. Atomic-resolution structure of a disease-relevant A β (1-42) amyloid fibril. *Proc Natl Acad Sci USA.* **113**, E4976-4984 (2016).
7. Sgourakis, N.G., Yau, W.M., Qiang, W. Modeling an in-register, parallel "iowa" $\alpha\beta$ fibril structure using solid-state NMR data from labeled samples with rosetta. *Structure.* **23**, 216-227 (2015).
8. Xiao, Y. et al. A β (1-42) fibril structure illuminates self-recognition and replication of amyloid in Alzheimer's disease. *Nat Struct Mol Biol.* **22**, 499-505 (2015).
9. Paravastu, A.K., Leapman, R.D., Yau, W.M., Tycko, R. Molecular structural basis for polymorphism in Alzheimer's beta-amyloid fibrils. *Proc Natl Acad Sci USA.* **105**, 18349-18354 (2008).
10. Lührs, T. et al. 3D structure of Alzheimer's amyloid-beta(1-42) fibrils. *Proc Natl Acad Sci USA.* **102**, 17342-17347 (2005).
11. Colvin, M.T. et al., Atomic Resolution Structure of Monomorphic A β 42 Amyloid Fibrils. *J Am Chem Soc.* **138**, 9663-9674 (2016).
12. Gremer, L. et al. Fibril structure of amyloid- β (1-42) by cryo-electron microscopy. *Science.* **358**, 116-119 (2017).
13. Schütz, A.K. et al. Atomic-resolution three-dimensional structure of amyloid β fibrils bearing the Osaka mutation. *Angew Chem Int Ed Engl.* **54**, 331-335 (2015).

Appendix B Preface

In this work, fast polymerizing fjord-edge graphene nanoribbons were synthesized and characterized by Li, Y. et al. A structure of one crystalline intermediate was determined by traditional X-ray diffraction. This proved an arduous task as the polymerization of the diyne precursor could be initiated with ambient light or other mild conditions (Figure 1).

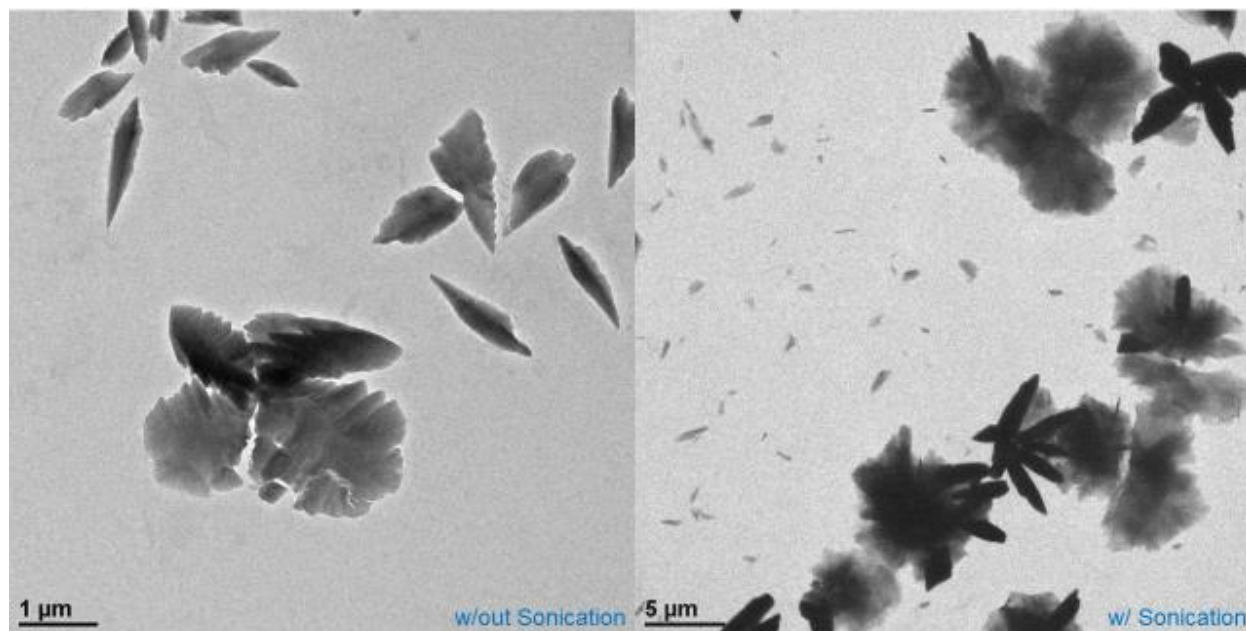


Figure 1: Electron micrograph of crystals obtained from recrystallization in methanol. When deposited on EM grids without sonication, individual crystals were discernable from each other (left). However, if the suspended crystals were sonicated prior to deposition, the crystals appear to undergo polymerization and aggregation into star-shaped clumps (right).

It was originally thought that electron diffraction would be necessary to obtain a structure for the crystalline precursor. Fortunately, well-ordered crystals of appropriate size were grown and traditional X-ray techniques were used to determine the diyne crystal packing prior to topochemical polymerization. Our structure confirmed the expected intermolecular distances between the triple bonded carbons.

Appendix B

Fjord-Edge Graphene Nanoribbons with Site-Specific Nitrogen Substitution

The following is a reprint of a research article from

Journal of the American Chemical Society

142(42), 18093-18102 (2020)

DOI: 10.1021/jacs.0c07657

Reprinted by permission from American Chemical Society.

© 2020 American Chemical Society

Fjord-Edge Graphene Nanoribbons with Site-Specific Nitrogen Substitution

Yolanda L. Li, Chih-Te Zee, Janice B. Lin, Victoria M. Basile, Mit Muni, Maria D. Flores, Julen Munárriz, Richard B. Kaner, Anastassia N. Alexandrova, K. N. Houk, Sarah H. Tolbert, and Yves Rubin*

Cite This: *J. Am. Chem. Soc.* 2020, 142, 18093–18102

Read Online

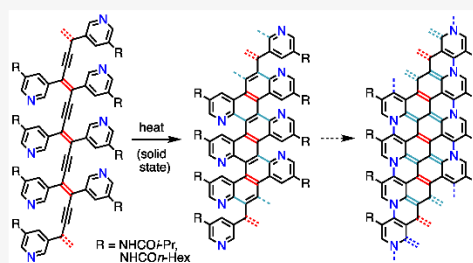
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: The synthesis of graphene nanoribbons (GNRs) that contain site-specifically substituted backbone heteroatoms is one of the essential goals that must be achieved in order to control the electronic properties of these next generation organic materials. We have exploited our recently reported solid-state topochemical polymerization/cyclization-aromatization strategy to convert the simple 1,4-bis(3-pyridyl)butadiynes **3a,b** into the fjord-edge nitrogen-doped graphene nanoribbon structures **1a,b** (fjord-edge $N_2[8]$ GNRs). Structural assignments are confirmed by CP/MAS ^{13}C NMR, Raman, and XPS spectroscopy. The fjord-edge $N_2[8]$ GNRs **1a,b** are promising precursors for the novel backbone nitrogen-substituted $N_2[8]_A$ GNRs **2a,b**. Geometry and band calculations on $N_2[8]_A$ GNR **2c** indicate that this class of nanoribbons should have unusual bonding topology and metallicity.



INTRODUCTION

Graphene nanoribbons (GNRs) are expected to usher in the ultimate nanosizing of electronics^{1–4} and sensors^{5,6} for next generation devices. The electronic properties of GNRs are exquisitely tuned by modification of their width, backbone, and edge structure.^{1,7–10} In the past decade, both on-surface and in-solution bottom-up syntheses have achieved precise structural control over these benchmarks.^{11–15} Early bottom-up syntheses have focused on GNRs with armchair^{16–20} or zigzag²¹ edges. More recently, intricate edge or interior configurations, such as chevron,^{11,22–24} cove,^{25–27} fjord,²⁸ or holey,^{29–31} have been obtained. These novel topologies significantly alter the electronic or magnetic properties of GNRs, as do atomically precise³² substitutions of carbons with heteroatoms such as boron,^{33,34} sulfur,^{35,36} or nitrogen.^{30,37,38} Crucially, site-specific doping at the GNR backbone produces a dramatic alteration of its electronics, making such structures the most desirable targets for synthesis.^{32,39–41} Nitrogen doped GNRs are of particular interest as they produce p-doped materials.^{24,37,38,42–45}

Herein, we describe the synthesis of the first eight-atom wide, fjord-edge nitrogen-doped graphene nanoribbons (fjord-edge $N_2[8]$ GNRs **1a,b**; Figure 1). Fjord-edge $N_2[8]$ GNRs **1a,b** were obtained in a facile two-step conversion starting from dipyrindyl diynes **3a,b**. Photochemically induced topochemical polymerization in the crystalline state afforded polydiacetylenes (PDAs) **4a,b**, which were thermally converted to GNRs **1a,b** with no loss of the side chains. The Hopf

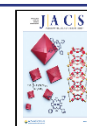
cyclization and ensuing aromatization from PDAs **4a,b** to GNRs **1a,b** was monitored by cross-polarization magic angle spinning (CP/MAS) solid state ^{13}C NMR. X-ray photoelectron spectroscopy (XPS) revealed both the pyridinic and amide bonding states of the nitrogen atoms. Raman spectroscopy further confirmed the structural integrity of the fjord-edge $N_2[8]$ GNRs **1a,b**.

RESULTS AND DISCUSSION

Diyne Monomer Synthesis. The topochemical polymerization of diynes requires suitable packing of the monomers in the crystal to trigger subsequent chain reactions.^{46–48} Here, the dipyrindyl diyne units of **3a,b,d** (Scheme 1, Figure 1) needed to have the diyne 1,4'-carbons within van-der-Waals contact distance (~ 3.5 Å) to promote facile formation of intermolecular bonds, a process that often occurs under ambient light.^{49,50} Although we synthesized several isomeric dipyrindyl diyne systems,⁵¹ only one series, based on 3-amino-5-alkynylpyridine, gave the polymerizable diynes **3a,b**. Accordingly, 3-amino-5-bromopyridine was coupled with trimethylsilylacetylene under Sonogashira conditions, followed by

Received: July 16, 2020

Published: September 8, 2020



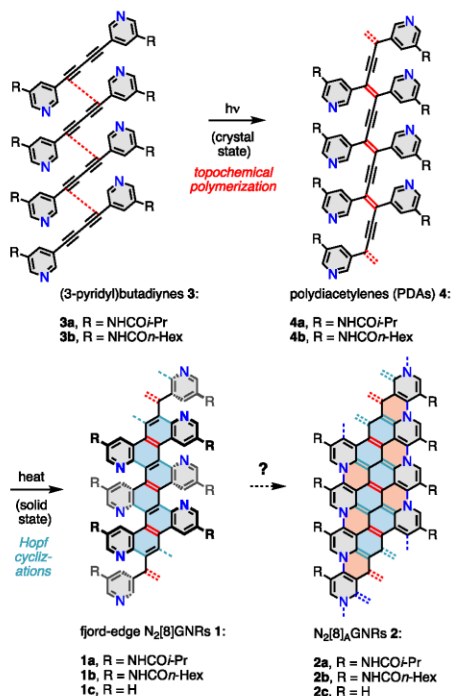
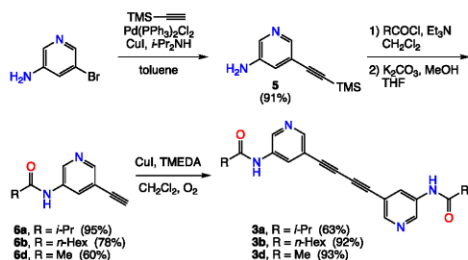


Figure 1. Synthesis of fjord-edge nitrogen-doped graphene nanoribbons (fjord-edge N₂[8]GNRs **1a,b**).

Scheme 1. Synthesis of the 1,4-Bis(3-pyridyl)butadiynes **3a,b,d**.



acylation of amine **5** with the corresponding acid chlorides (Scheme 1, R = *i*-Pr, *n*-Hex, Me). Removal of the trimethylsilyl protecting group gave alkyne amides **6a,b,d** in good to excellent yields. Oxidative coupling under the Hay conditions afforded diyne amides **3a,b,d** in good to high yields.

X-ray Structure. Crude diyne **3a** afforded single crystals after slow evaporation from methanol (Figure S28a,b).⁵¹ X-ray diffraction at the Advanced Photon Source (APS) of the Argonne National Lab (Figure S29) afforded a 1.0 Å resolution crystal structure of diyne **3a** (Figure 2, Table S1).⁵¹

The crystal packing geometry for molecules of diyne **3a** validates the desired short, nonbonded C1–C4' distance of 3.45 Å (Figure 2a). The hydrogen bonds between the carbonyl

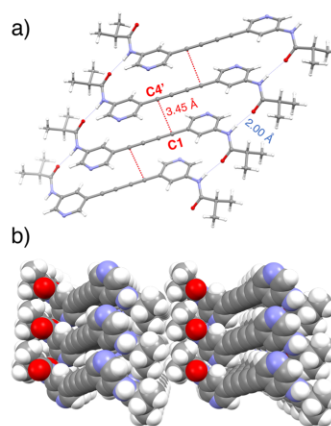


Figure 2. (a) Crystal packing structure for diyne **3a** displaying the short C1–C4' distance directed by the C=O...H–N hydrogen-bonded network. (b) View of **3a** down the H-bonding axis.

oxygens and amide hydrogens have an optimal distance of 2.00 Å, guiding the assembly of diyne units in **3a** along the unit cell vector *a*. The relative strength of these intermolecular interactions is reflected in the crystal morphology and powder diffraction (Figures S28 and S30). To accommodate the H-bonding motif, the polymer growth axis exhibits a horizontal offset between each molecule, organizing the diynes into an optimal arrangement for topochemical polymerization (Figure 2b). While we were not able to obtain the single crystal structure of **3b**, its powder diffraction displayed a similar packing arrangement to **3a** (Figure S31).

Topochemical Polymerization of Diynes 3a,b. Both dipyridyl diynes **3a,b** quickly polymerized to dipyridyl PDAs **4a,b** when subjected to UV light, as well as under ambient light, whereas diyne **3c** was unreactive. The polymerizations were carried out by the irradiation of finely pulverized dispersions of the crystals in hexanes using a medium pressure Hanovia lamp (Pyrex filter), typically for 12 h, producing a deep purple/black material. Dissolution of unreacted monomer from the polymerized crystals gave the nearly insoluble, pristine polydiacetylenes **4a,b** (18 and 4%, respectively) as fibrous powders after filtration. The low polymerization yield for **3b** appears to be inherent to this derivative, since repeated attempts to increase yields by using nanocrystalline material could not raise the conversion yield for this substrate.

Conversion of PDAs 4a,b to GNRs 1a,b. Thermal conversion experiments were carried out in separate runs on PDAs **4a,b** under increasingly higher temperatures in argon.⁵¹ This transformation could be conveniently monitored by CP/MAS ¹³C solid state NMR, focusing on the four distinct carbon signal ranges corresponding to the four functional groups of interest (Figure 3b,c): amide carbonyls (160–170 ppm), aromatic carbons (110–150 ppm), alkyne carbons (~100 ppm), and amide side chains (10–40 ppm). As the PDAs **4a,b** were heated under increasingly higher temperatures (1h each per run), the distinct ¹³C NMR signals tracked an initial Hopf cyclization, as evidenced by the disappearance of the alkyne peak at temperatures between 300 and 350 °C (Figure 3b,c). Thus, the Hopf cyclization reactions of PDAs **4a,b** occur more

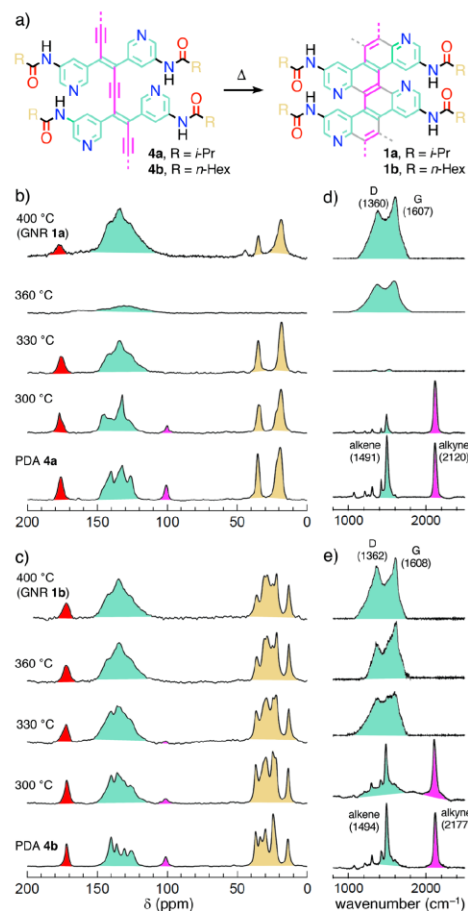


Figure 3. CP/MAS solid-state ^{13}C NMR and Raman spectra for the products obtained by heating PDAs **4a,b**. Each heating experiment was carried out for 1 h using a fresh sample of PDA. (a) PDAs **4a,b** and GNRs **1a,b** are color-coded by chemical shift region to reflect spectral changes in parts b and c. (b) PDA **4a** conversion to fjord-edge $\text{N}_2[8]$ GNR **1a**, and (c) PDA **4b** conversion to fjord-edge $\text{N}_2[8]$ GNR **1b**. The very broad, weak signal for the sample heated at 360 °C in part b is due to an air-stable π -radical intermediate formed during heating.⁵² Baseline-corrected Raman spectra for the conversion of (d) PDA **4a** to fjord-edge $\text{N}_2[8]$ GNR **1a**, and (e) PDA **4b** to fjord-edge $\text{N}_2[8]$ GNR **1b**. The very weak Raman signal for the sample heated at 330 °C in part d is due to the strong background fluorescence of this sample, which was smoothed out through baseline correction.

readily, ~ 100 °C lower than for our phenyl analogs.⁵⁰ The Hopf cyclization step is followed by further aromatization reactions that form fjord-edge $\text{N}_2[8]$ GNRs **1a** and **1b** at temperatures between 350 and 400 °C (Figure 3b,c), as revealed by the changes in the overall envelope for the aromatic signals between 110–150 ppm, which adopt an underlying intensity ratio of 1:2:1 for both **1a** and **1b** (Figure

3b,c, Table 1a). Curve fitting of the experimental spectrum of **1b** in the 110–150 ppm range with seven Gaussian curves of equal intensity and width, representing the expected number of aromatic ^{13}C signals for fjord-edge $\text{N}_2[8]$ GNR **1b**, affords the fitted peaks in Table 1. These values compare rather well with the calculated values for model compound **1e** (Table 1b, top). Furthermore, the clusters of peaks for the aromatic carbons for each of the alternate possible model structures, i.e., **7c** and **8c** (Table 1b, middle and bottom), which are the structural alternatives in the conversion of polydiacetylenes **4a,b** to fjord-edge $\text{N}_2[8]$ GNRs **1a,b** (see discussion below and Figure 5), do not fit the experimental curve as well. In particular, structure **7c** has its A,B peaks clustered around 144 ppm, which leaves a large empty void between these peaks and the C–G peaks. This gap is even wider for structure **8c**, which has a ~ 20 ppm gap between the clusters of A,B and C–G peaks. Thus, the experimental CP/MAS ^{13}C solid state NMR spectrum of fjord-edge $\text{N}_2[8]$ GNR **1b** matches best the calculated chemical shifts of model structure **1e**.

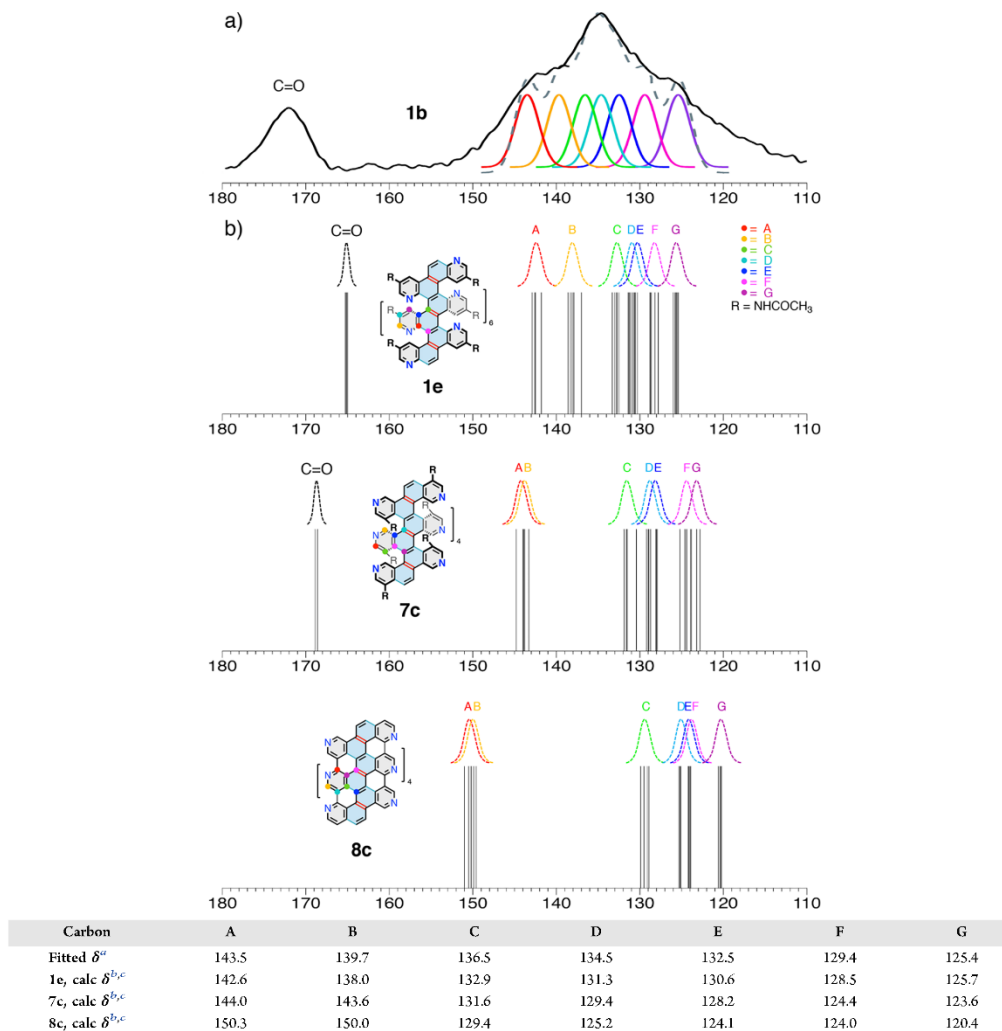
Furthermore, the convergence of both PDAs **4a** and **4b** to identical aromatic peak shapes upon heating indicates that they both likely undergo similar processes to form the same fjord-edge $\text{N}_2[8]$ GNR core. Subsequent heating of **4a** and **4b** to temperatures as high as 400 °C shows no change in the ^{13}C CP/MAS spectra, suggesting that the fjord-edge $\text{N}_2[8]$ GNRs **1a,b** have fully formed at 330 and 360 °C. These lower conversion temperatures prevent side chain fragmentation, unlike in our previous work on $[8]_A$ GNR, which required temperatures as high as 500 °C for full conversion.⁵⁰

As expected, the presence of side chains in fjord-edge $\text{N}_2[8]$ GNRs **1a,b** increases their solubility to a small extent: sonication in strongly hydrogen-bonding *N*-methyl-2-pyrrolidone (NMP), followed by filtration through a 0.2 μm Teflon membrane, gives yellowish solutions of fjord-edge $\text{N}_2[8]$ GNRs **1a,b** (UV–vis, see Figure S27).⁵¹

Raman Spectroscopy. The precursor PDAs **4a,b** both exhibit strong alkene peaks at 1494 and 1491 cm^{-1} , respectively, as well as alkyne peaks at 2117 and 2120 cm^{-1} , respectively, which are typical of the enyne backbone (Figure 3d,e, bottom traces). Upon heating, these signals disappear, while the signature D and G peaks of GNRs **1a,b** appear (Figure 3d,e, top traces). The Raman spectra for fjord-edge $\text{N}_2[8]$ GNRs **1a,b** show D peaks at 1360 and 1362 cm^{-1} , and G peaks at 1607 and 1608 cm^{-1} , respectively. The G peaks of these GNRs are upshifted by ~ 30 cm^{-1} from graphene (1580 cm^{-1}), due to the confined, strongly aggregated nature of GNRs **1a,b** (see HR-TEM imaging, Figure S32).⁵¹ Prior reports on GNRs show similar shifts in the D peak as well.⁵³ Furthermore, the broad feature of the observed D peaks can be attributed to the fjord-edge structure and the high levels of site-specific heteroatom substitutions, as noted below.

Broadening of the D peak is generally caused by defects within the graphene lattice.^{54–56} The fjord-edge GNRs **1a,b** have two inherent structural features that broaden the D peak beyond previously reported GNR examples. Specifically, fjord-edges represent bond vacancies along the edges of pristine graphene. Broad D and G peaks are characteristic of a large numbers of defects, which can also be seen in the all-carbon fjord edge $[8]$ GNR.²⁸ Further broadening of the D peaks is caused by distortion of the lattice from the nitrogen dopant. Reports for both nitrogen doped graphene and top-down synthesized doped GNRs have shown broad D peaks with various levels of dopant atoms.^{57,58} The bottom-up approach

Table 1. Comparison of (a) the Curve-Fitted Experimental^a Spectrum of Fjord-Edge N₂[8]GNR 1b, and (b) Calculated ¹³C NMR Chemical Shifts for Model Structures 1e, 7c, and 8c.⁵¹



^aThe experimental CP/MAS ¹³C NMR spectrum is plotted as a solid black line, while the sum of the curve fits is a gray dashed line. ^bAverage of the DFT calculated chemical shifts (GIAO/B3LYP//6-31G(d)) for each of the seven distinct, nonsymmetrically related carbons defined in part b). Gaussian curves of arbitrary but constant width and height are displayed for better visualization of the groups of chemical shifts belonging to each of the seven types of nonsymmetrically related carbons for each structure. ^cSee Figures S39–S41.

of our synthesis incorporates higher levels of nitrogen doping, 10% by atom (based on core aromatic ring C,N,H atoms only), which likely further accounts for the broad D peak.⁵⁹

XPS Spectra. To confirm the formation of a fjord-edge topology in compounds 1a,b, we examined the nitrogen bonding environment present in GNRs 1a,b using XPS spectroscopy (Figure 4). GNRs 1a,b should contain solely pyridinic and amide nitrogens if the fjord edge structure is exclusively formed. As expected, the XPS spectra of GNRs 1a,b

show signatures corresponding to the pyridinic^{60,61} (398.7 eV) and amide⁶² (399.7 eV) bonding, and no other species (Figure 4). Notably, we do not see the presence of internal graphitic (401.7 eV)^{60,61,63,64} or pyridinium species (402.5 eV),^{65,66} which would indicate further cyclizations have occurred beyond the fjord edge structure (see discussion below and Figure 5). Additionally, the retention of the side chain substituents indicates that the alternate path of cyclization to an edge-like topology (8), via intermediates 7a,b (Figures 5

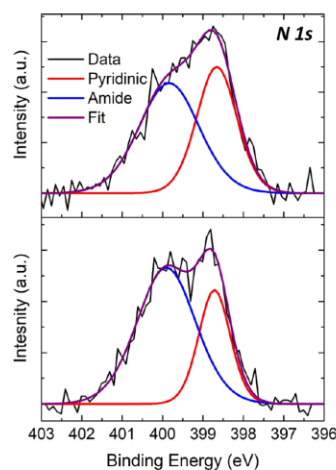


Figure 4. N 1s XPS spectra for fjord-edge $N_2[8]$ GNR **1a** and **1b**, with pyridinic and amide nitrogens centered at 398.7 and 399.7 eV, respectively.

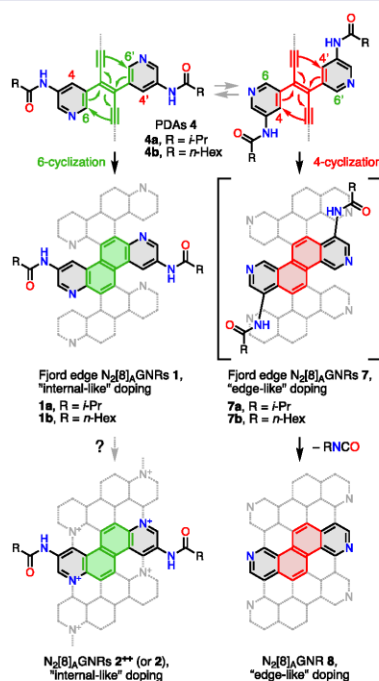


Figure 5. Two Hopf cyclization pathways for PDAs **4a,b**, which can occur at either the 4- or 6-positions of the pyridyl rings, resulting in the nitrogen atoms being located at either internal (**1a,b** or **2,2****) or at edge locations (**7a,b** or **8**).

and **6**), is unlikely to have occurred, in agreement with the calculations, which show this pathway to be highly unfavorable (see below). The increased width of the amide peak relative to the pyridinic peak is consistent with conformational disorder in the side-chains only. Together, these results all indicate that the structure most consistent with the data is that of the fjord-edge $N_2[8]$ GNRs **1a,b**.

Conformation Preference, Cyclization Pathways, and Reaction Barriers. The conformational preference in a model system for the fjord-edge $N_2[8]$ GNR **1**, R = NHCOMe) was investigated with its two extreme cases, all-zigzag and helical (see Section S4.8).⁵¹ The fjord edge structure of **1a,b** imparts strong steric repulsion between the pyridine nitrogen lone pairs and the C–H bonds of adjacent diazachrysene units. Thus, the two possible key conformations, all-zigzag (alternating up-down pyridyl units) and helical (all pyridyl units staggered in a nonalternating fashion) were calculated at the semiempirical (PM3) and DFT (B3LYP) levels. The differences in energy (35.9 and 60.2 kcal·mol⁻¹ for PM3 and B3LYP, respectively) between the two conformers is very high, thus it is likely that only the zigzag conformation exists in the fjord-edge $N_2[8]$ GNRs **1a,b** as represented in Figure 1.

There are two possible Hopf cyclization pathways for PDAs **4a,b** (Figure 5), which can afford fjord-edge GNRs with two different topologies. The internal-like topology (**1a,b**) has the nitrogen atoms opposite to C–H bonds of the next "diazachrysene" units, while the edge-like topology has them at the edges of the fjord-edge nanoribbons **7a,b**, or the ensuing $N_2[8]$ GNR **8**, since it can be expected that the amide side chains should be easily lost from structure **7a,b** under our heating conditions. Unlike our previous work on $[8]_A$ GNR,⁵⁰ which forms the same structure regardless of the initial cyclization pathway at the 2- or 4-positions of the PDAs' *m*-amidophenyl rings owing to subsequent thermally induced loss of sidechains, cyclization at either the similarly related 4- or 6-positions of the pyridyl rings in PDAs **4a,b** could give two different fjord-edge GNRs or a statistical mixture alternating both pathways along the nanoribbon length. However, cyclization at the 4-position should be strongly disfavored owing to the severe steric clash between the amide groups and adjacent pyridyl units during Hopf cyclization (structures **7a,b**, Figure 5). Furthermore, aromatization to the edge-doped armchair $N_2[8]$ GNR **8** should easily ensue if the edge-like pathway is followed, resulting in total loss of the side chains. This is not the case, based on our experimental data (CP/MAS solid-state ¹³C NMR and XPS, Figures 3 and 4). In fact, cyclization appears to favor the 6-position, which produces the lesser strained, internally doped fjord-edge $N_2[8]$ GNRs **1a,b** with their amide side chains pointing away from the series of fused diazachrysene units. The transition state calculations reported below lend further strong support for the fact that the Hopf cyclization occurs at the 6-pyridyl positions to yield fjord-edge $N_2[8]$ GNRs **1a,b**.

To further understand the energetics and pathway of the Hopf cyclization of PDAs **4a,b**, we base our theoretical considerations on previous results by Prall *et al.*⁶⁷ and our own work.⁵⁰ The Hopf cyclization mechanism has been calculated to proceed through an initial 6 π -electrocyclization, followed by two consecutive [1,2]-H shifts, with the first H-shift as the rate-determining step.⁶⁸ As discussed above (Figure 5), the enediyne units of PDAs **4a,b** can undergo cyclization at either the 6-position (*para* to the amide group) or the 4-position (*ortho* to the amide group), the latter of which is intuitively

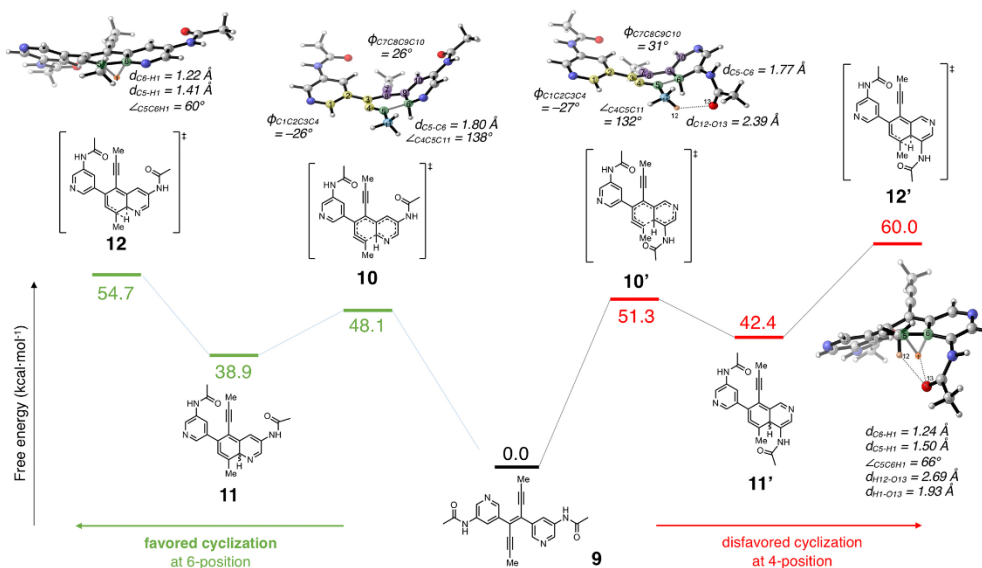


Figure 6. Free energies (in kcal·mol⁻¹) of the intermediates and transition states for both the favored (left) and disfavored (right) Hopf cyclization pathways relative to starting structure 9.

unfavorable due to the large steric bulk of the amide group, compared to only H in 1a,b (Figure 5).

Using density functional theory (DFT), we computed the geometries of the model system 9 (Figure 6), the transition states for the initial 6π electrocyclizations (10 and 10'), the strained allenic intermediates 11 and 11', and the transition states for the rate-determining 1,2-shifts (12 and 12'). These structures were optimized in the gas-phase using B3LYP/6-31G(d), and single-point energy calculations were performed using M06-2X/6-311+G(d,p) with B3LYP frequencies to obtain free energy values. The potential energy surfaces for the two cyclization pathways are shown in Figure 6.

The energetic trends for this bispyridyl system are similar to the all-carbon system previously studied by us.⁵⁰ As expected, the barriers for cyclization at the 4-position are higher than those at the 6-position. The activation free energies for the more favorable transition states 10 and 12 are 48.1 and 54.7 kcal·mol⁻¹, respectively, while the analogous transition states 10' and 12' have higher barriers of 51.3 and 60.0 kcal·mol⁻¹, respectively. Like the all-carbon PDAs previously studied, the 1,2-shift following electrocyclization is the rate-determining step in both pathways, and the preference for the pathway with cyclization at the 6-position is substantial (5.3 kcal·mol⁻¹). This energy difference virtually ensures that the analogous series of cyclizations within PDAs 4a,b should occur exclusively at the 6-positions (*para* to the amide group) of the pyridyl rings.

The geometries of the transition states 10 and 10' are similar to each other and to their all-carbon variants. In 10 and 10', the π -system of the alkynes is planar, while the flanking pyridyl groups are out-of-plane by approximately 30°. The slightly higher barrier of 10' is likely due to the close proximity of the amide oxygen and methyl group on the forming C–C bond.

More ample differences can be seen between the rate-determining states 12 and 12'. In 12, the quinoline intermediate is planar, and the C6–H1 bond stretches to 1.22 Å from its normal C–H bond length of 1.09 Å. However, in the less-favorable transition state 12', the bulky amide group forces the quinoline out of plane, largely due to unfavorable steric interactions between the amide oxygen and the shifting hydrogen and adjacent methyl group, which accounts for the intrinsic preference for the 6-position cyclization.

Calculations for the unsubstituted variant (no amide groups) were also computed to probe the intrinsic preference for the 6-position (Figure S35).⁵¹ The barriers of the rate-determining 1,2-shift for the unsubstituted system are 53.7 and 55.7 kcal·mol⁻¹ for the 6 and 4-positions, respectively, with less strong of a preference (2.0 kcal·mol⁻¹) for the 6-position. Thus, the increased preference for the 6-position in the substituted system 9 can be wholly attributed to the unfavorable steric interactions between substituents in the rate-determining transition state structure 12'.

We note that the barrier of the rate-determining step in pyridyl system 9 is ~ 3 kcal·mol⁻¹ lower than that of the all-carbon system, which correlates well with the ~ 100 °C lower conversion temperature for this N-based system.

Electronic Properties of Fully Cyclized N₂[8]_AGNR. In order to better understand the electronic properties of the fully cyclized (graphitic) N₂[8]_AGNR system 2c, we calculated the electronic Density of States (DOS) for the parent systems 1c and 2c by means of periodic DFT (Figures 1 and 7).⁵¹ The DOS of 1c and 2c, obtained from the HSE06 functional on the basis of PBE geometries,^{68,69} are depicted in Figure 7. As we can see, fjord-edge N₂[8]_AGNR 1c is expected to be a semiconductor, with a bandgap of 2.04 eV (Figure 7a). Fully cyclized N₂[8]_AGNR 2c clearly displays metallic behavior, as

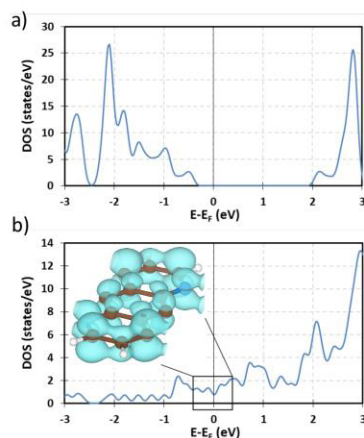


Figure 7. DOS plots calculated at the HSE06 level for (a) fjord-edge GNR **1c** and (b) fully cyclized $N_2[8]_A$ GNR **2c**. The inset shows a 3D representation of the partial charge density at the Fermi level in the $[-0.2, 0.2]$ eV range for **2c**.

indicated by the continuous DOS landscape at the Fermi level depicted in Figure 7b. Notice that the HSE06 results are in agreement with the results at the PBE level of theory (Figure S37).⁵¹ Namely, both predict **2c** to be metallic and **1c** to be a semiconductor, although the band gap predicted by PBE is smaller than by HSE06 (1.43 eV vs 2.04 eV, respectively), as expected from the DFT delocalization error.⁷⁰ Furthermore, bond-equalization of many of the carbon–carbon bonds within the structure of fully fused $N_2[8]_A$ GNR **2c**, unlike those in fjord edge structure **1c** (Figure S38),⁵¹ indicates that $N_2[8]_A$ GNR **2c** has quinoid character, which incites its metallic nature. This fact is in agreement with the charge density at the Fermi level, which exhibits a delocalized π -character (Figure 7b, inset).

These calculations indicate that experimental isolation of the fully cyclized, metallic $N_2[8]_A$ GNRs **2a,b** could be challenging, and despite our best efforts so far, this step is awaiting further work. Formation and characterization of these novel, exciting graphene nanoribbons is currently under investigation.

CONCLUSION

In summary, we have demonstrated the synthesis of a novel fjord-edge $N_2[8]$ GNR system with site-specific nitrogen substitution. The stepwise conversion from dipyrindyl diynes **3a,b** to the nitrogen doped, fjord edge $N_2[8]$ GNRs **1a,b** via topochemical polymerization of PDAs **4a,b**, followed by Hopf cyclizations to the GNRs, proceeded at relatively moderate temperatures of 330–360 °C. The formation of the fjord-edge structure in the GNRs **1a,b** was confirmed via CP/MAS ¹³C NMR, XPS, and Raman spectroscopy. The success of our bottom-up method demonstrates the versatility of the crystalline state topochemical polymerization method to incorporate heteroatom substitution and structural diversity into GNR structures.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/jacs.0c07657>.

Experimental procedures and spectroscopic characterization data, crystallization procedures, HR-TEM imaging, and details of the DFT and Molecular Mechanics calculations (PDF)

Crystallographic data (CIF)

AUTHOR INFORMATION

Corresponding Author

Yves Rubin – Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095–1567, United States; orcid.org/0000-0003-0187-9689; Email: rubin@chem.ucla.edu

Authors

Yolanda L. Li – Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095–1567, United States; orcid.org/0000-0002-2180-3086

Chih-Te Zee – Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095–1567, United States

Janice B. Lin – Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095–1567, United States; orcid.org/0000-0002-9333-5760

Victoria M. Basile – Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095–1567, United States

Mit Muni – Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095–1567, United States

Maria D. Flores – Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095–1567, United States

Julen Munárriz – Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095–1567, United States; orcid.org/0000-0001-6089-6126

Richard B. Kaner – Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095–1567, United States; orcid.org/0000-0003-0345-4924

Anastassia N. Alexandrova – Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095–1567, United States; orcid.org/0000-0002-3003-1911

K. N. Houk – Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095–1567, United States; orcid.org/0000-0002-8387-5261

Sarah H. Tolbert – Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095–1567, United States; orcid.org/0000-0001-9969-1582

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/jacs.0c07657>

<https://dx.doi.org/10.1021/jacs.0c07657>
J. Am. Chem. Soc. 2020, 142, 18093–18102

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by grants from the National Science Foundation to Y.R. (NSF-CHE-1808836), K.N.H. (NSF-DMR-1335645 and NSF-CHE-1254897), A.N.A. (NSF-CHE-1351968), S.H.T. and Y.R. (NSF-CHE-1608957) and NSF-MRI-1532232. This work used computational and storage services associated with the Hoffman2 cluster provided by the UCLA Institute for Digital Research and Education's Research Technology Group, the NSF-supported Extreme Science and Engineering Discovery Environment (XSEDE) (NSF-OCI-1054575 and TG-CHE160054), and high-performance research computing resources by Texas A&M University (<http://hprc.tamu.edu>). M.D.F. is funded by a Ruth L. Kirschstein National Research Service Award GM007185 and a National Science Foundation Graduate Research Fellowship. C.-T.Z. is funded by a George Gregory Fellowship. R.B.K. thanks the Dr. Myung Ki Hong Chair in Materials Innovation. Data collection at the Argonne National Lab was funded by DOE grant no. DE-FC02-02ER63421. We thank Drs. Duilio Cascio and Michael R. Sawaya for organizing the single crystal data collection at the Advanced Photon Source of Argonne National Laboratory, Dr. Ta-Chung Ong and Dr. Robert Taylor for their assistance with the CP/MAS NMR experiments, and Prof. Jose A. Rodriguez for advice and assistance with the crystal data collection and structure analysis.

REFERENCES

- (1) Zhou, X.; Yu, G. Modified Engineering of Graphene Nanoribbons Prepared via On-Surface Synthesis. *Adv. Mater.* **2020**, *32*, 1905957.
- (2) Llinas, J. P.; Fairbrother, A.; Barin, G. B.; Shi, W.; Lee, K.; Wu, S.; Choi, B. Y.; Braganza, R.; Lear, J.; Kau, N.; Choi, W.; Chen, C.; Pedramrazi, Z.; Dumlaff, T.; Narita, A.; Feng, X.; Müllen, K.; Fischer, E.; Zettl, A.; Ruffieux, P.; Yablonovitch, E.; Crommie, M.; Fasel, R.; Bokor, J. Short-Channel Field-Effect Transistors with 9-Atom and 13-Atom Wide Graphene Nanoribbons. *Nat. Commun.* **2017**, *8*, 633.
- (3) Li, X.; Wang, X.; Zhang, L.; Lee, S.; Dai, H. Chemically Derived, Ultrasoft Graphene Nanoribbon Semiconductors. *Science* **2008**, *319*, 1229–1232.
- (4) Shen, H.; Shi, Y.; Wang, X. Synthesis, Charge Transport and Device Applications of Graphene Nanoribbons. *Synth. Met.* **2015**, *210*, 109–122.
- (5) Rajaji, U.; Arumugam, R.; Chen, S.-M.; Chen, T.-W.; Tseng, T.-W.; Chinnapaiyan, S.; Lee, S.-Y.; Chang, W.-H. Graphene Nanoribbons in Electrochemical Sensors and Biosensors: A Review. *Int. J. Electrochem. Sci.* **2018**, *13*, 6643–6654.
- (6) Cho, K. M.; Cho, S.-Y.; Chong, S.; Koh, H.-J.; Kim, D. W.; Kim, J.; Jung, H.-T. Edge-Functionalized Graphene Nanoribbon Chemical Sensor: Comparison with Carbon Nanotube and Graphene. *ACS Appl. Mater. Interfaces* **2018**, *10*, 42905–42914.
- (7) Nakada, K.; Fujita, M.; Dresselhaus, G.; Dresselhaus, M. S. Edge State in Graphene Ribbons: Nanometer Size Effect and Edge Shape Dependence. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 17954–17961.
- (8) Fujita, M.; Wakabayashi, K.; Nakada, K.; Kusakabe, K. Peculiar Localized State at Zigzag Graphite Edge. *J. Phys. Soc. Jpn.* **1996**, *65*, 1920–1923.
- (9) Hao, Z.; Zhang, H.; Ruan, Z.; Yan, C.; Lu, J.; Cai, J. Tuning Electronic Properties of Atomically Precise Graphene Nanoribbons by Bottom-up Fabrications. *ChemNanoMat* **2020**, *6*, 493–515.
- (10) Wang, X.-Y.; Yao, X.; Müllen, K. Polycyclic Aromatic Hydrocarbons in the Graphene Era. *Sci. China: Chem.* **2019**, *62*, 1099–1144.
- (11) Cai, J.; Ruffieux, P.; Jaafar, R.; Bieri, M.; Braun, T.; Blankenburg, S.; Muoth, M.; Seitsonen, A.; Saleh, M.; Feng, X.; Müllen, K.; Fasel, R. Atomically Precise Bottom-up Fabrication of Graphene Nanoribbons. *Nature* **2010**, *466*, 470–473.
- (12) Bronner, C.; Durr, R. A.; Rizzo, D. J.; Lee, Y.-L.; Marangoni, T.; Kalayjian, A. M.; Rodriguez, H.; Zhao, W.; Louie, S. G.; Fischer, F. R.; Crommie, M. F. Hierarchical On-Surface Synthesis of Graphene Nanoribbon Heterojunction. *ACS Nano* **2018**, *12*, 2193–2200.
- (13) Beyer, D.; Wang, S.; Pignedoli, C. A.; Melidoni, J.; Yuan, B.; Li, C.; Wilhelm, J.; Ruffieux, P.; Berger, R.; Müllen, K.; Fasel, R.; Feng, X. Graphene Nanoribbons Derived from Zigzag Edge-Encased Poly(para-2,9-dibenzo[bc,k]coronene-ylene) Polymer Chains. *J. Am. Chem. Soc.* **2019**, *141*, 2843–2846.
- (14) Radocea, A.; Sun, T.; Vo, T.; Sinitskii, A.; Aluru, N.; Lyding, J. Solution-Synthesized Chevron Graphene Nanoribbons Exfoliated onto H:Si(100). *Nano Lett.* **2017**, *17*, 170–178.
- (15) Gao, J.; Uribe-Romo, F. J.; Saathoff, J. D.; Arslan, H.; Crick, C. R.; Hein, S. J.; Itin, B.; Clancy, P.; Dichtel, W. R.; Loo, Y.-L. Ambipolar Transport in Solution-Synthesized Graphene Nanoribbons. *ACS Nano* **2016**, *10*, 4847–4856.
- (16) Li, G.; Yoon, K.-Y.; Zhong, X.; Zhu, X.; Dong, G. Efficient Bottom-Up Preparation of Graphene Nanoribbons by Mild Suzuki–Miyaura Polymerization of Simple Triaryl Monomers. *Chem. - Eur. J.* **2016**, *22*, 9116–9120.
- (17) Yang, W.; Lucotti, A.; Tommasini, M.; Chalifoux, W. A. Bottom-Up Synthesis of Soluble and Narrow Graphene Nanoribbons Using Alkyne Benzannulations. *J. Am. Chem. Soc.* **2016**, *138*, 9137–9144.
- (18) Hayashi, H.; Yamaguchi, J.; Jippo, H.; Hayashi, R.; Aratani, N.; Ohfuchi, M.; Sato, S.; Yamada, H. Experimental and Theoretical Investigations of Surface-Assisted Graphene Nanoribbon Synthesis Featuring Carbon-Fluorine Bond Cleavage. *ACS Nano* **2017**, *11*, 6204–6210.
- (19) Di Giovannantonio, M.; Deniz, O.; Urgel, J. I.; Widmer, R.; Diemel, T.; Stolz, S.; Sánchez-Sánchez, C.; Muntwiler, M.; Dumlaff, T.; Berger, R.; Narita, A.; Feng, X.; Müllen, K.; Ruffieux, P.; Fasel, R. On-Surface Growth Dynamics of Graphene Nanoribbons: The Role of Halogen Functionalization. *ACS Nano* **2018**, *12*, 74–81.
- (20) Kolmer, M.; Steiner, A. K.; Izydorczyk, I.; Ko, W.; Englund, M.; Szymonki, M.; Amsharov, K. Rational Synthesis of Atomically Precise Graphene Nanoribbons Directly on Metal Oxide Surfaces. *Science* **2020**, *369*, 571–575.
- (21) Ruffieux, P.; Wang, S.; Yang, B.; Sánchez-Sánchez, C.; Liu, J.; Diemel, T.; Talirz, L.; Shinde, P.; Pignedoli, C. A.; Passerone, D.; Dumlaff, T.; Feng, X.; Müllen, K.; Fasel, R. On-Surface Synthesis of Graphene Nanoribbons with Zigzag Edge Topology. *Nature* **2016**, *531*, 489–492.
- (22) Lee, Y.-L.; Zhao, F.; Cao, T.; Ihm, J.; Louie, S. G. Topological Phases in Cove-Edged and Chevron Graphene Nanoribbons: Geometric Structures, z_2 Invariants, and Junction States. *Nano Lett.* **2018**, *18*, 7247–7253.
- (23) Teeter, J. D.; Costa, P. S.; Pour, M. M.; Miller, D. P.; Zurek, E.; Enders, A.; Sinitskii, A. Epitaxial Growth of Aligned Atomically Precise Chevron Graphene Nanoribbons on Cu(111). *Chem. Commun.* **2017**, *53*, 8463–8466.
- (24) Bronner, C.; Stremlau, S.; Gille, M.; Brauße, F.; Haase, A.; Hecht, S.; Tegeder, P. Aligning the Band Gap of Graphene Nanoribbons by Monomer Doping. *Angew. Chem., Int. Ed.* **2013**, *52*, 4422–4425.
- (25) Liu, J.; Li, B.-W.; Tan, Y.-Z.; Giannakopoulos, A.; Sánchez-Sánchez, C.; Belljonne, C.; Ruffieux, P.; Fasel, R.; Feng, X.; Müllen, K. Toward Cove-Edged Low Band Gap Graphene Nanoribbons. *J. Am. Chem. Soc.* **2015**, *137*, 6097–6103.
- (26) Verzhbitskiy, I. A.; De Corato, M.; Ruini, A.; Molinari, E.; Narita, A.; Hu, Y.; Schwab, M. G.; Bruna, M.; Yoon, D.; Milana, S.; Feng, X.; Müllen, K.; Ferrari, A. C.; Casiraghi, C.; Prezzi, D. Raman

- Fingerprints of Atomically Precise Graphene Nanoribbons. *Nano Lett.* **2016**, *16*, 3442–3447.
- (27) Wu, J.; Gu, Y.; Muñoz-Mármol, R.; Wu, S.; Han, Y.; Ni, Y.; Díaz-García, M.; Casado, J. Cove-Edged Nanographenes with Localized Double Bonds. *Angew. Chem., Int. Ed.* **2020**, *59*, 8113–8117.
- (28) Yano, Y.; Mitoma, N.; Matsushima, K.; Wang, F.; Matsui, K.; Takakura, A.; Miyauchi, Y.; Ito, H.; Itami, K. Living Annulative π -Extension Polymerization for Graphene Nanoribbon Synthesis. *Nature* **2019**, *571*, 387–392.
- (29) Moreno, C.; Vilas-Varela, M.; Kretz, B.; Garcia-Lekue, A.; Costache, M. V.; Paradinas, M.; Panighel, M.; Ceballos, G.; Valenzuela, S. O.; Peña, D.; Mugarza, A. Bottom-up Synthesis of Multifunctional Nanoporous Graphene. *Science* **2018**, *360*, 199–203.
- (30) Pawlak, R.; Liu, X.; Ninova, S.; D'Astolfo, P.; Drechsel, C.; Sangtarash, S.; Häner, R.; Decurtins, S.; Sadeghi, H.; Lambert, C. J.; Aschauer, U.; Liu, S.-X.; Meyer, E. Bottom-up Synthesis of Nitrogen-Doped Porous Graphene Nanoribbons. *J. Am. Chem. Soc.* **2020**, *142*, 12568–12573.
- (31) Jacobse, P. H.; McCurdy, R. D.; Jiang, J.; Rizzo, D. J.; Veber, G.; Butler, P.; Zuzak, R.; Louie, S. G.; Fischer, F. R.; Crommie, M. F. Bottom-up Assembly of Nanoporous Graphene with Emergent Electronic States. *J. Am. Chem. Soc.* **2020**, *142*, 13507–13514.
- (32) Wang, X.-Y.; Yao, X.; Narita, A.; Müllen, K. Heteroatom-Doped Nanographenes with Structural Precision. *Acc. Chem. Res.* **2019**, *52*, 2491–2505.
- (33) Kawai, S.; Saito, S.; Osumi, S.; Yamaguchi, S.; Foster, A. S.; Spijker, P.; Meyer, E. Atomically Controlled Substitutional Boron-Doping of Graphene Nanoribbons. *Nat. Commun.* **2015**, *6*, 8098.
- (34) Wang, X.-Y.; Urgel, J. I.; Barin, G. B.; Eimre, K.; Di Giovannantonio, M.; Milani, A.; Tommasini, M.; Pignedoli, C. A.; Ruffieux, P.; Feng, X.; Fasel, R.; Müllen, K.; Narita, A. Bottom-Up Synthesis of Heteroatom-Doped Chiral Graphene Nanoribbons. *J. Am. Chem. Soc.* **2018**, *140*, 9104–9107.
- (35) Cao, Y.; Qi, J.; Zhang, Y.-F.; Huang, L.; Zheng, Q.; Lin, X.; Cheng, Z.; Zhang, Y.-Y.; Feng, X.; Du, S.; Pantelides, S. T.; Gao, H.-J. Tuning the Morphology of Chevron-Type Graphene Nanoribbons by Choice of Annealing Temperature. *Nano Res.* **2018**, *11*, 6190–6196.
- (36) Nguyen, G. D.; Toma, F. M.; Cao, T.; Pedramrazi, Z.; Chen, C.; Rizzo, D. J.; Joshi, T.; Bronner, C.; Chen, Y.-C.; Favaro, M.; Louie, S. G.; Fischer, F. R.; Crommie, M. F. Bottom-Up Synthesis of N = 13 Sulfur-Doped Graphene Nanoribbons. *J. Phys. Chem. C* **2016**, *120*, 2684–2687.
- (37) Kim, K. T.; Lee, J. W.; Jo, W. H. Charge-Transport Tuning of Solution-Processable Graphene Nanoribbons by Substitutional Nitrogen Doping. *Macromol. Chem. Phys.* **2013**, *214*, 2768–2773.
- (38) Vo, T. H.; Perera, U. G. E.; Shekhirev, M.; Pour, M. M.; Kunkel, D. A.; Lu, H.; Gruberman, A.; Sutter, E.; Cotlet, M.; Nykypanchuk, D.; Zahl, P.; Enders, A.; Sinitskii, A.; Sutter, P. Nitrogen-Doping Induced Self-Assembly of Graphene Nanoribbon-Based Two-Dimensional and Three-Dimensional Metamaterials. *Nano Lett.* **2015**, *15*, 5770–5777.
- (39) Kawai, S.; Nakatsuka, S.; Hatakeyama, T.; Pawlak, R.; Meier, T.; Tracey, J.; Meyer, E.; Foster, A. S. Multiple Heteroatom Substitution to Graphene Nanoribbon. *Science Advances* **2018**, *4*, eaar7181.
- (40) Cloke, R. R.; Marangoni, T.; Nguyen, G. D.; Joshi, T.; Rizzo, D. J.; Bronner, C.; Cao, T.; Louie, S. G.; Crommie, M. F.; Fischer, F. R. Site-Specific Substitutional Boron Doping of Semiconducting Armchair Graphene Nanoribbons. *J. Am. Chem. Soc.* **2015**, *137*, 8872–8875.
- (41) Piskun, L.; Blackwell, R.; Jornet-Somoza, J.; Zhao, F.; Rubio, A.; Louie, S.; Fischer, F. Covalent C-N Bond Formation through a Surface Catalyzed Thermal Cyclodehydrogenation. *J. Am. Chem. Soc.* **2020**, *142*, 3696–3700.
- (42) Cai, J.; Pignedoli, C. A.; Talirz, L.; Ruffieux, P.; Sode, H.; Liang, L.; Meunier, V.; Berger, R.; Li, R.; Feng, X.; Müllen, K.; Fasel, R. Graphene Nanoribbon Heterojunctions. *Nat. Nanotechnol.* **2014**, *9*, 896–900.
- (43) Zhang, Y.; Zhang, Y.; Li, G.; Lu, J.; Lin, X.; Du, S.; Berger, R.; Feng, X.; Müllen, K.; Gao, H.-J. Direct Visualization of Atomically Precise Nitrogen-Doped Graphene Nanoribbons. *Appl. Phys. Lett.* **2014**, *105*, No. 023101.
- (44) Durr, R. A.; Haberer, D.; Lee, Y.-L.; Blackwell, R.; Kalayjian, A. M.; Marangoni, T.; Ihm, J.; Louie, S. G.; Fischer, F. R. Orbital Matched Edge-Doping in Graphene Nanoribbons. *J. Am. Chem. Soc.* **2018**, *140*, 807–813.
- (45) Vo, T. H.; Shekhirev, M.; Kunkel, D. A.; Orange, F.; Guinel, M. F.-F.; Enders, A.; Sinitskii, A. Bottom-Up Solution Synthesis of Narrow Nitrogen-Doped Graphene Nanoribbons. *Chem. Commun.* **2014**, *50*, 4172–4174.
- (46) Wegner, G. Topochemical Polymerization of Monomers with Conjugated Triple Bonds. *Makromol. Chem.* **1972**, *154*, 35–48.
- (47) Enkelmann, V. (1984) Structural aspects of the topochemical polymerization of diacetylenes. In: Cantow, H. J. (eds) *Polydiacetylenes. Advances in Polymer Science*, vol 63. Springer: Berlin, Heidelberg, DOI: 10.1007/BFb0017652.
- (48) Lauher, J.; Fowler, E.; Goroff, N. Single-Crystal-to-Single-Crystal Topochemical Polymerizations by Design. *Acc. Chem. Res.* **2008**, *41*, 1215–1229.
- (49) Jordan, R. S.; Wang, Y.; McCurdy, R. D.; Yeung, M. T.; Marsh, K. L.; Kahn, S. L.; Kaner, R. B.; Rubin, Y. Synthesis of Graphene Nanoribbons via the Topochemical Polymerization and Subsequent Aromatization of a Diacetylene Precursor. *Chem.* **2016**, *1*, 78–90.
- (50) Jordan, R. S.; Li, Y. L.; Lin, C.-W.; McCurdy, R. D.; Lin, J. B.; Brosmer, J. L.; Marsh, K. L.; Khan, S. I.; Houk, K. N.; Kaner, R. B.; Rubin, Y. Synthesis of N = 8 Armchair Graphene Nanoribbons from Four Distinct Polydiacetylenes. *J. Am. Chem. Soc.* **2017**, *139*, 15878–15890.
- (51) See Supporting Information.
- (52) See reference 49 for a similar observation on an intermediate stage to [8]_hGNR.
- (53) Ryu, S.; Maultzsch, J.; Han, M. Y.; Kim, P.; Brus, L. E. Raman Spectroscopy of Lithographically Patterned Graphene Nanoribbons. *ACS Nano* **2011**, *5*, 4123–4130.
- (54) Ortiz-Medina, J.; García-Betancourt, M. L.; Jia, X.; Martínez-Gordillo, R.; Pelagio-Flores, M. A.; Swandon, D.; Elias, A. L.; Gutiérrez, H. R.; García-Espino, E.; Meunier, V.; Owens, J.; Sumpter, B. G.; Cruz-Silva, E.; Rodríguez-Macías, F. J.; López-Urías, F.; Muñoz-Sandoval, E.; Dresselhaus, M. S.; Terrones, H.; Terrones, M. Nitrogen-Doped Graphitic Nanoribbons: Synthesis, Characterization, and Transport. *Adv. Funct. Mater.* **2013**, *23*, 3755–3762.
- (55) Daigle, M.; Maio, D.; Lucotti, A.; Tommasini, M.; Morin, J.-F. Helically Coiled Graphene Nanoribbons. *Angew. Chem.* **2017**, *129*, 6309–6313.
- (56) Zhang, C.; Fu, L.; Liu, N.; Lui, M.; Wang, Y.; Liu, Z. Synthesis of Nitrogen-Doped Graphene Using Embedded Carbon and Nitrogen Sources. *Adv. Mater.* **2011**, *23*, 1020–1024.
- (57) Lu, Y.-F.; Lo, S.-T.; Lin, J.-C.; Zhang, W.; Lu, J.-Y.; Liu, F.-H.; Tseng, C.-M.; Lee, Y.-H.; Liang, C.-T.; Li, L.-L. Nitrogen-Doped Graphene Sheets Grown by Chemical Vapor Deposition: Synthesis and Influence of Nitrogen Impurities on Carrier Transport. *ACS Nano* **2013**, *7*, 6522–6532.
- (58) Liu, M.; Song, Y.; He, S.; Tjui, W. W.; Pan, J.; Xia, Y.-Y.; Liu, T. Nitrogen-Doped Graphene Nanoribbons as Efficient Metal-Free Electrocatalysts for Oxygen Reduction. *ACS Appl. Mater. Interfaces* **2014**, *6*, 4212–4222.
- (59) Wang, G.; Maiyalagan, T.; Wang, X. Review on Recent Progress in Nitrogen-Doped Graphene: Synthesis, Characterization, and Its Potential Applications. *ACS Catal.* **2012**, *2*, 781–794.
- (60) Geng, J.; Si, L.; Guo, H.; Lin, C.; Xi, Y.; Li, Y.; Yan, X.; Wang, B.; Chen, L. 3D Nitrogen-Doped Graphene Gels as Robust and Sustainable Adsorbents for Dyes. *New J. Chem.* **2017**, *41*, 15447–15457.
- (61) Kondo, T.; Casolo, S.; Suzuki, T.; Shikano, T.; Sakurai, M.; Harada, Y.; Saito, M.; Oshima, M.; Trioni, M. I.; Tantardini, G. F.; Nakamura, J. Atomic-Scale Characterization of Nitrogen-Doped Graphite: Effects of Dopant Nitrogen on the Local Electronic

Structure of the Surrounding Carbon Atoms. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2012**, *86*, No. 035436.

(62) Jansen, R. J. J.; van Bekkum, H. XPS of Nitrogen-Containing Functional Groups on Activated Carbon. *Carbon* **1995**, *8*, 1021–1027.

(63) Tian, Z.; Dai, S.; Jiang, D. Stability and Core-Level Signature of Nitrogen Dopants in Carbonaceous Materials. *Chem. Mater.* **2015**, *27*, 5775–5781.

(64) Kapteijn, F.; Moulijn, J. A.; Matzner, S.; Boehm, H.-P. The Development of Nitrogen Functionality in Model Chars during Gasification in CO₂ and O₂. *Carbon* **1999**, *37*, 1143–1150.

(65) Matanovic, L.; Artyushkova, K.; Strand, M. B.; Dzara, M. J.; Pylypenko, S.; Atanassov, P. Core Level Shifts of Hydrogenated Pyridinic and Pyrrolic Nitrogen in the Nitrogen-Containing Graphene-Based Electrocatalysts: In-Place vs Edge Defects. *J. Phys. Chem. C* **2016**, *120*, 29225–29232.

(66) Men, S.; Mitchell, D. S.; Lovelock, K. R. J.; License, P. X-ray Photoelectron Spectroscopy of Pyridinium-Based Ionic Liquids: Comparison to Imidazolium- and Pyrrolidinium-Based Analogues. *ChemPhysChem* **2015**, *16*, 2211–2218.

(67) Prall, M.; Krüger, A.; Schreiner, P. R.; Hopf, H. The Cyclization of Parent and Cyclic Hexa-1,3-dien-5-yne—A Combined Theoretical and Experimental Study. *Chem. - Eur. J.* **2001**, *7*, 4386–4394.

(68) Krukau, A. V.; Vydrov, O. A.; Izmaylov, A. F.; Scuseria, G. E. Influence of the exchange screening parameter on the performance of screened hybrid functionals. *J. Chem. Phys.* **2006**, *125*, 224106.

(69) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(70) Mori-Sanchez, P.; Cohen, A. J.; Yang, W. Localization and Delocalization Errors in Density Functional Theory and Implications for Band-Gap Prediction. *Phys. Rev. Lett.* **2008**, *100*, 146401.

Fjord-edge Graphene Nanoribbons with Site-Specific Nitrogen Substitution

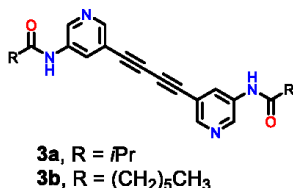
Yolanda L. Li, Chih-Te Zee, Janice B. Lin, Victoria M. Basile, Mit Muni, Maria D. Flores, Julen Munárriz, Richard B. Kaner, Anastassia N. Alexandrova, K. N. Houk, Sarah H. Tolbert, and Yves Rubin*

Table of Contents

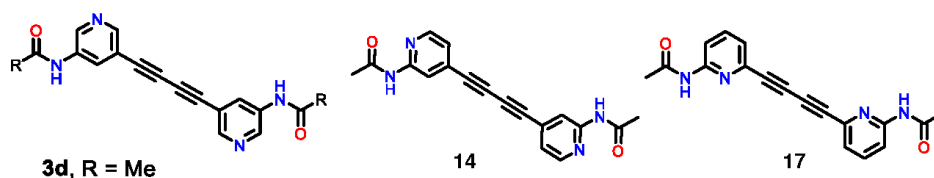
Section 1:	Experimental Details	
	1.1. Experimental details	S2
	1.2. NMR spectra.....	S10
	1.3. FT-IR spectra of GNRs 1a and 1b	S35
	1.4. UV-Vis spectra of GNRs 1a and 1b	S37
Section 2:	Crystallography Details	S38
Section 3:	High-Resolution Transmission Microscopy (HR-TEM).....	S43
Section 4:	Computational Details	S45
	4.1. Molecular Mechanics Calculations of the Packing Modes.....	S45
	4.2. Packing Structure Coordinates for fjord-edge N ₂ [8]GNR Acetamide Derivative 1c	S48
	4.3. Packing Structure Coordinates for fjord-edge N ₂ [8]GNR Parent System 1d	S76
	4.4. Computational Methods for the Preferred Hopf Cyclization Pathway	S111
	4.5. XYZ Geometries	S112
	Substituted Model System	S112
	Results for Unsubstituted N-PDA Model System	S124
	4.6. Calculation of the Electronic Density of States (DOS) for 1d and 2d	S134
	Geometrical coordinates for fjord-edge N ₂ [8]GNR 1d and N ₂ [8] _A GNR 2d	S138
	4.7. Calculation of ¹³ C NMR Chemical Shifts for Model Compounds	S141
	4.8. Conformational Preference Calculations on Model Compound 1e	S145
Section 5:	References.....	S147

Section 1: Experimental Details

Diynes undergoing successful topochemical polymerization:



Diynes that did not polymerize topochemically:



General procedures: Chemical reagents were obtained from commercial sources and used without further purification. Unless otherwise noted, all reactions were performed under argon at 25 °C. Silicycle silicagel 60 (Siliflash P60, 240-400 mesh) was used for all column chromatographic purifications.

Solution NMR spectra were obtained on a Bruker AV500 instrument. CP/MAS ¹³C NMR spectra were obtained on a Bruker AV600 using a 3.2 mm zirconia rotor spinning at 18 kHz. A 514 nm excitation laser was used on pellets of material to obtain Raman spectra. The x-ray photoelectron spectroscopy (XPS) was performed using a Kratos Axis Ultra DLD spectrometer with a monochromatic Al K α radiation source. A pass energy of 20 eV was used with a step size of 0.1 eV, a dwell time of 300 ms, and a charge neutralizer filament. The obtained spectra were calibrated to the advantageous carbon 1s peak at 284.8 eV and analysis of the spectra was performed using CasaXPS.

3-Amino-5-((trimethylsilyl)ethynyl)pyridine (5).

3-Amino-5-bromopyridine (5.00 g, 28.9 mmol, 1.0 eq) was added to a round-bottomed flask with a stirring bar under argon. Toluene (75 mL), bis(triphenylphosphine) palladium dichloride (507 mg, 0.722 mmol, 0.025 eq) and copper(I) iodide (79 mg, 0.72 mmol, 0.025 eq) were added and the mixture was sparged with argon for 15 minutes. Trimethylsilylacetylene (5.37 g, 57.7 mmol, 2.0 eq) and diisopropylamine (5.84 g, 57.7 mmol, 2.0 eq) were added before refluxing overnight. After reaction completion as ascertained by ¹H NMR of the crude, the mixture was cooled to 25 °C, filtered over Celite and concentrated on the Rotavap. Purification by flash column

chromatography with SiO₂ yielded 5.02 g (91%) of **5** as a pale tan solid. ¹H NMR (400 MHz, CDCl₃): δ (ppm) 0.25 (s, 9H), 3.68 (br s, 2H), 7.03 (dd, *J* = 2.7, 1.9 Hz, 1H), 8.00 (d, *J* = 2.7 Hz, 1H), 8.10 (d, *J* = 1.9 Hz, 1H); ¹³C NMR (125 MHz, CDCl₃): δ (ppm) 0.0, 97.5, 102.0, 120.28, 124.0, 137.0, 141.8, 143.1; HRMS (DART) Calculated for C₁₀H₁₄N₂Si [M+H]: 191.09902; found 191.10066.

N-(5-Ethynyl-3-pyridyl)isobutyramide (6a).

3-Amino-5-((trimethylsilyl)ethynyl)pyridine **5** (5.02 g, 26.4 mmol, ~1.0 eq) was added to a round-bottomed flask with a stirring bar under argon and dissolved in anhydrous CH₂Cl₂ (75 mL). Triethylamine (5.34 g, 52.8 mmol, 2.0 eq) was added and the flask was cooled to 0 °C. Isobutryl chloride (4.22 g, 39.6 mmol, 1.5 eq) was dissolved in CH₂Cl₂ (15 mL) and added dropwise. The mixture was allowed to warm to 25 °C while stirring. Upon completion by TLC, the reaction was washed with an aqueous solution of NaHCO₃ and extracted twice with CH₂Cl₂. Organic layers were combined and washed with brine before drying over MgSO₄ and concentrating *in vacuo*. The crude product was then dissolved in MeOH and THF (each 130 mL), anhydrous K₂CO₃ added (~0.5 g) and stirred at 25 °C until completion by TLC. The solvents were removed *in vacuo* before partitioning between water and CH₂Cl₂. The aqueous layer was extracted with CH₂Cl₂ twice, organic layers combined, washed with brine, dried over MgSO₄ and concentrated. A total of 4.73 g (95% over 2 steps) of amide **6a** was collected as pale yellow crystals. ¹H NMR (400 MHz, CDCl₃): δ (ppm) 1.26 (d, *J* = 6.9 Hz, 6H), 2.58 (sept, *J* = 6.9 Hz, 1H), 3.19 (s, 1H), 7.24 (br s, 1H, NH), 7.56, (br s, 1H), 8.30 (dd, *J* = 2.6, 1.9 Hz, 1H), 8.43 (d, *J* = 1.9 Hz, 1H), 8.52 (d, *J* = 2.6, 1H); ¹³C NMR (125 MHz, CDCl₃): δ (ppm) 19.5, 36.6, 79.9, 80.9, 129.9, 133.6, 134.4, 140.3, 147.9, 175.7; HRMS (DART) Calculated for C₁₁H₁₂N₂O [M⁺]: 188.09496; found 188.10645.

N,N'-(Buta-1,3-diyne-1,4-diylbis(pyridine-5,3-diyl))bis(2-methylpropanamide) (3a).

N-(5-ethynyl-3-pyridyl)isobutyramide **6a** (3.33 g, 17.7 mmol, 1.0 eq) was dissolved in CH₂Cl₂ (53 mL) before adding N,N,N',N'-tetramethylethylene-1,2-diamine (TMEDA) (205 mg, 1.77 mmol, 0.1 eq) and copper(I) iodide (337 mg, 1.77 mmol, 0.1 eq). Air was bubbled through the mixture for 15 minutes before allowing to stir overnight open to air. The product was filtered and washed with water to give 2.07 g (63%) of diyne **3a**. The material was then recrystallized from MeOH to yield clear colorless needles which rapidly turn deep blue upon exposure to light. ¹H NMR (400 MHz, DMSO): δ (ppm) 1.08 (d, *J* = 6.9 Hz, 12H), 2.58 (sept, *J* = 6.9 Hz, 2H), 8.26 (dd, *J* = 1.8, 1.4 Hz, 2H), 8.45 (d, *J* = 1.4 Hz, 2H), 8.74 (d, *J* = 1.8 Hz, 2H), 10.24 (s, 2H); ¹³C NMR (125 MHz, DMSO): δ (ppm) 19.8, 35.4, 76.1, 80.0, 117.7, 128.9, 136.2, 141.8, 147.3, 176.7; HRMS (DART) Calculated for C₂₂H₂₂N₄O₂ [M⁺]: 374.1743; found 374.1596.

N-(5-Ethynyl-3-pyridyl)heptanamide (6b).

3-Amino-5-((trimethylsilyl)ethynyl)pyridine **5** (4.90 g, 25.8 mmol, ~1.0 eq) was added to a round-bottomed flask with a stirring bar under argon and dissolved in anhydrous CH₂Cl₂ (72 mL). Triethylamine (5.21 g, 52.0 mmol, 2.0 eq) was added and the flask cooled to 0 °C. Heptanoyl chloride (5.74 g, 38.6 mmol, 1.5 eq) was dissolved in CH₂Cl₂ (15 mL) and added dropwise. The mixture was allowed to warm to 25 °C while stirring. Upon completion by TLC, the reaction was washed with an aqueous solution of NaHCO₃ and extracted twice with CH₂Cl₂. Organic layers were combined and washed with brine before drying over MgSO₄ and concentrating *in vacuo*. The crude product was then dissolved in MeOH (100 mL) and THF (100 mL), anhydrous K₂CO₃ added (~0.5 g) and the solution stirred at 25 °C until reaction completion (desilylation) as monitored by TLC. The solvents were removed *in vacuo* before partitioning between water and CH₂Cl₂. The aqueous layer was extracted with CH₂Cl₂ twice, organic layers combined, washed with brine, dried over MgSO₄ and concentrated. A total of 4.65 g (78% over 2 steps) of amide **6b** was collected as pale yellow crystals. ¹H NMR (400 MHz, CDCl₃): δ (ppm) 0.88 (t, *J* = 7.0 Hz, 3H), 1.31 (m, 4H), 1.36 (t, *J* = 7.0 Hz, 2 H), 1.72 (quint, *J* = 7.5 Hz, 2H), 2.38 (t, *J* = 7.5 Hz, 2H), 3.19 (s, 1H), 7.37 (br s, 1H), 8.29 (dd, *J* = 2.6, 1.6 Hz, 1H), 8.42 (d, *J* = 1.6 Hz, 1H), 8.50 (d, *J* = 2.6 Hz, 1H); ¹³C NMR (125 MHz, CDCl₃): δ (ppm) 14.0, 22.5, 25.4, 28.9, 31.5, 80.0, 80.9, 119.5, 129.9, 134.5, 140.3, 147.8, 172.1; HRMS (DART) Calculated for C₁₄H₁₈N₂O [M⁺]: 230.1419; found 230.1312.

N,N'-(Buta-1,3-diyne-1,4-diylbis(pyridine-5,3-diyl))diheptanamide (3b).

N-(5-Ethynyl-3-pyridyl)heptanamide **6b** (3.92 g, 17.0 mmol, 1.0 eq) was dissolved in CH₂Cl₂ (50 mL) before adding N,N,N',N'-tetramethylethylene-1,2-diamine (TMEDA) (197 mg, 1.70 mmol, 0.1 eq) and copper(I) iodide (324 mg, 1.70 mmol, 0.1 eq). Air was bubbled through the mixture for 15 minutes before allowing to stir overnight open to air. The product was filtered and washed with water to give 3.60 g (92%) of diyne **3b**. The material was then recrystallized from isopropanol to yield clear needles which rapidly turn blue upon exposure to light. ¹H NMR (500 MHz, DMSO): δ (ppm) 0.86 (t, *J* = 7.0, 6H), 1.28 (m, 6H), 1.59 (quint, *J* = 7.0 Hz, 4H), 2.34 (t, *J* = 7.5 Hz, 4H), 8.28 (dd, *J* = 2.3, 1.4 Hz, 2H), 8.48 (d, *J* = 1.4 Hz, 2H), 8.73 (d, *J* = 2.3, 2H), 10.31 (s, 2H); ¹³C NMR (125 MHz, DMSO): δ (ppm) 19.1, 27.2, 30.0, 33.5, 36.2, 41.5, 80.8, 84.7, 122.4, 133.6, 140.8, 146.4, 152.1, 177.6; HRMS (DART) Calculated for C₂₈H₃₄N₄O₂ [M⁺]: 458.2682; found 458.3872.

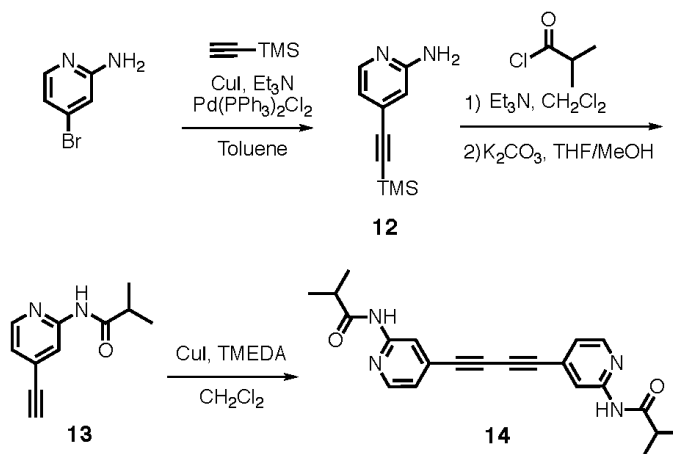
N-(5-Ethynyl-3-pyridyl)acetamide (6d).

3-Amino-5-((trimethylsilyl)ethynyl)pyridine **5** (11.00 g, 57.8 mmol, 1.0 eq) was added to a round-bottomed flask with a stirring bar under argon and dissolved in anhydrous THF (115 mL). Acetic anhydride (8.843 g, 86.7 mmol, 1.5 eq) was added slowly at 0 °C. The mixture was allowed to warm to 25 °C while stirring overnight. The reaction was washed with an aqueous solution of

NaHCO₃ and extracted twice with CH₂Cl₂. Organic layers were combined and washed with brine before drying over MgSO₄ and concentrating *in vacuo*. The residue was dissolved in methanol and THF (85 mL each) before adding anhydrous potassium carbonate (17.53 g, 126.8 mmol, 3.0 eq) and stirring until completion. The aqueous layer was extracted with CH₂Cl₂ twice, organic layers combined, washed with brine, dried over MgSO₄ and concentrated to yield 4.559 g (60%) of **6d**. ¹H NMR (400 MHz, CDCl₃): δ (ppm) 8.49 (br s, 1H), 8.45 (br s, 1H), 8.26 (s, 1H), 7.15 (br s, 1H), 3.21 (s, 1H), 2.22 (s, 3H).

N,N'-(Buta-1,3-diyne-1,4-diylbis(pyridine-5,3-diyl))diacetamide (3d).

N-(5-ethynyl-3-pyridyl)acetamide **6c** (4.559 g, 34.7 mmol, 1.0 eq) was dissolved in CH₂Cl₂ (105 mL) before adding N,N,N',N'-tetramethylethylenediamine (TMEDA) (403 mg, 3.47 mmol, 0.1 eq) and copper(I) iodide (661 mg, 3.47 mmol, 0.1 eq). Air was bubbled through the mixture for 15 minutes before allowing to stir overnight open to air. The product was filtered and washed with water to give 5.156 g (94%) of diyne **3d**. ¹H NMR (400 MHz, d₆-DMSO): δ (ppm) 10.35 (br s, 2H), 8.69 (s, 2H), 8.46 (s, 2H), 8.24 (s, 2H), 2.07 (s, 6H); ¹³C NMR (125 MHz, d₆-DMSO): δ (ppm) 169.84, 147.4, 141.6, 136.0, 128.9, 117.6, 79.9, 76.1, 24.4.



Scheme S1. Synthesis of 2,4-disubstituted dipyridyl diyne **14**.

2-Amino-4-((trimethylsilyl)ethynyl)pyridine (12).

2-Amino-4-bromopyridine (20.0 g, 115 mmol, 1.0 eq) was added to a round-bottomed flask fitted with a stirring bar. Toluene (250 mL), bis(triphenylphosphine) palladium dichloride (4.06 g, 5.78

mmol, 0.05 eq) and copper(I) iodide (2.20 g, 11.5 mmol, 0.1 eq) were added, then the mixture was sparged with argon for 15 minutes. Trimethylsilylacetylene (22.7 g, 231 mmol, 2.0 eq) and triethylamine (23.4 g, 231 mmol, 2.0 eq) were added before refluxing overnight. After reaction completion as ascertained by ^1H NMR of the crude, the mixture was cooled to 25 °C, then filtered over Celite. The filtrate was then washed with water and acidified with HCl to pH = 1. The aqueous phase was then washed with CH_2Cl_2 , followed by neutralization to pH = 7 using aqueous NaHCO_3 . The aqueous phase was extracted with CH_2Cl_2 three times before all the organic layers were combined, dried over MgSO_4 and concentrated. The product **12**, 18.8 g (86%), was obtained as a tan colored solid. ^1H NMR (400 MHz, CDCl_3): δ (ppm) 8.00 (d, $J = 2.6$ Hz, 1H), 6.67 (dd, $J = 2.6$, 0.6 Hz, 1H), 6.54 (d, $J = 0.6$ Hz, 1H), 4.42 (br s, 2H), 0.24 (s, 9H); ^{13}C NMR (125 MHz, CDCl_3): δ (ppm) 158.2, 148.2, 132.5, 116.4, 110.8, 102.6, 98.2, 0.3.

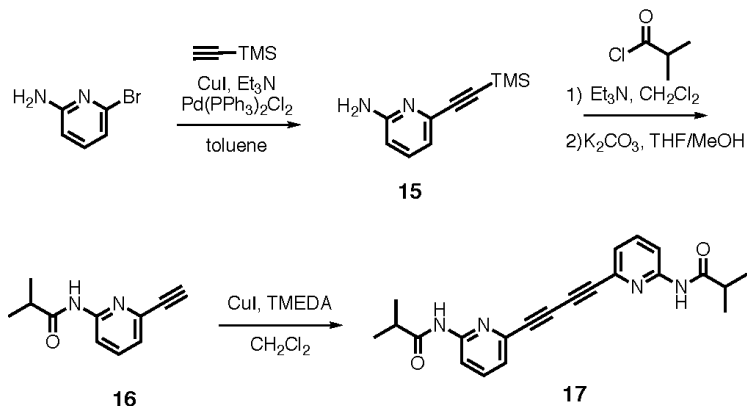
N-(4-Ethynyl-2-pyridyl)isobutyramide (13).

2-Amino-4-((trimethylsilyl)ethynyl)pyridine **12** (500 mg, 2.63 mmol, 1.0 eq) was added to a round-bottomed flask fitted with a stirring bar under argon. After dissolution into anhydrous CH_2Cl_2 (8 mL), triethylamine (663 mg, 6.57 mmol, 2.5 eq) was added. The flask was cooled to 0 °C, then isobutyryl chloride (616 mg, 5.78 mmol, 2.2 eq), dissolved in CH_2Cl_2 (2.3 mL), was added dropwise. The mixture was allowed to warm to 25 °C while stirring. Upon completion by TLC, the reaction was washed with an aqueous solution of NaHCO_3 and extracted twice with CH_2Cl_2 . Organic layers were combined and washed with brine before drying over MgSO_4 and concentrating *in vacuo*. The crude product was then dissolved in methanol and THF (10 mL each), anhydrous K_2CO_3 (868 mg, 6.29 mmol, 3.0 eq) was added and the mixture was stirred at 25 °C until the removal of the trimethylsilyl group was completed, as judged by TLC. The solvents were removed *in vacuo* before partitioning between water and CH_2Cl_2 . The aqueous layer was extracted with CH_2Cl_2 twice, organic layers combined, washed with brine, dried over MgSO_4 and concentrated to yield 286 mg (59%) of compound **13**. ^1H NMR (400 MHz, CDCl_3): δ (ppm) 8.35 (t, $J = 1.1$ Hz, 1H), 8.21 (dd, $J = 2.6$, 0.4 Hz, 1H), 7.84 (br s, 1H), 7.07 (dd, $J = 2.6$, 0.8 Hz, 1H), 3.26 (s, 1H), 2.54 (septet, $J = 6.9$ Hz, 1H), 1.26 (d, $J = 6.9$ Hz, 6H). ^{13}C NMR (125 MHz, CDCl_3): δ (ppm) 175.5, 151.5, 147.8, 132.2, 122.2, 116.4, 81.7, 81.1, 36.8, 19.4.

N,N'-(buta-1,3-diyne-1,4-diylbis(pyridine-4,2-diyl))bis(2-methylpropanamide) (14).

N-(4-Ethynyl-2-pyridyl)isobutyramide **13** (290 mg, 1.54 mmol, 1.0 eq) was dissolved in CH_2Cl_2 (5 mL) before adding N,N,N',N'-tetramethylethylene-1,2-diamine (TMEDA) (17.8 mg, 0.154 mmol, 0.1 eq) and copper(I) iodide (29 mg, 0.154 mmol, 0.1 eq). Air was bubbled through the mixture for 15 minutes before allowing to stir overnight open to air. The product was filtered and washed with water to give 253 mg (88%) of diyne **14** as off-white fibrous crystals. ^1H NMR (400 MHz, d_6 -DMSO): δ (ppm) 10.64 (br s, 2H), 8.38 (dd, $J = 2.6$, 0.3 Hz, 2H), 8.23 (t, $J = 1.1$ Hz, 2H),

7.28 (dd, $J = 2.6$ Hz, 2H), 2.74 (septet, $J = 6.9$ Hz, 2H), 1.07 (d, $J = 6.9$ Hz, 12H); ^{13}C NMR (125 MHz, d_6 -DMSO): δ (ppm) 177.1, 153.0, 149.3, 129.9, 121.8, 115.9, 81.2, 76.2, 34.9, 19.8.



Scheme S2. Synthesis of 2,6-disubstituted dipyridyl diyne **16**.

2-Amino-6-((trimethylsilyl)ethynyl)pyridine (**15**).

2-Amino-6-bromopyridine (1.00 g, 5.78 mmol, 1.0 eq) was dissolved in toluene (15 mL) under argon along with bis(triphenylphosphine) palladium dichloride (203 mg, 0.29 mmol, 0.05 eq) and copper(I) iodide (110 mg, 0.589 mmol, 0.1 eq) before sparging with argon for 15 minutes. Trimethylsilylacetylene (1.14 g, 11.5 mmol, 2.0 eq) and triethylamine (1.17 g, 11.5 mmol, 2.0 eq) were added before refluxing overnight. After reaction completion as ascertained by ^1H NMR of the crude, the mixture was cooled to 25 °C, filtered over Celite. The filtrate was then washed with water and acidified with HCl to pH = 1. The aqueous phase was then washed with CH_2Cl_2 then neutralized using aqueous NaHCO_3 . The aqueous phase was then extracted with CH_2Cl_2 three times before organic layers were combined, dried over MgSO_4 and concentrated. A total of 767 mg (70%) of compound **15** was obtained. ^1H NMR (400 MHz, CDCl_3): δ (ppm) 7.36 (dd, $J = 4.4$, 3.6 Hz, 1H), 6.85 (dd, $J = 3.7$, 0.4 Hz, 1H), 6.64 (dd, $J = 4.1$, 0.4 Hz, 1H), 4.48 (br s, 2H), 0.25 (s, 9H); ^{13}C NMR (125 MHz, CDCl_3): δ (ppm) 158.1, 140.9, 137.7, 118.1, 108.8, 104.0, 93.5, 0.2.

N-(6-ethynyl-2-pyridyl)isobutyramide (**16**)

2-Amino-6-((trimethylsilyl)ethynyl)pyridine **15** (200 mg, 1.05 mmol, 1.0 eq) was dissolved in anhydrous CH_2Cl_2 (3 mL) under argon along with triethylamine (319 mg, 2.63 mmol, 3.0 eq) at 0 °C. A solution of isobutryl chloride (280 mg, 2.63 mmol, 2.5 eq) in CH_2Cl_2 (1 mL) was added

dropwise. The mixture was allowed to warm to 25 °C while stirring, then washed with an aqueous solution of NaHCO₃ and extracted twice with CH₂Cl₂. Organic layers were combined and washed with brine before drying over MgSO₄ and concentrating *in vacuo*. The residue was dissolved in methanol and THF (7 mL each) before adding anhydrous potassium carbonate (601 mg, 4.35 mmol, 3.0 eq). After stirring at room temperature for 3 hours, the mixture was concentrated and partitioned between water and CH₂Cl₂. The aqueous layer was further extracted with CH₂Cl₂ before the organic layers were combined and washed with water and brine. After drying over MgSO₄ and concentrating, 210 mg (99%) of product **16** was obtained as an off-white solid. ¹H NMR (400 MHz, CDCl₃): δ (ppm) 8.24 (dd, *J* = 4.3, 0.4 Hz, 1H), 7.95 (br s, 1H), 7.67 (t, *J* = 8.0 Hz, 1H), 7.21 (dd, *J* = 3.8, 0.4 Hz, 1H), 3.14 (s, 1H), 2.52 (septet, *J* = 6.9 Hz, 1H), 1.25 (d, *J* = 6.9 Hz, 6H); ¹³C NMR (125 MHz, CDCl₃): δ (ppm) 175.7, 151.5, 139.9, 138.7, 123.3, 114.2, 82.2, 77.1, 36.8, 19.4.

N,N'-(buta-1,3-diyne-1,4-diylbis(pyridine-6,2-diyl))bis(2-methylpropanamide) (17)

A solution of N-(6-ethynyl-2-pyridyl)isobutyramide **15** (210 mg, 1.12 mmol, 1.0 eq) in CH₂Cl₂ (4 mL), N,N,N',N'-tetramethylethylene-1,2-diamine (TMEDA) (22 mg, 0.11 mmol, 0.1 eq), and copper(I) iodide (13 mg, 0.11 mmol, 0.1 eq) was stirred at room temperature. Air was bubbled through the reaction for 15 minutes before being left open to air overnight. The product was filtered and washed with water to give 200 mg (96%) of diyne **17**. ¹H NMR (400 MHz, CDCl₃): δ (ppm) 8.29 (dd, *J* = 4.2, 0.4 Hz, 1H), 7.96 (br s, 1H), 7.70 (t, *J* = 8.1 Hz, 1H), 7.28 (d, *J* = 3.8 Hz, 1H), 2.54 (septet, *J* = 6.9 Hz, 1H), 1.25 (d, *J* = 6.9 Hz, 6H); ¹³C NMR (125 MHz, CDCl₃): δ (ppm) 175.8, 151.7, 139.3, 138.8, 124.3, 114.8, 85.8, 80.4, 36.8, 19.4.

General procedure for the topochemical polymerizations:

Finely crushed crystals were suspended in hexanes (100 mL/g) in a capped Erlenmeyer flask. This was placed in a photoreactor and stirred overnight under exposure to a medium pressure Hanovia mercury lamp. The dark brown solids were then filtered off and subsequently heating to boiling in the appropriate solvent to remove residual monomeric starting material after filtration. Accordingly, the remaining insoluble PDAs were filtered over filter paper and peeled off as a film. The recovered filtrate was evaporated *in vacuo* and the starting material was recrystallized to repeat the polymerization process.

PDA 4a.

Polymerization yield: 587 mg of PDA **4a** from 3.30 g of **3a** (18%)

Solvent: Boiling in ethanol and filtration followed by additional boiling in DMF, filtering and washing with diethyl ether.

PDA 4b.

Polymerization yield: 130 mg of PDA **4b** from 2.97 g of **3b** (4%)

Solvent: Boiling in chloroform and filtration followed by additional boiling in DMF, filtering and washing with diethyl ether.

General procedure for the graphitization of PDAs 4a,b:

All heating experiments were done using a 1-inch wide quartz tube and programmable tube furnace (MTI OTF-1200X0S0NT). PDAs **4a** and **4b** were placed in an alumina boat and heated under positive argon flow. The furnace was programmed to ramp from 25 °C to the desired temperature over 40 minutes, remaining at the desired temperature for 1 hour, then cooled down to 25 °C over 40 minutes. Once fully cooled, the samples were removed from the furnace and weighted.

1.2. NMR spectra

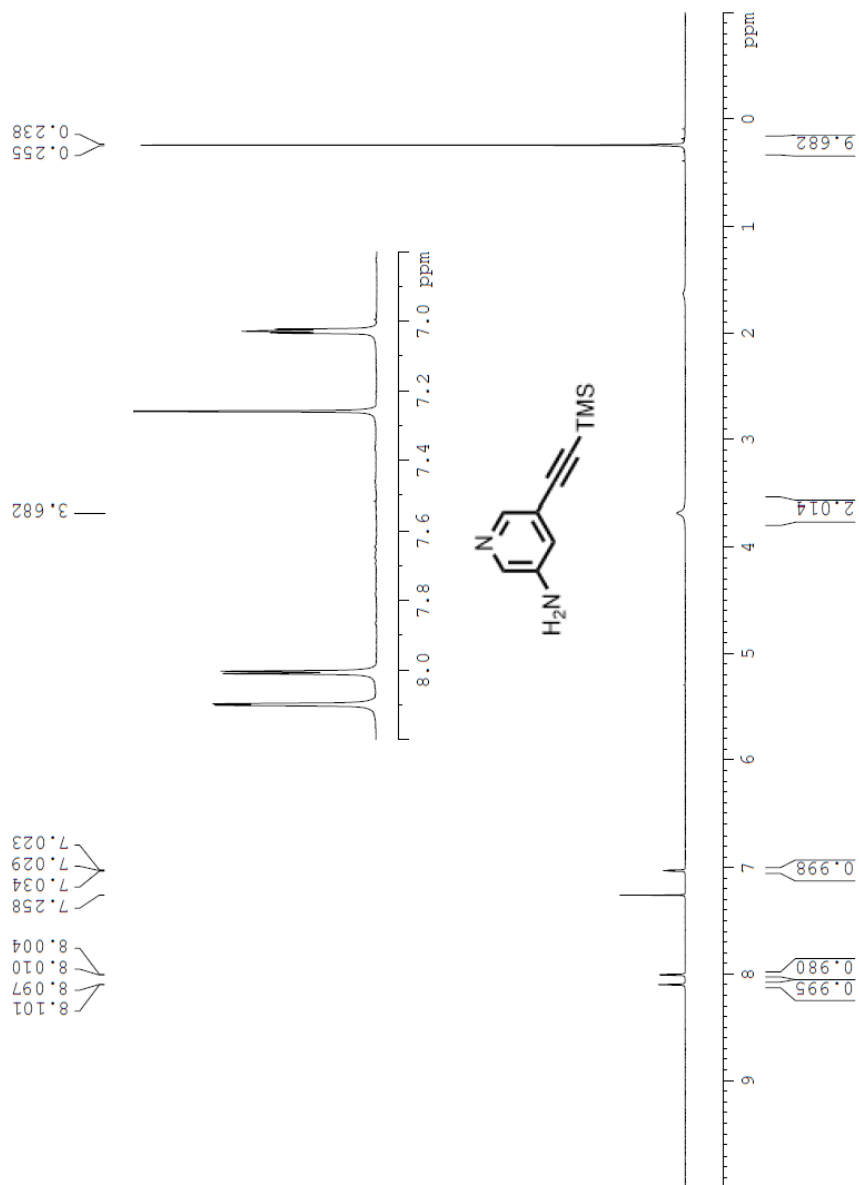


Figure S1. ^1H NMR of compound **5** in CDCl_3 .

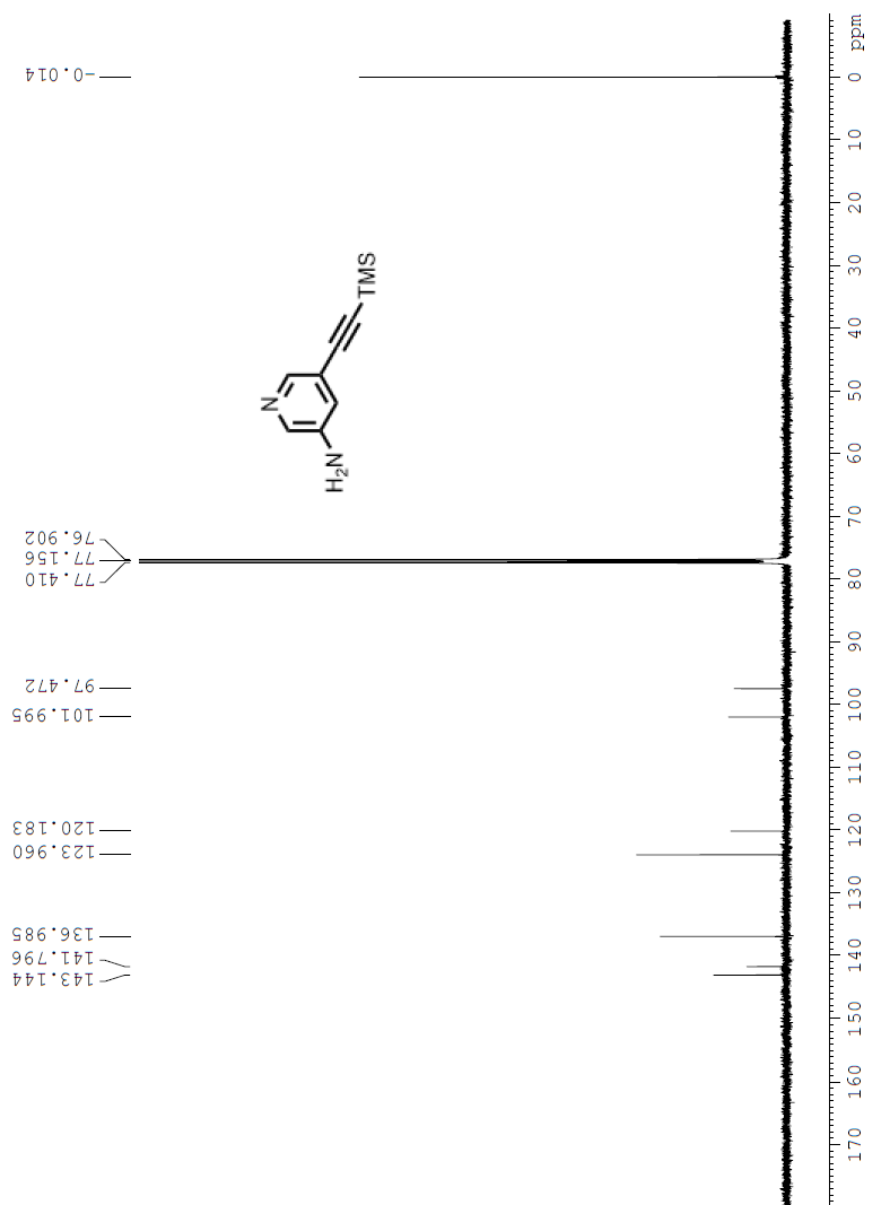


Figure S2. ^{13}C NMR of compound 5 in CDCl_3 .



Figure S3. ^1H NMR of compound **6a** in CDCl_3 . *Inset:* Expansion of the spectrum (7.1–8.3 ppm) after apodization of FID (LB = -1, GB = 0.3), showing the broad NH peak next to the sharpened peak of CHCl_3 .

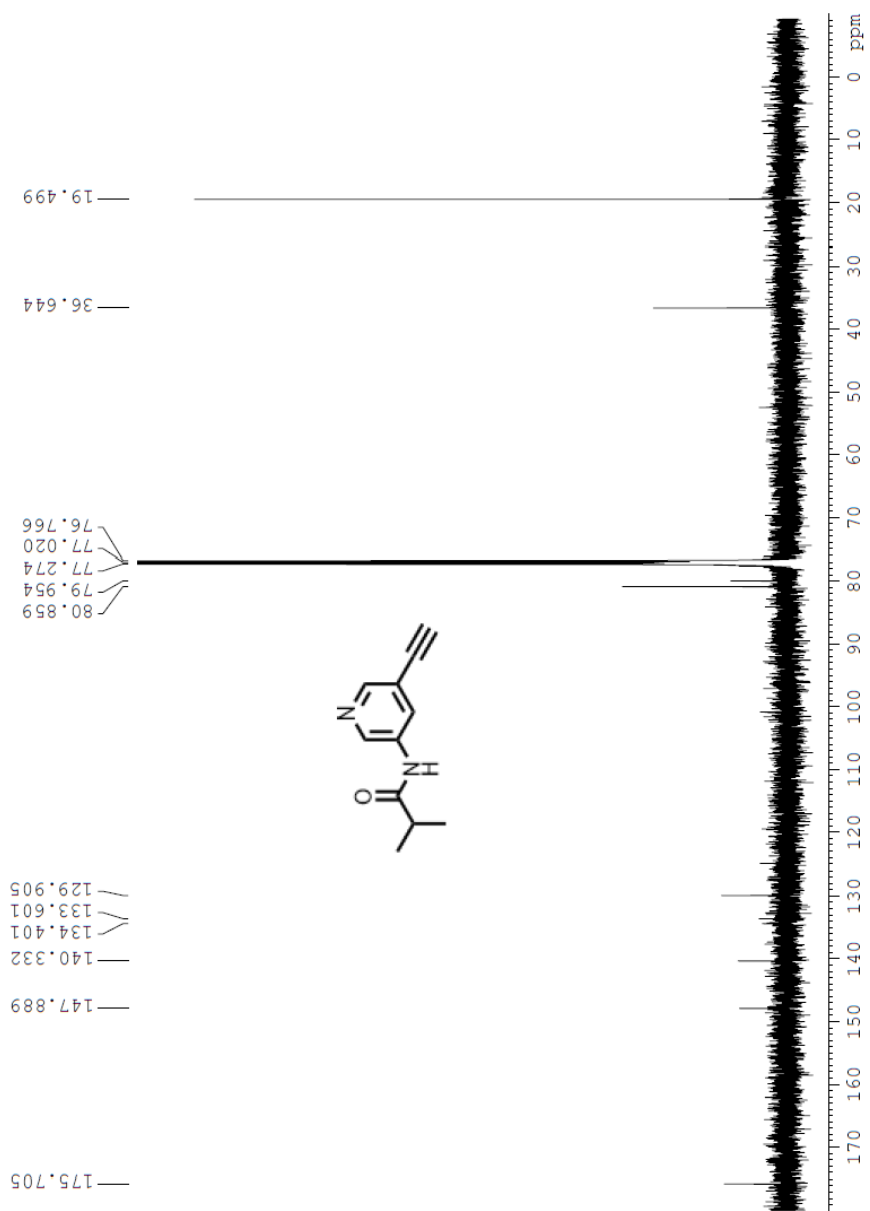


Figure S4. ^{13}C NMR of compound 6a in CDCl_3 .

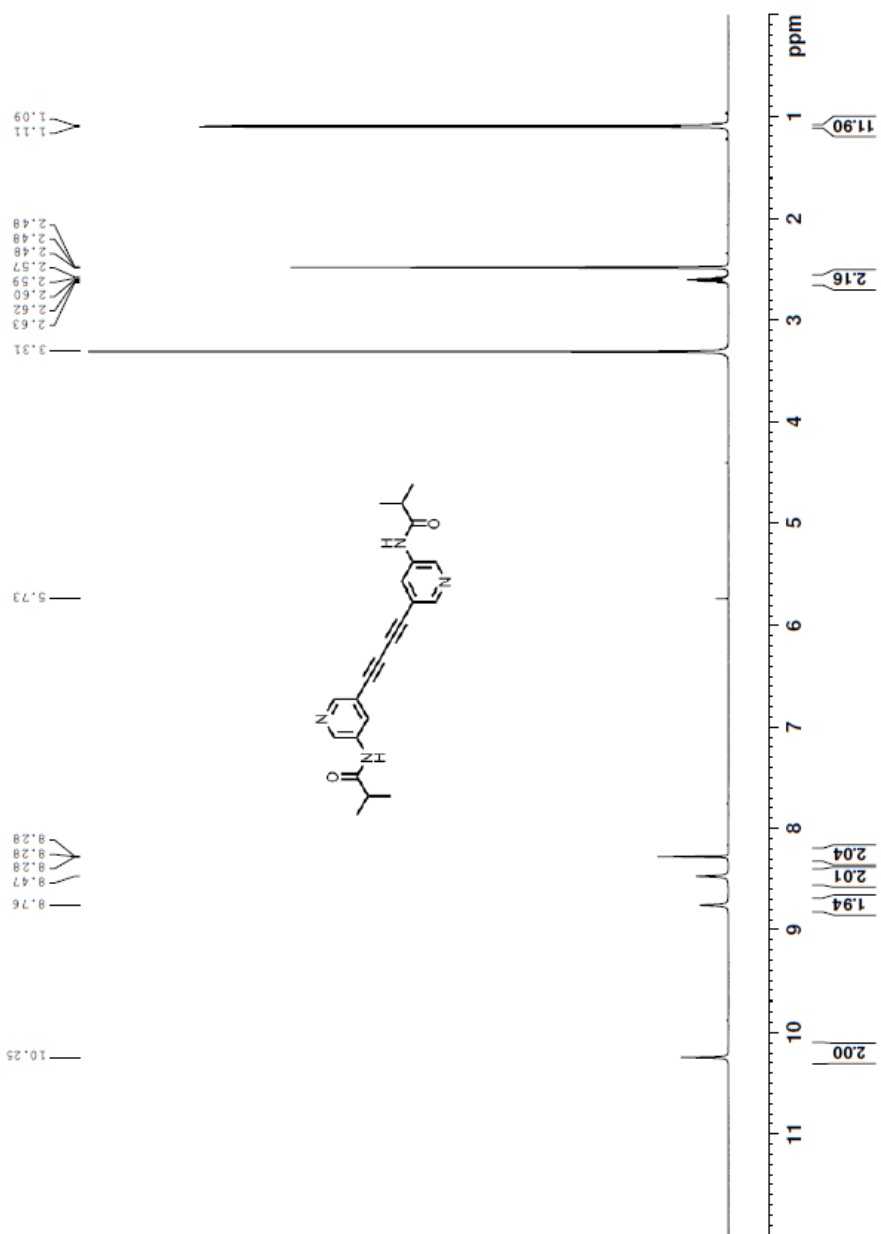


Figure S5. ¹H NMR of compound **3a** in d₆-DMSO.

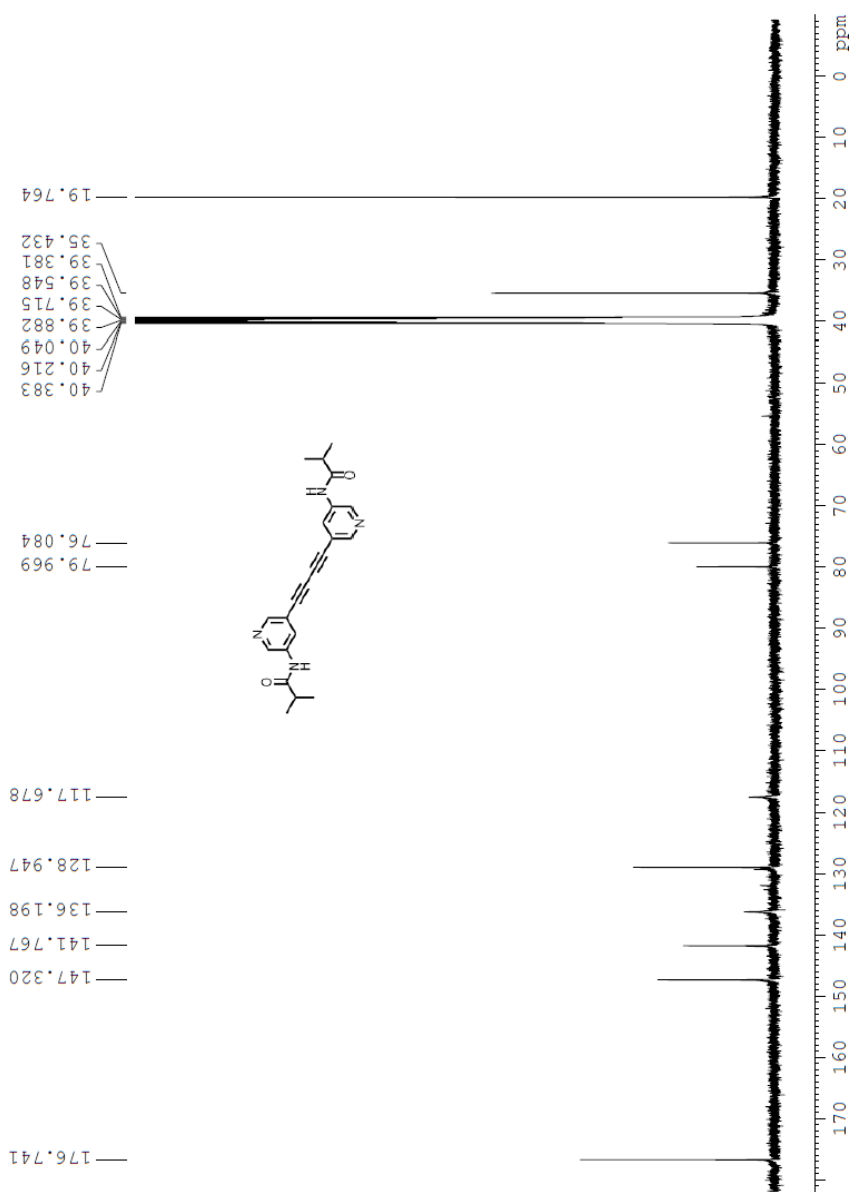


Figure S6. ^{13}C NMR of compound **3a** in $\text{d}_6\text{-DMSO}$.

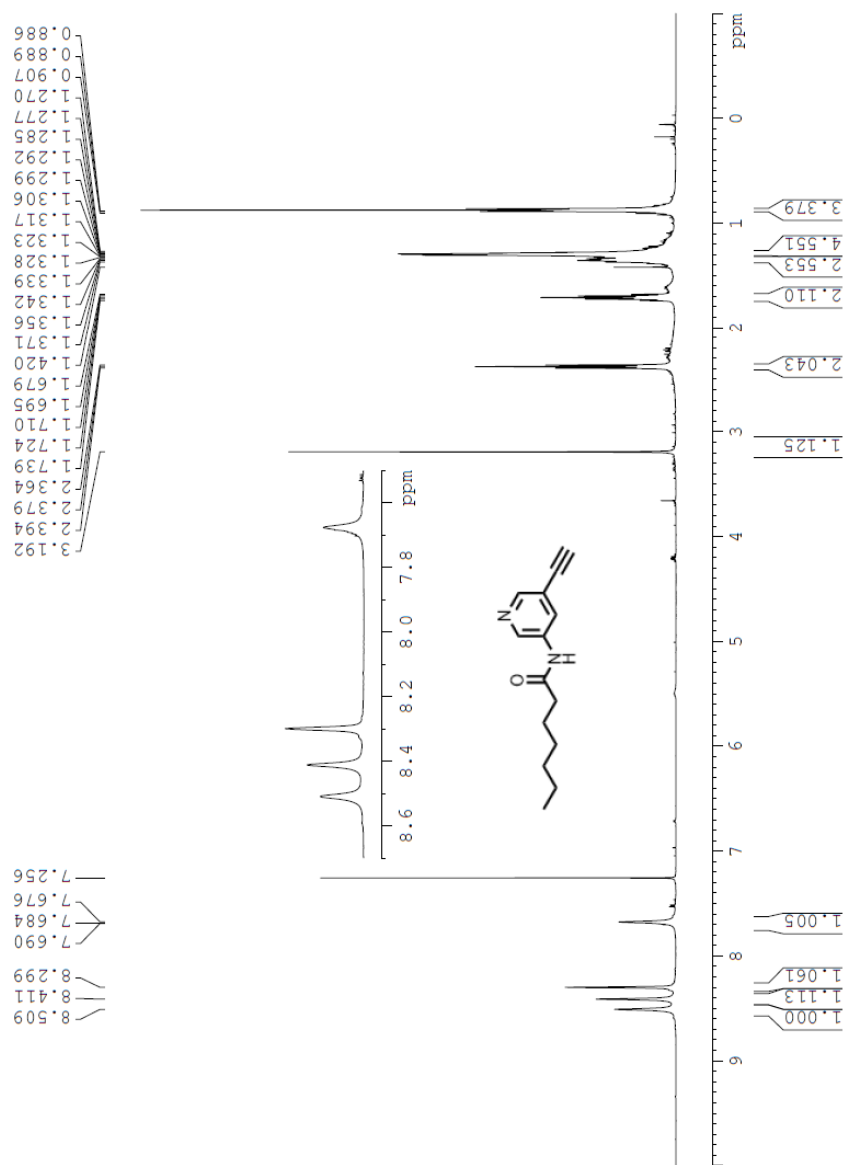


Figure S7. ¹H NMR of compound **6b** in CDCl₃.

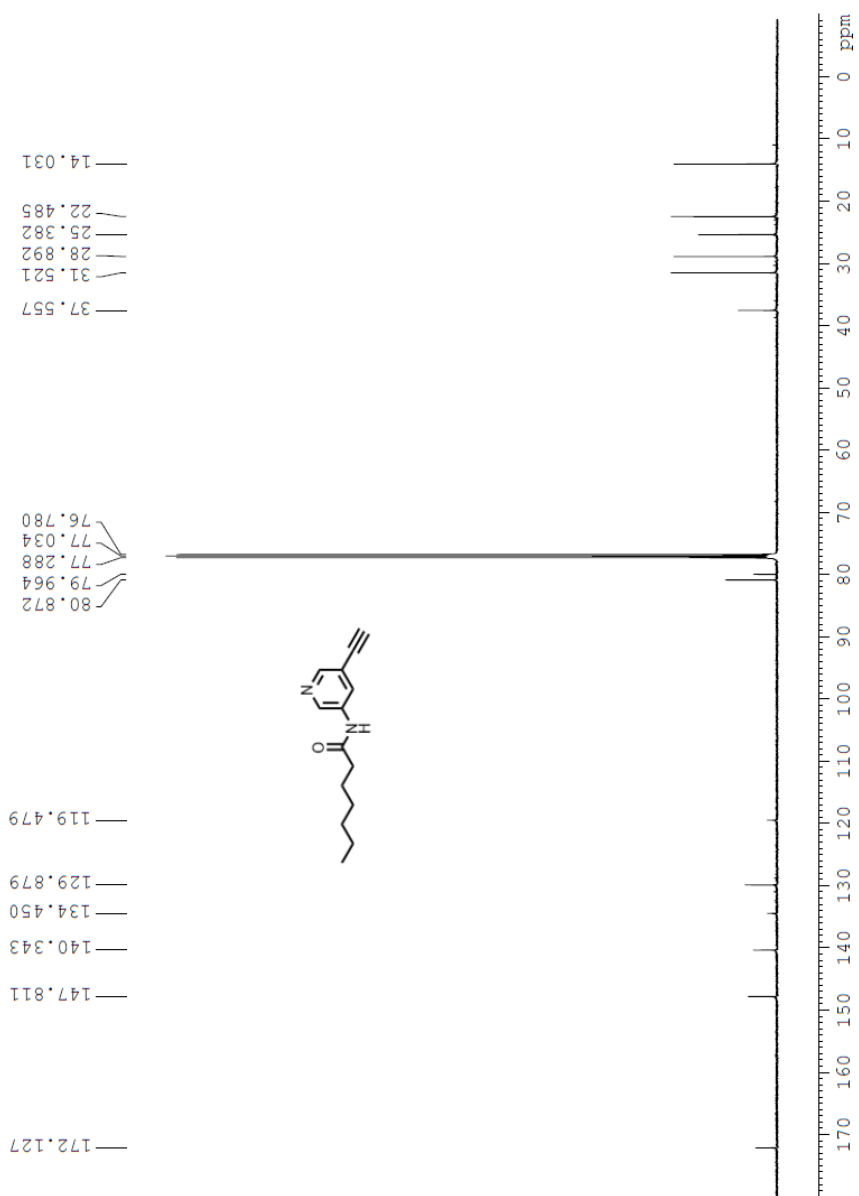


Figure S8. ^{13}C NMR of compound **6b** in CDCl_3 .

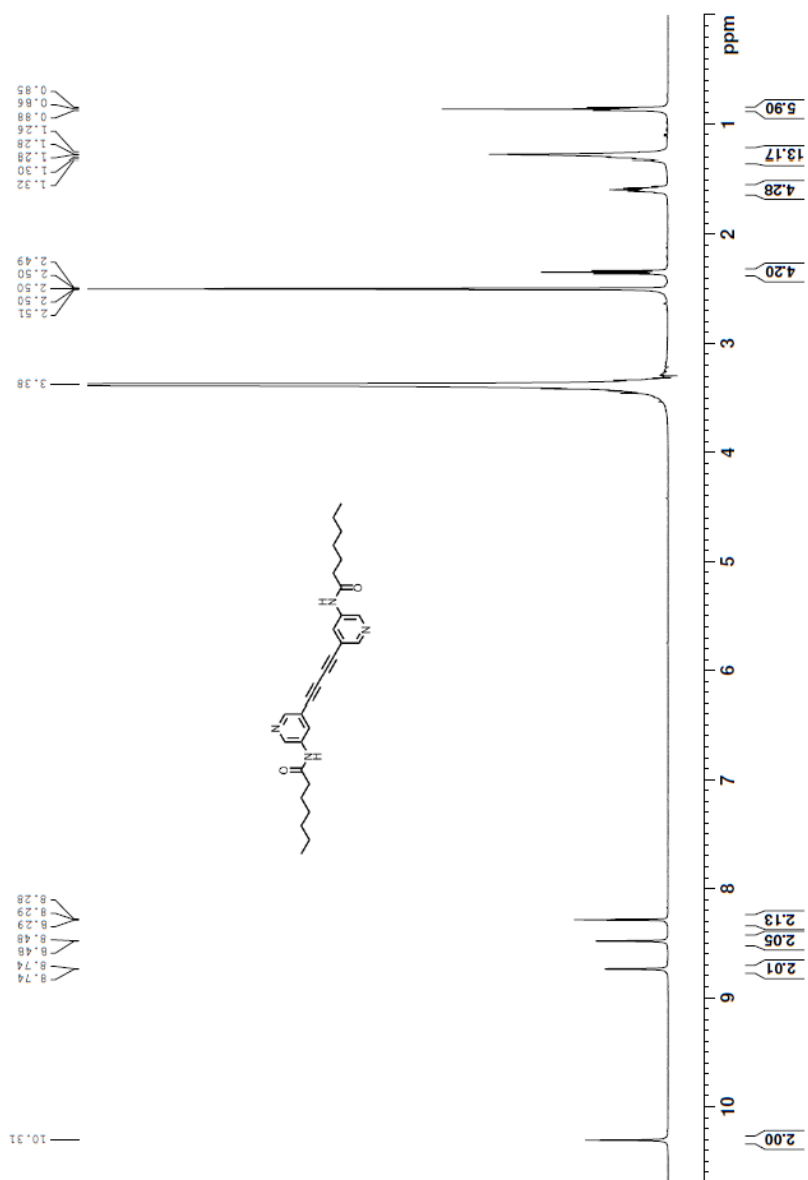


Figure S9. ¹H NMR of compound **3b** in d₆-DMSO.

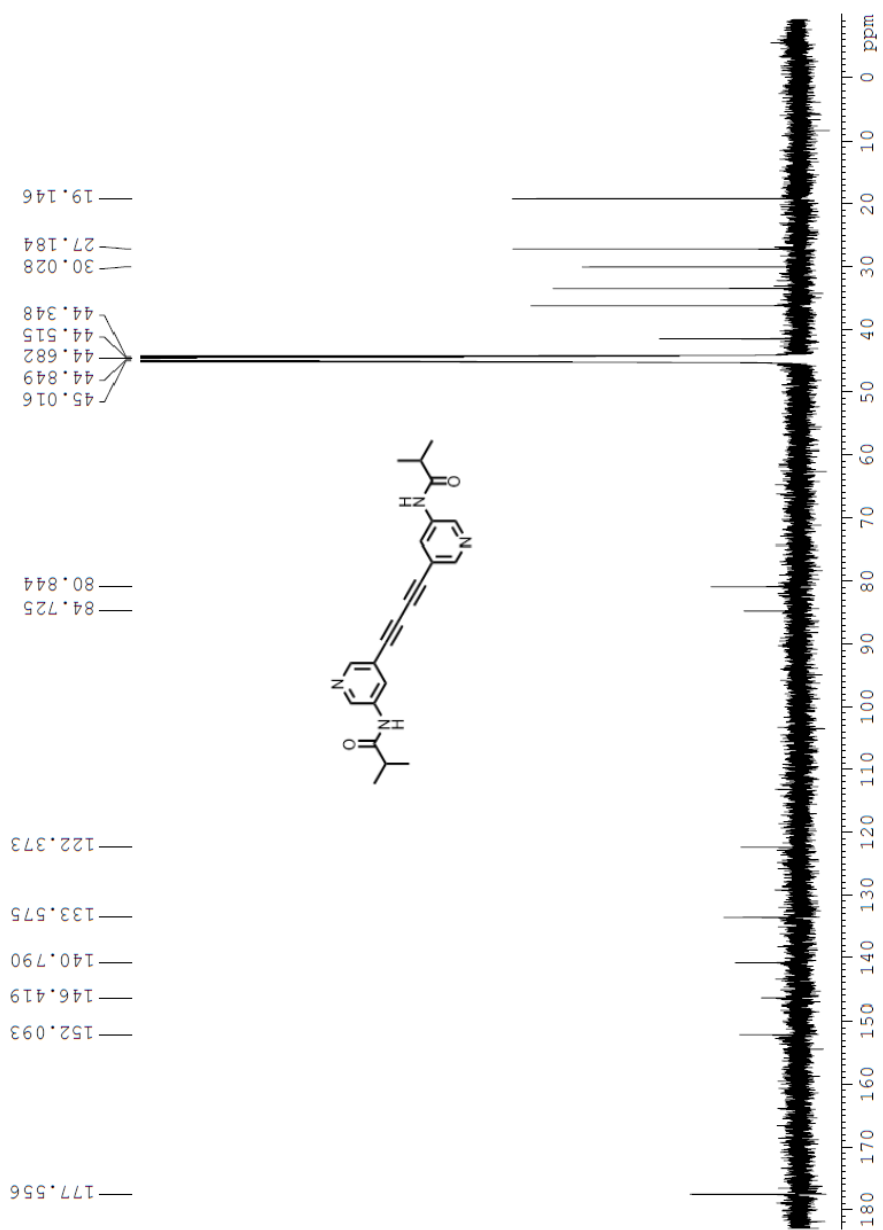


Figure S10. ^{13}C NMR of compound **3b** in $\text{d}_9\text{-DMSO}$.

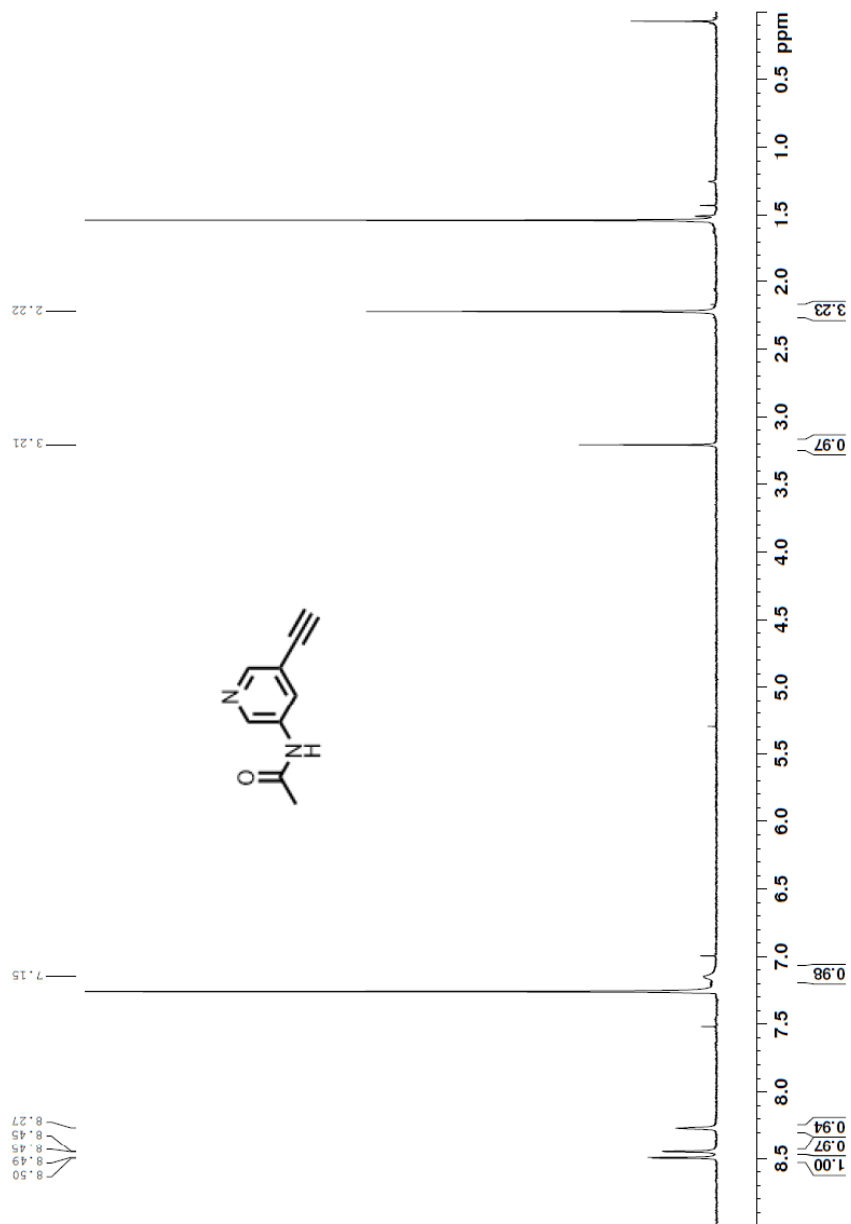


Figure S11. ¹H NMR of compound **6d** in CDCl₃.

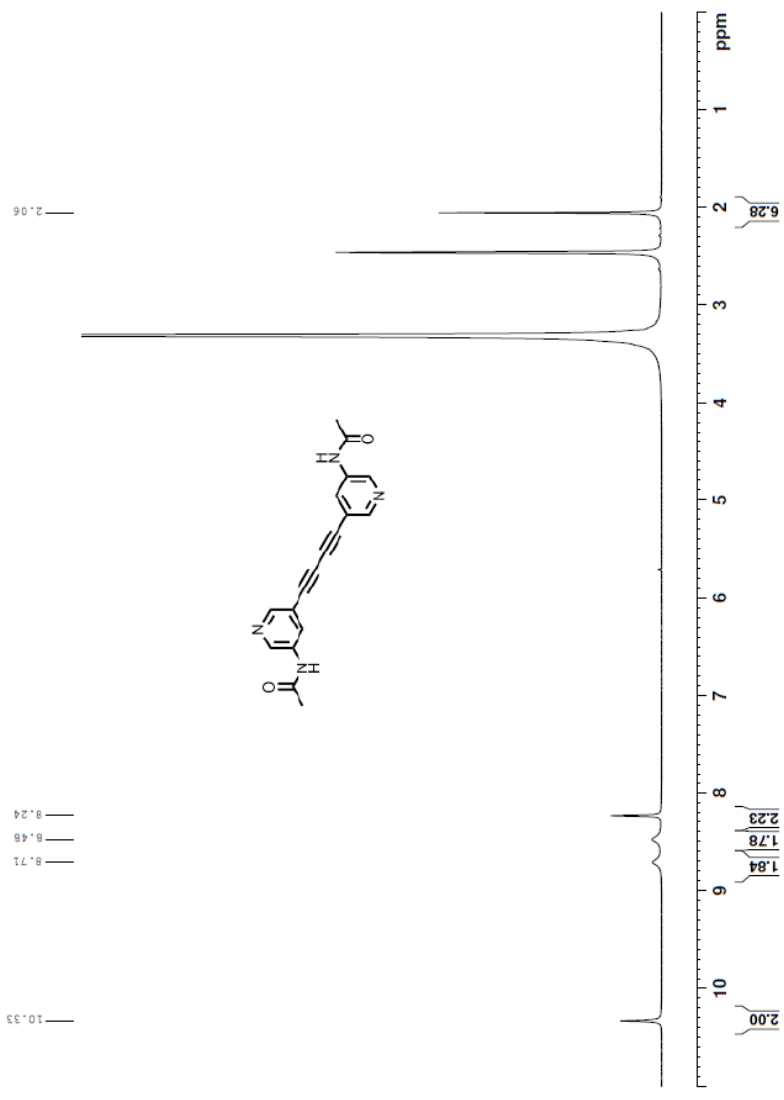


Figure S12. ¹H NMR of compound **3d** in d₆-DMSO.

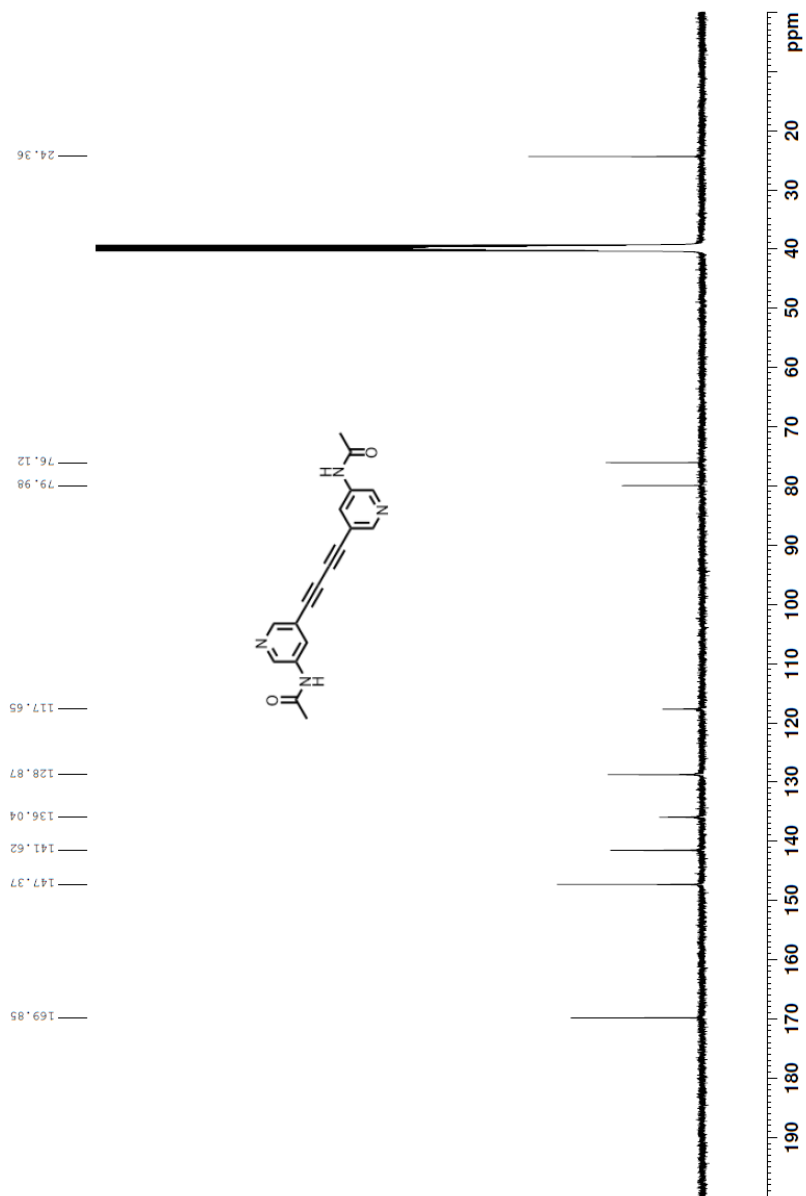


Figure S13. ^{13}C NMR of compound **3d** in d_6 -DMSO.

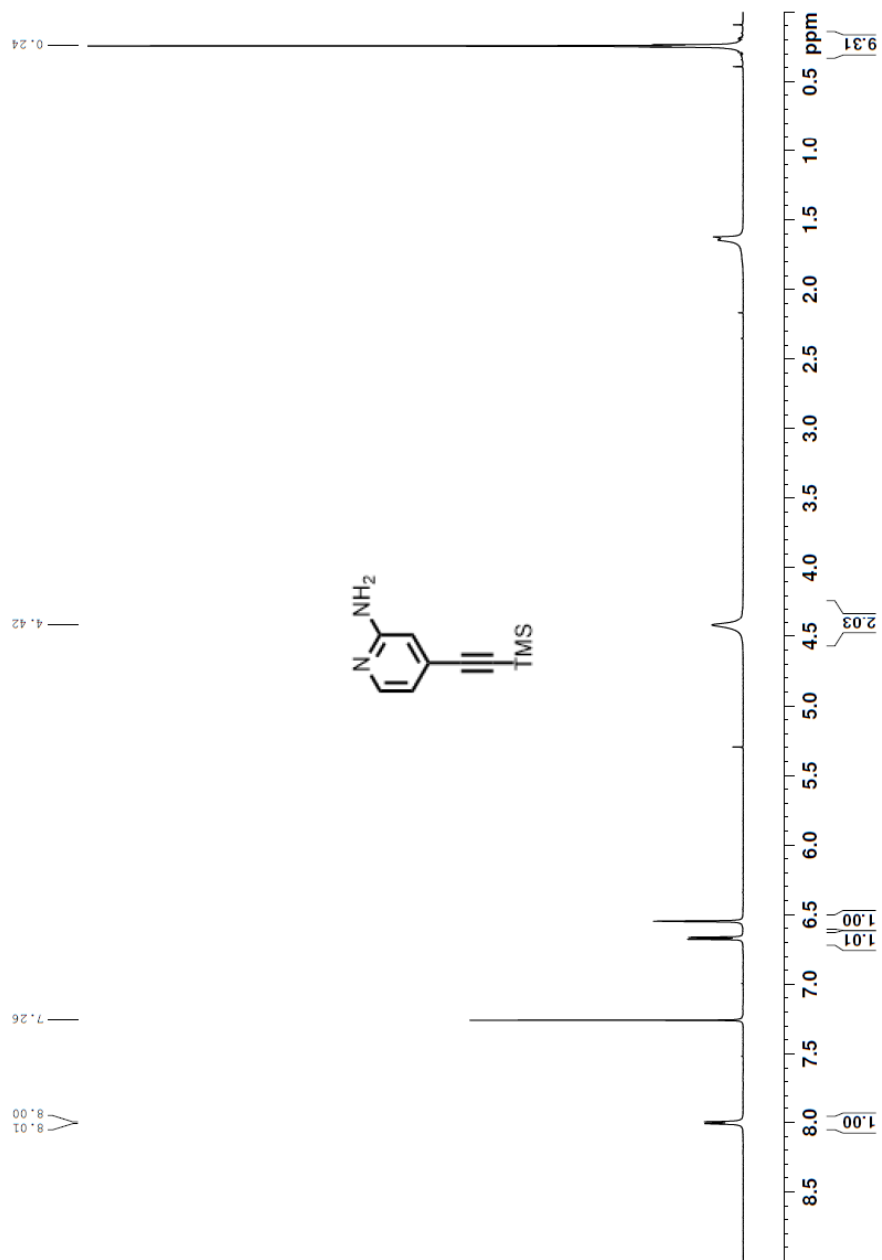


Figure S14. ¹H NMR of compound 11 in CDCl₃.

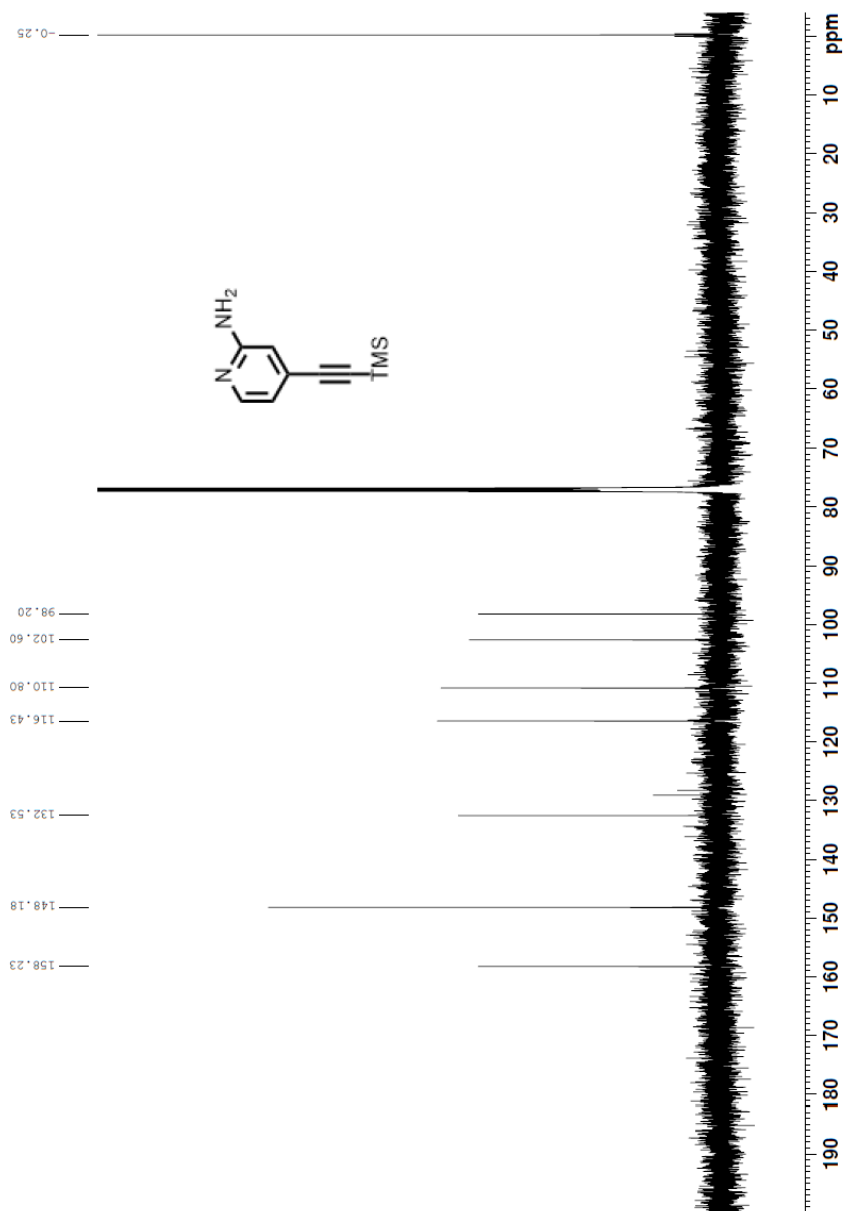


Figure S15. ^{13}C NMR of compound **11** in CDCl_3 .

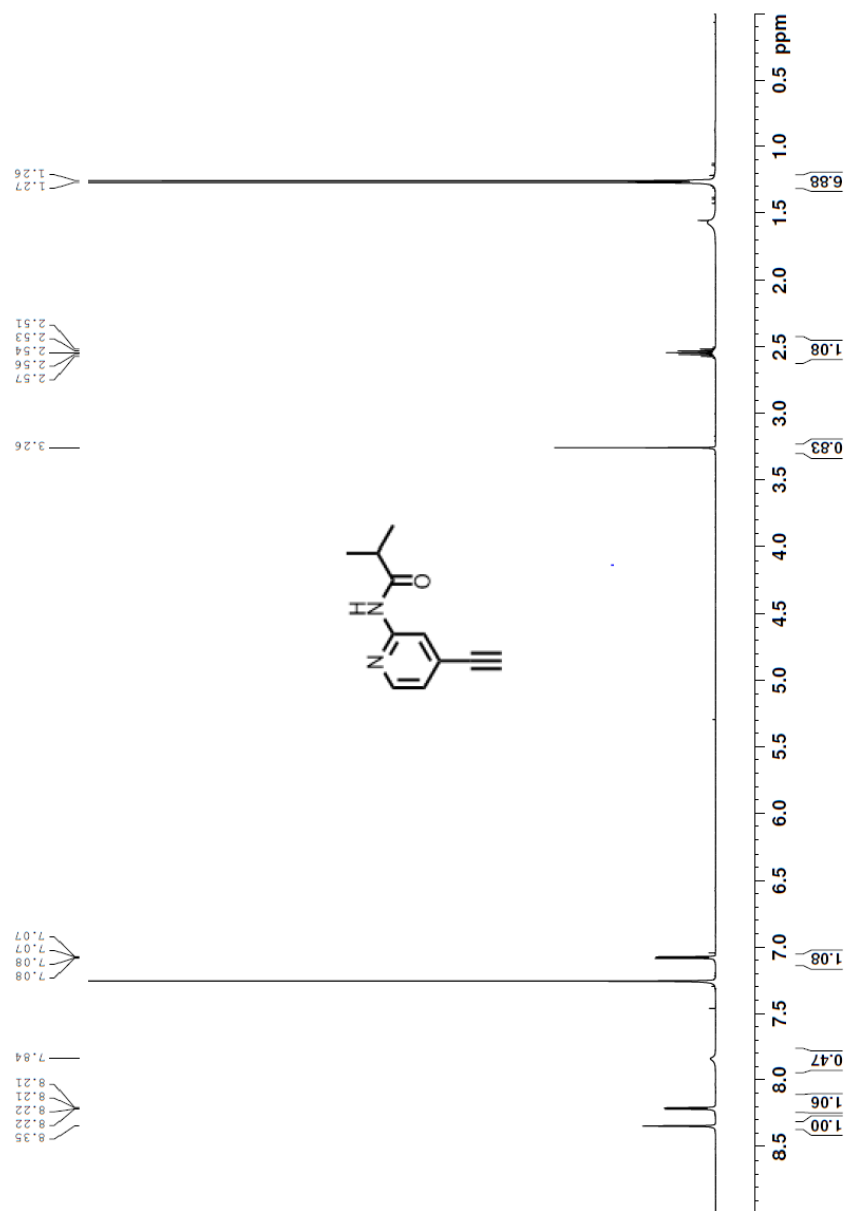


Figure S16. ¹H NMR of compound **12** in CDCl₃.

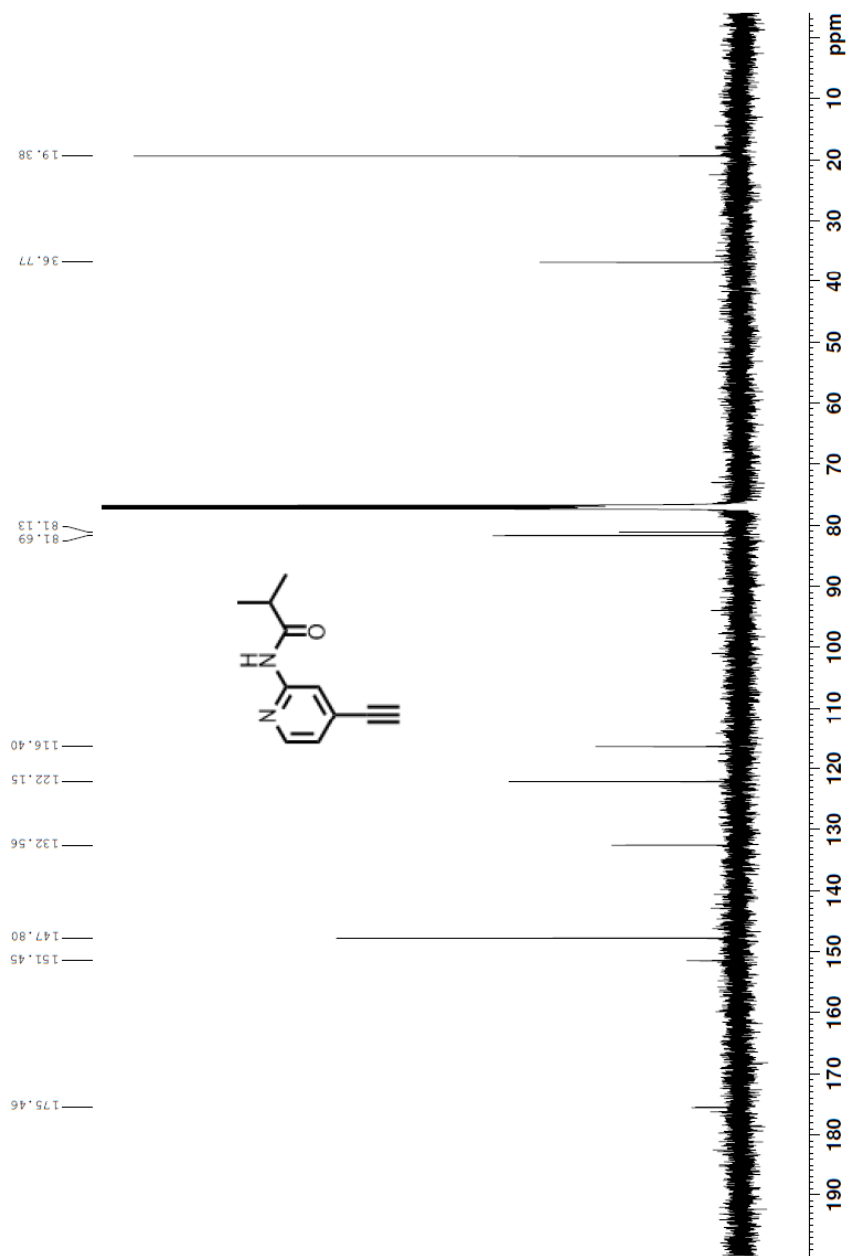


Figure S17. ¹³C NMR of compound **12** in CDCl₃.

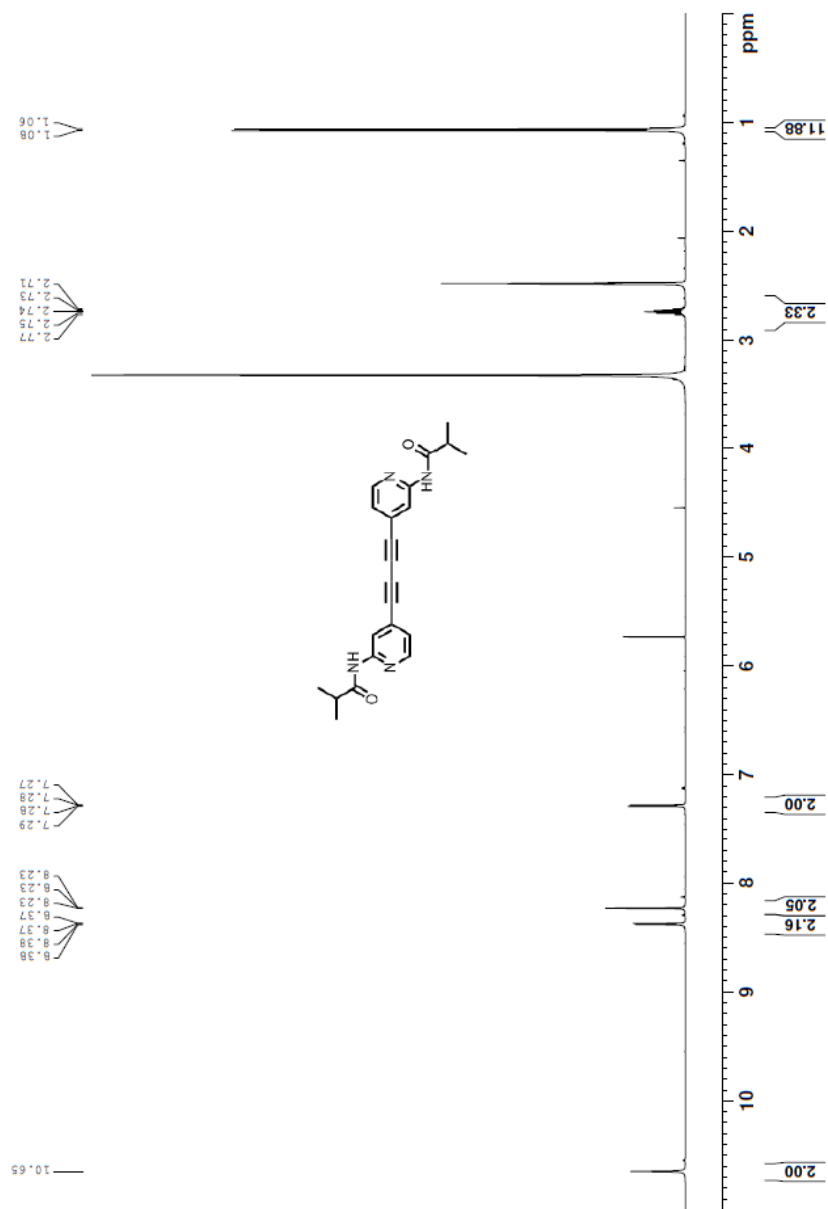


Figure S18. ¹H NMR of compound **13** in d₆-DMSO.

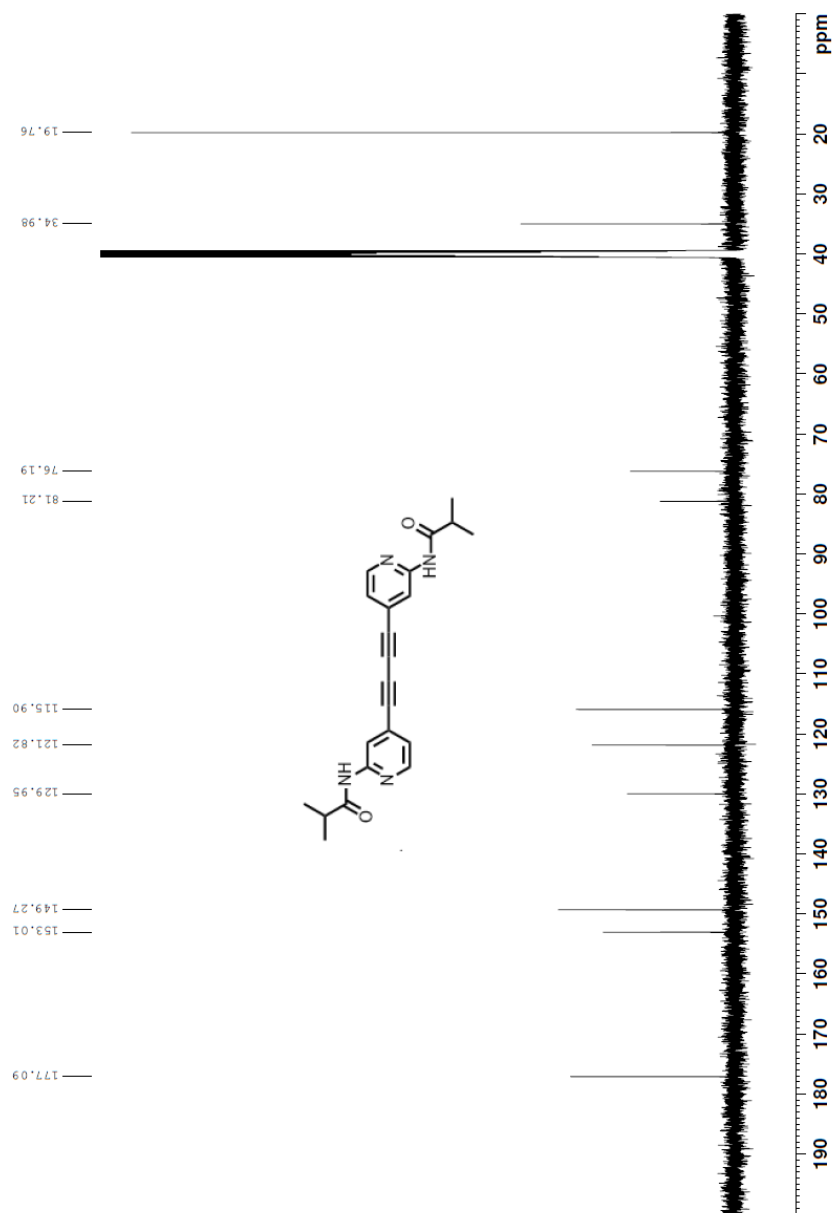


Figure S19. ¹³C NMR of compound **13** in d₆-DMSO.

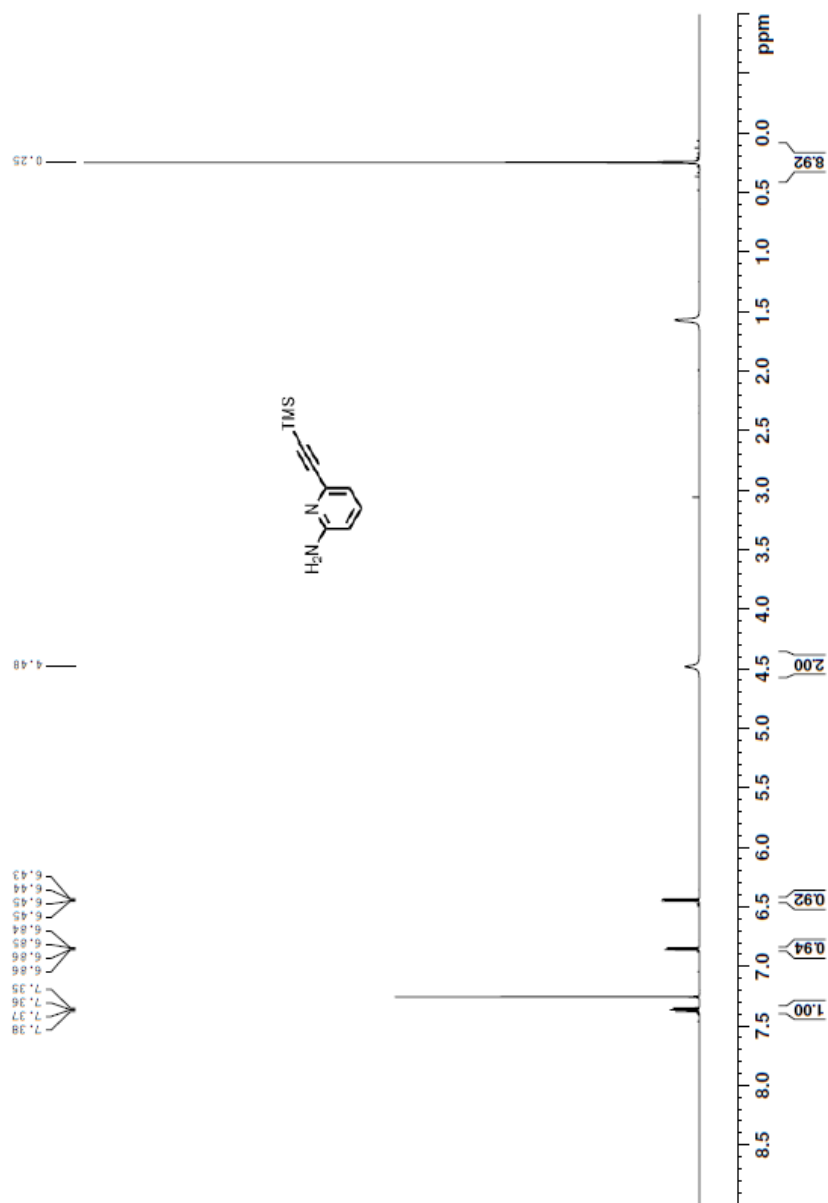


Figure S20. $^1\text{H NMR}$ of compound 14 in CDCl_3 .

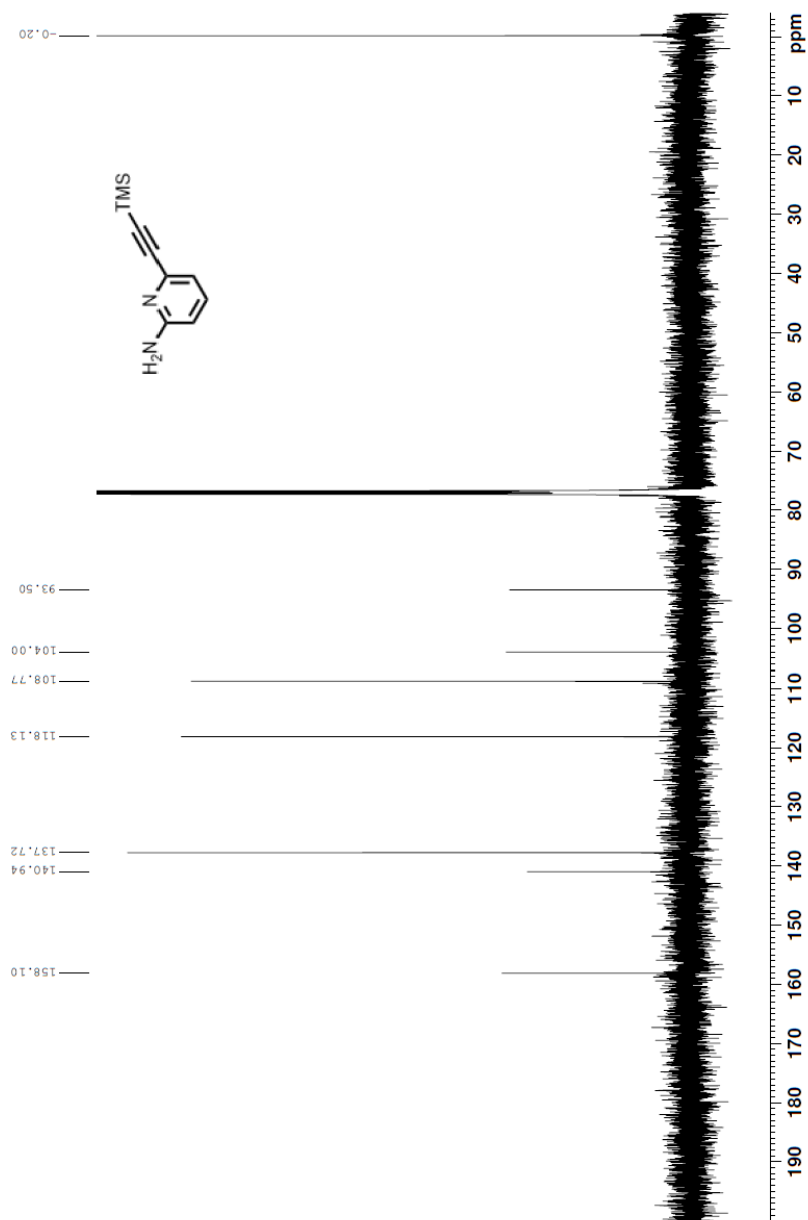


Figure S21. ^{13}C NMR of compound 14 in CDCl_3 .



Figure S22. ¹H NMR of compound 15 in CDCl₃.

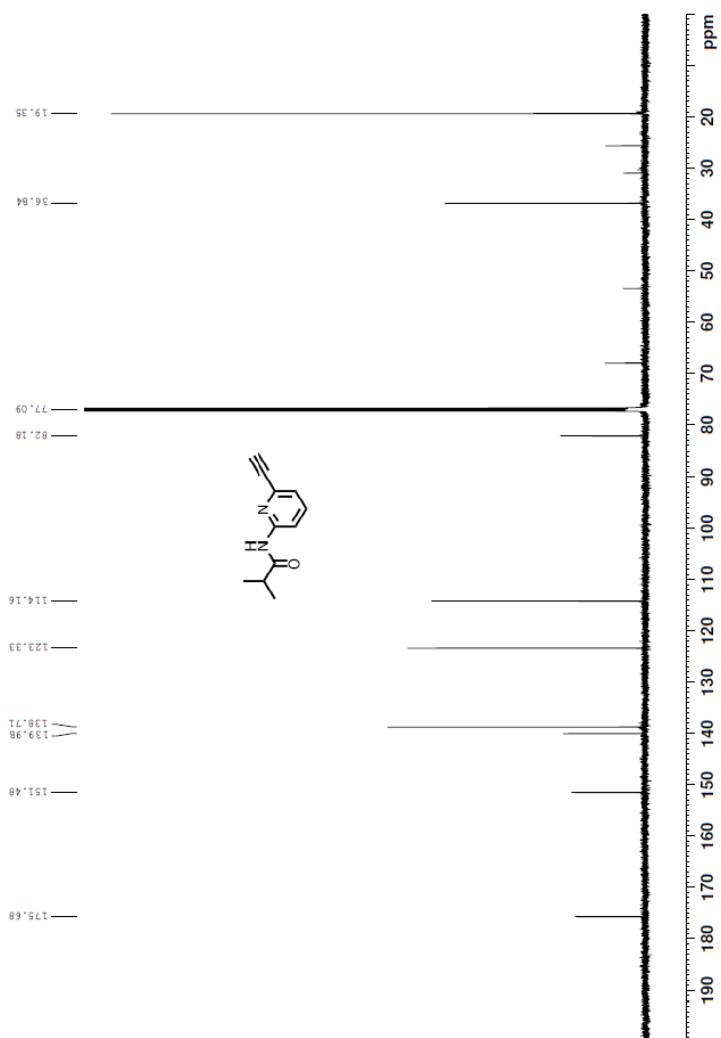


Figure S23. ^{13}C NMR of compound **15** in CDCl_3 .

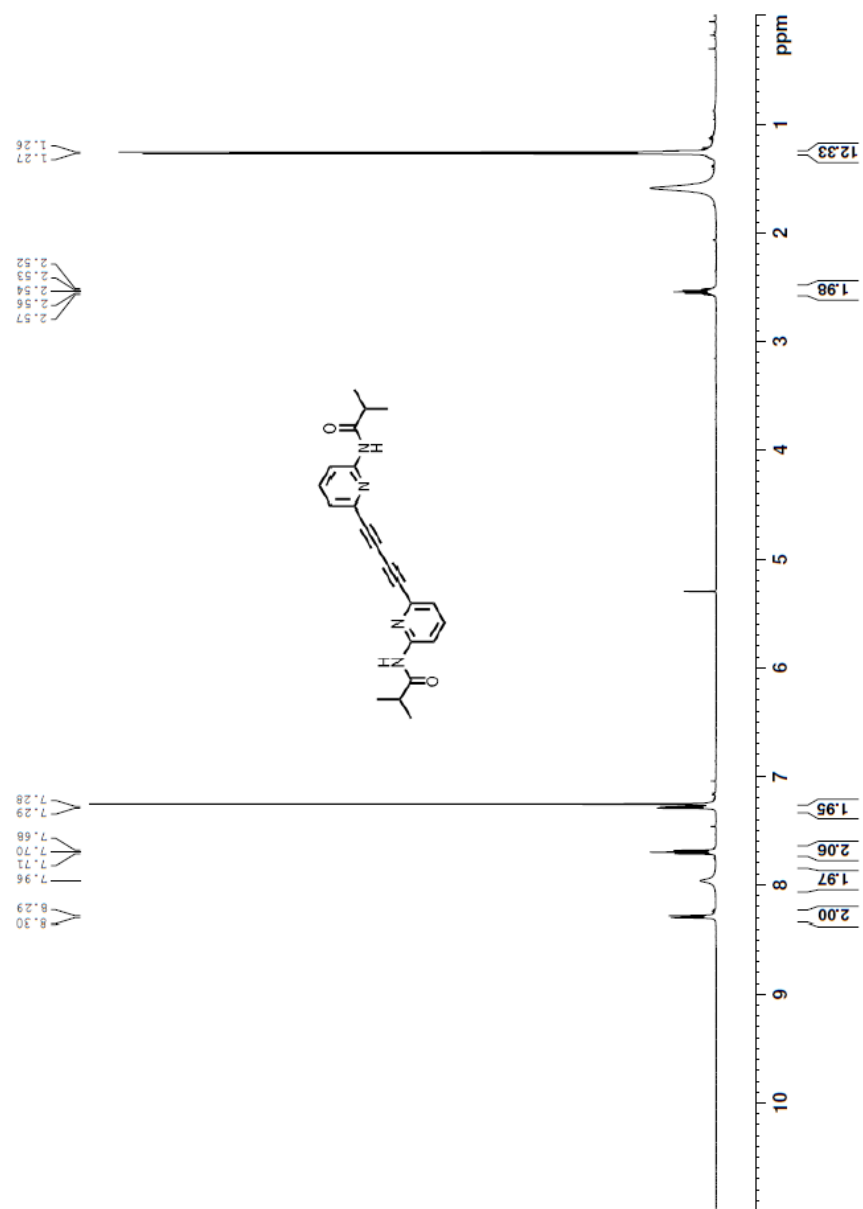


Figure S24. ¹H NMR of compound 16 in CDCl₃.

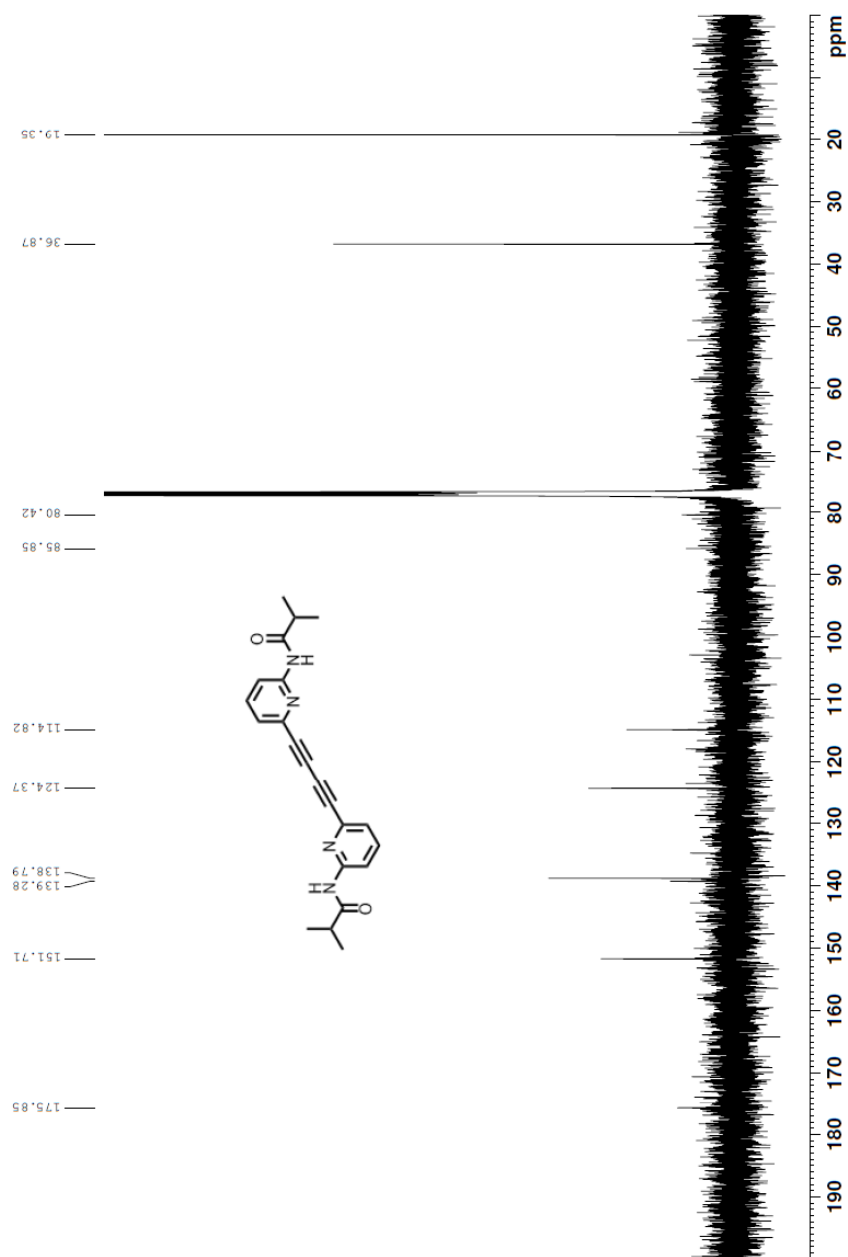
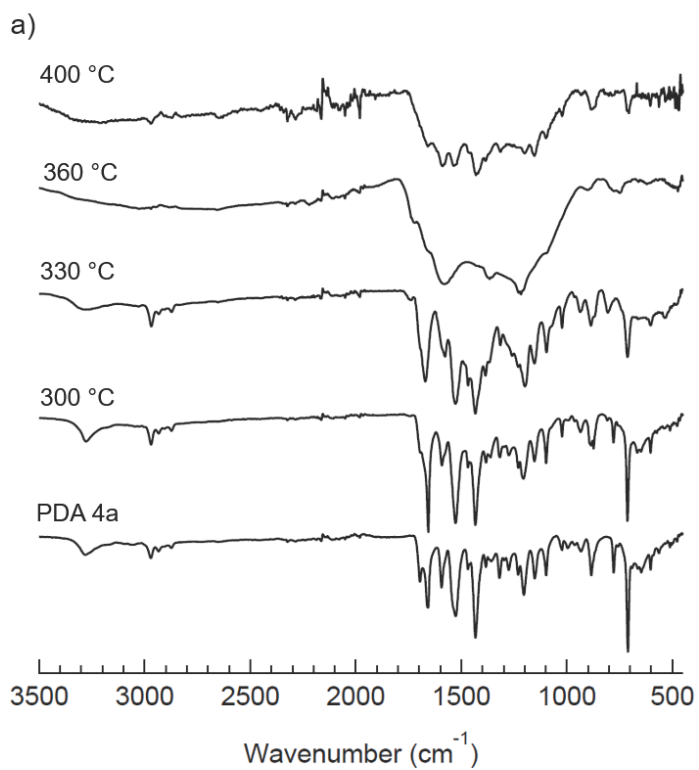


Figure S25. ¹³C NMR of compound 16 in CDCl₃.

1.3. FT-IR spectra of GNRs 1a and 1b.



(see Figure caption next page)

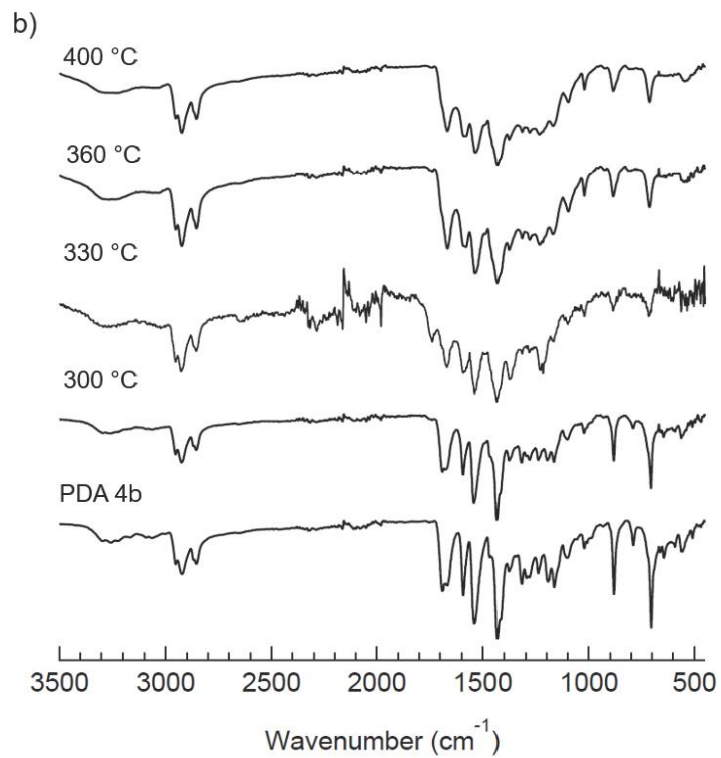


Figure S26. FT-IR Spectra of a) PDA **4a** and b) PDA **4b**, respectively, and of their thermally generated intermediates during their conversion to GNRs **1a** and **1b**, respectively.

1.4. UV-Vis spectra of GNRs 1a and 1b.

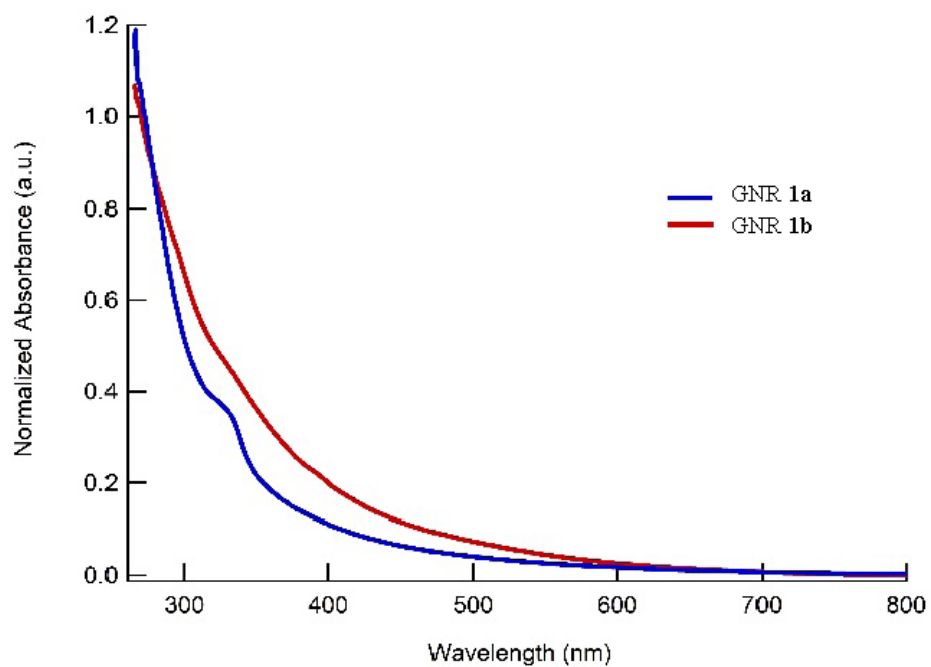


Figure S27. Normalized UV-Vis spectra of extracts of GNRs 1a,b in N-methyl-2-pyrrolidone. 1 mg/mL solutions were sonicated overnight then passed through a 0.2 μm Teflon membrane filter.

Section 2: Crystallography Details

Microfocus X-ray data collection and structure determination

To obtain crystals of appropriate size and order for diffraction, crude product was dissolved in boiling methanol at 1 mg/mL and filtered while hot through packed cotton. The filtrate was sealed in a standard 20-mL scintillation vial and protected from light with aluminum foil. Large crystal clusters formed after about a week (Figure S11a). For extraction, the crystal clusters were transferred to glass 9-well spot plates with fresh methanol constantly added to prevent the crystals from drying. During extraction, crystals were mounted on CryoLoops™ from Hampton Research after dragging them through Parabar 10312 (Paratone®) for cryoprotection and flash-cooled in liquid nitrogen.

Diffraction data were collected under cryogenic conditions (100 K) on beamline 24-ID-C at the Advanced Photon Source (APS) equipped with a Dectris PILATUS 6MF pixel detector array using a 10 μm beam with wavelength 0.775 \AA at a detector distance of 150 mm (0.85 \AA resolution at detector edge). 720 diffraction images were collected with an oscillation range of 0.5° from a single crystal. Three image ranges, consisting of 542/720 frames, were indexed and integrated separately by *XDS*¹. The outputs from *XDS* were sorted and merged in *XSCALE* and yielded *ab initio* solutions using *SHELXD*² and *SHELXT*³. The atomic positions were refined against measured data using *SHELXL*⁴.

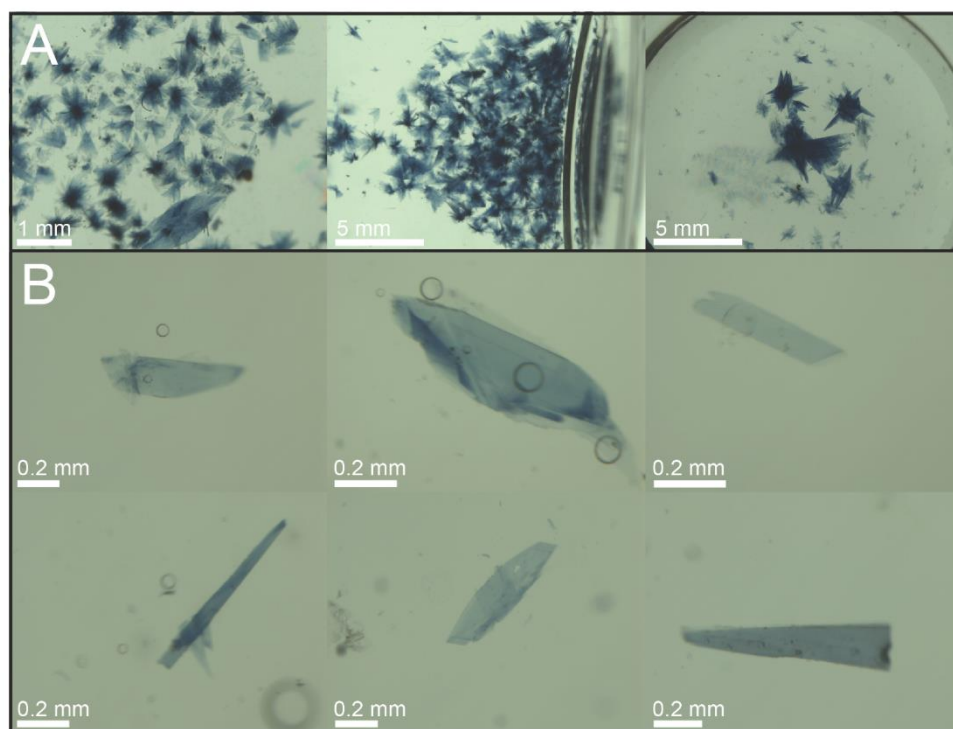


Figure S28. A) Crude compound **3a** was recrystallized from methanol, resulting in large crystal clusters. B) Individual crystals were isolated from the crystal clusters and flash-frozen for microfocus x-ray data collection. The plate-like morphology suggests preferential growth along two unit cell vectors.

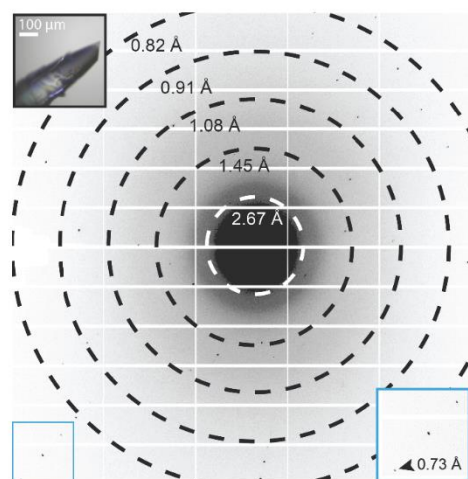


Figure S29. Single crystal diffraction pattern of compound **3a** measured during microfocus x-ray data collection. The pattern corresponds to a 0.5° wedge of reciprocal space. The blue inset on the bottom right shows a magnified image of the boxed region at the bottom left of the diffraction pattern. Diffraction was observed to the corners of the detector. The black inset contains the in-line crystal image with a scale bar corresponding to $100\ \mu\text{m}$. Resolution circles are indicated by the dotted rings.

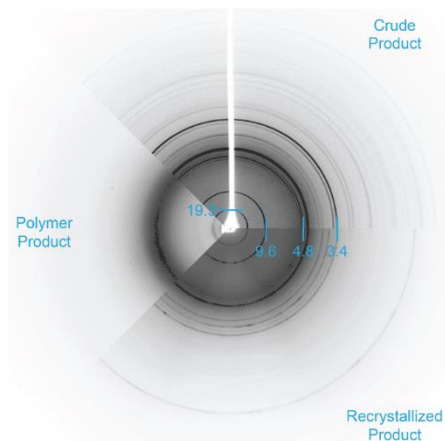


Figure S30. Comparison of powder diffraction patterns of compound **3a**, measured from crude product (top right), recrystallized product (bottom right), and polymer product (left). Powder rings at $4.8\ \text{\AA}$, $9.6\ \text{\AA}$, and $19.3\ \text{\AA}$ correspond to unit cell vectors a and c while rings related to unit cell vector b are notably missing, suggesting that the intermolecular interactions in the direction of vector b are significantly weaker than those associated to vectors a and c . In addition to the rings corresponding to the unit cell dimensions, a ring at $3.4\ \text{\AA}$ reflects arrangement of the diynes into a polymer growth axis.

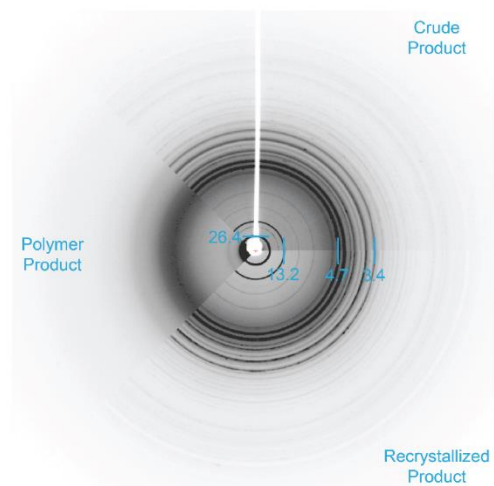


Figure S31. Comparison of powder diffraction patterns of compound **3b**, measured from crude product (top right), recrystallized product (bottom right), and polymer product (left). Strong powder rings at 4.7 Å and 26.4 Å likely correspond to unit cell vectors. The rings in the lower resolution region are commensurate to the presence of a larger R-group in **2b** compared to **2a**. Furthermore, a ring at 3.4 Å suggests a similar arrangement of diyne units into the polymer growth axis seen in **2a** (see Figure 2 in main text).

Table S1. Crystallographic Data for **3a**.

Data Collection and Processing	
Molecular Formula	C ₂₂ H ₂₂ N ₄ O ₂
No. of crystals merged	1 (3 wedges)
Space Group	P -1
Cell dimensions	
a, b, c (Å)	4.80 5.29 19.31
α , β , γ (°)	89.882 87.884 76.199
Resolution Limit (Å)	0.85 (0.88-0.85)
Wavelength (Å)	0.77490
Rmerge (%)	3.1 (4.6)
I/ σ ₁	25.57 (21.32)
Completeness	83.0 (81.7)
CC _{1/2}	99.9 (99.5)
No. total reflections	3822 (478)
No. unique reflections	1356 (165)
Multiplicity	2.82 (2.90)
Refinement	
Resolution (Å)	0.85
No. of Reflections	1356
R1 (%)	4.58
wR2 (%)	13.14
GooF (Restrained GooF)	1.087 (1.087)

Section 3: High-Resolution Transmission Microscopy (HR-TEM)

HR-TEM samples of graphene nanoribbons fjord-edge N₂[8]GNRs **1a** and **1b** were prepared by drop-casting dispersed samples of PDAs **4a** and **4b** onto silicon wafers, subsequently converting them to the GNRs through heating at 400 °C in a quartz tube furnace. Accordingly, dispersed suspensions obtained by sonicating the PDAs in N-methylpyrrolidone (NMP) for 12 hours were drop-cast onto silicon wafers, which after drying, were transferred to a 2-cm wide quartz tube placed in a high-temperature oven, followed by heating at 400 °C for 1 hour under argon flow to effect their conversion to fjord-edge N₂[8]GNRs **1a** and **1b**. These newly formed GNRs were then mechanically transferred in air to a TEM grid by brushing off the grid against the GNR-coated silicon wafer surface.

HR-TEM imaging of both fjord-edge N₂[8]GNRs **1a** and **1b** reveals tight bundles of stacked nanoribbons (Figure S32). The nanoribbons within the bundles have a stacking distance of ~0.35 nm, indicating cofacial π - π stacking, which is the expected π -stacking distance for graphitic materials. However, the best repeat distances calculated for idealized packing structures for the *non-planar* fjord-edge N₂[8]GNR acetamide **1c** or its parent structure **1d** do not fit this observation (Figures 1, S33, and S34). Incidentally, the parent structure **1d** was calculated for simplicity while searching for the best packing arrangement of these nanoribbons, as well as for the possibility that subsequent losses of sidechains during the TEM manipulation (via electron impact) might have occurred.

Accordingly, molecular mechanics calculations on tetrameric and octameric aggregates of either fjord-edge N₂[8]GNR acetamide **1c** or the parent **1d** (using GNR sections of midsize length) do not reproduce this short distance, instead giving shortest repeat stacking distances of roughly 8-10 Å and 10-12 Å between units of **1c**, respectively, and ~6.2 Å and 10 Å, respectively, between units of **1d** (Figures S33 and S34). Thus, we believe that there is a transformation of fjord-edge N₂[8]GNRs **1a** and **1b** occurring under the high intensity electron beam (200 kV accelerating voltage) in the HR-TEM experiment, which induces *in-situ* graphitization of **1a** and **1b** to the planar GNRs **2a** and **2b** (Figures 1 in manuscript, and S32 below). The flattening of the nanoribbons during that process would result in the observed ~0.35 nm distance corresponding to cofacial π - π stacking. Despite our repeated attempts, performing lattice-resolved HR-TEM experiments at lower electron beam energies were unsuccessful due to the reduced contrast at lower electron dosages and the small size of the GNRs. Because electron beam irradiation may be a method to obtain the exciting planar GNRs **2a** and **2b**, we are currently exploring ways to achieve this preparatively.

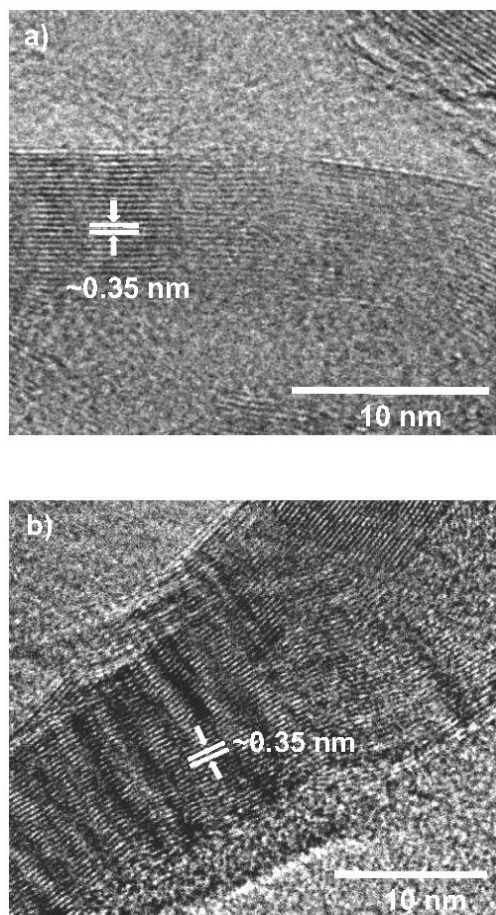


Figure S32. HR-TEM characterization of putative samples of (a) fjord-edge $N_2[8]$ GNR **1a**, and b) fjord-edge $N_2[8]$ GNR **1b**, both at higher magnification. Line profile across the GNR bundles shows that the average spacing between nanoribbons corresponds to the expected π - π stacking distance between flat nanoribbons. Thus, these images are more likely those of fully converted planar $N_2[8]_A$ GNR **2a** and **2b**.

Section 4: Computational Details

4.1. Molecular Mechanics Calculations of Packing Modes for fjord-edge N₂[8]GNRs **1c,d**

The geometries for the packing modes of fjord-edge N₂[8]GNR acetamide **1c** and the parent **1d** were obtained using the MMFF force field within the Spartan '18 Parallel Suite (Wavefunction, Inc).²⁹ Optimization of the relative position of each nanoribbon was carried out by moving one ribbon relative to the others using various geometries, but mainly keeping nanoribbons parallel in stacked or herringbone relationships. Initial minimization to obtain the tightest possible packing geometry exploited the overstated van der Waals interactions characteristic of the MMFF force field. The whole packing structure was then fully optimized at the MMFF level of theory. Only the most tightly packed structure for each of the tetrameric and octameric aggregates, respectively, for fjord-edge N₂[8]GNR acetamide **1c** or its parent structure **1d**, are shown here.

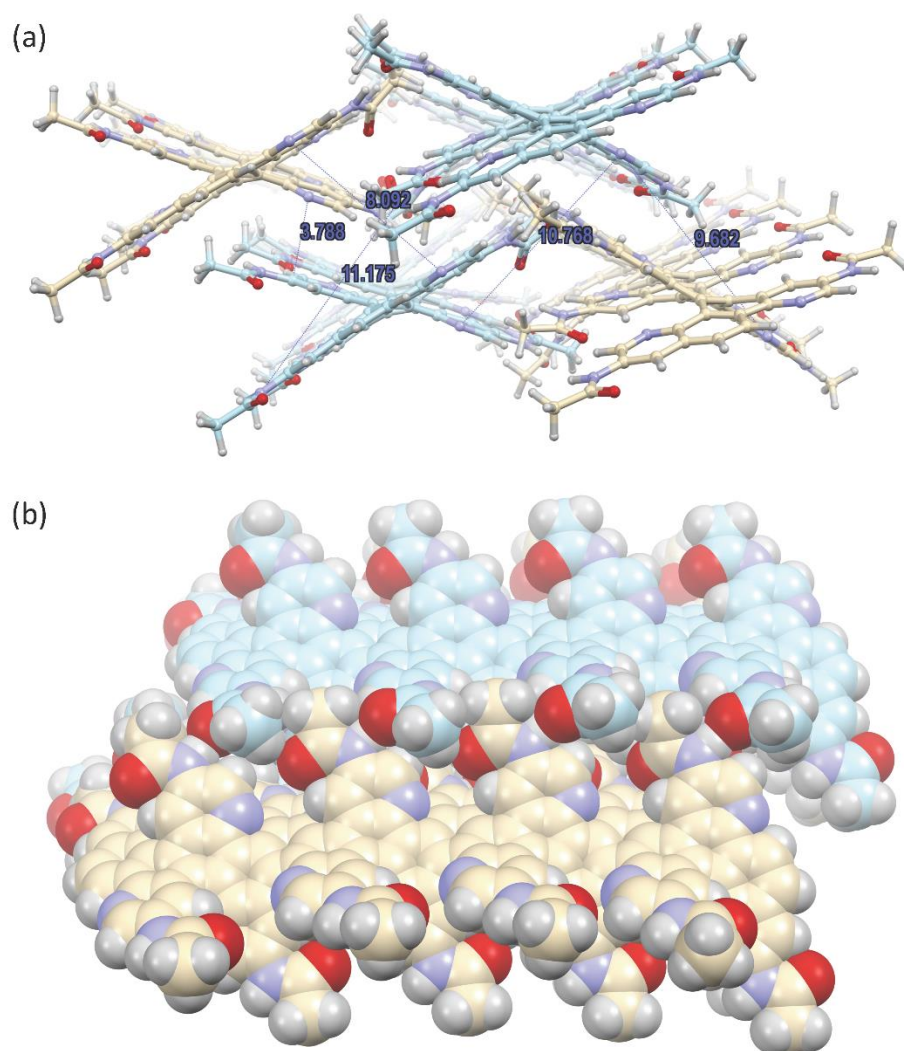


Figure S33. MMFF calculated packing structure for the tetrameric assembly of fjord-edge N₂[8]GNR acetamide model compound **1c**: (a) Ball and stick representation down the main axis of the nanoribbons, showing representative distances. The two repeat distances (not π - π interactions distances) are roughly 8-10 Å and 10-12 Å, respectively. (b) Space-filling representation highlighting the tight network of intermolecular H-bonding interactions between graphene nanoribbons. Light cyan and gold colors were used to distinguish each individual nanoribbon.

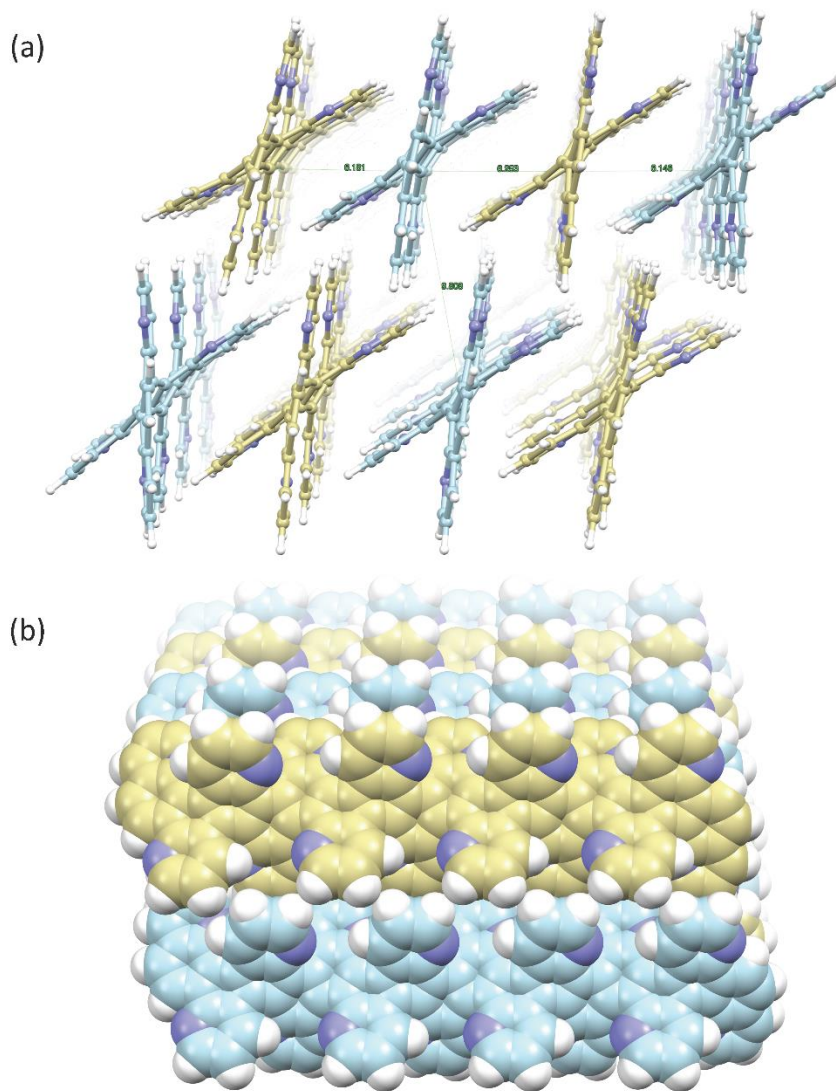
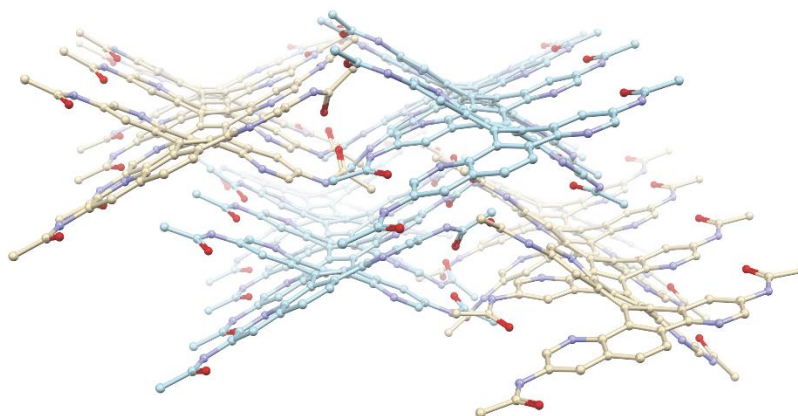


Figure S34. MMFF calculated packing structure for an octameric assembly of fjord-edge $N_2[8]GNR$ parent system **1d**: (a) Ball and stick representation down the main axis of the nanoribbons, showing representative distances. The two repeat distances (not π - π interactions distances) are ~ 6.2 Å and 10 Å, respectively. (b) Space-filling representation. Light cyan and gold colors were used to distinguish each individual nanoribbon.

4.2. Packing Structure Coordinates for fjord-edge N₂[8]GNR Acetamide Derivative 1c



1176

C	-9.579867	6.600609	-16.733153
C	-9.701898	6.385694	-15.371038
N	-8.809580	5.701270	-14.619227
C	-7.712704	5.160995	-15.190304
C	-7.490744	5.380670	-16.568354
C	-8.441286	6.083981	-17.331637
C	-6.326946	4.920045	-17.170084
C	-5.358999	4.297761	-16.397331
C	-6.764486	4.421618	-14.415412
C	-6.972578	4.055952	-13.032230
C	-5.858451	3.884771	-12.203055
C	-4.653845	3.395821	-12.867447
C	-8.242588	3.756695	-12.441627
C	-3.588284	2.711357	-12.207976
C	-2.485672	2.182905	-12.887941
C	-2.461575	2.377078	-14.248799
N	-3.431479	3.024915	-14.925919
C	-4.518644	3.484280	-14.269410
C	-9.345697	3.267702	-13.198076
C	-10.555903	2.914258	-12.617512
C	-10.629645	3.060965	-11.251130
N	-9.600319	3.472428	-10.485915
C	-8.422816	3.824523	-11.051659

C	-5.536978	4.075598	-15.016769
C	-7.356837	4.218765	-10.229877
C	-7.567245	4.546836	-8.833828
C	-6.518355	4.323577	-7.949085
C	-5.201194	4.541885	-8.509086
C	-8.744121	5.155363	-8.268771
C	-4.061615	4.900226	-7.724403
C	-2.864428	5.325502	-8.296323
C	-2.830540	5.352880	-9.670118
N	-3.851591	4.957389	-10.457701
C	-5.011244	4.544763	-9.912814
C	-9.651669	5.955000	-9.028948
C	-10.734915	6.606783	-8.452872
C	-10.867248	6.460632	-7.089351
N	-10.033321	5.729713	-6.325437
C	-8.999147	5.067569	-6.887443
C	-8.152920	4.307985	-6.063686
C	-12.185681	2.993370	-4.604402
C	-12.353469	3.093985	-3.240468
C	-6.875923	3.995273	-6.569440
C	-6.054576	4.173896	-10.777698
C	-9.142365	4.105527	-1.931211
C	-7.109199	4.457979	-0.077142
C	-6.072897	4.926207	0.780059
C	-4.871823	5.408182	0.275261
C	-4.695335	5.282848	-1.082206
C	-6.818992	4.406661	-1.458506
C	-7.804188	4.034664	-2.382088
C	-8.519770	3.944518	-4.704480
C	-7.499051	3.740407	-3.780539
C	-6.283770	3.181631	-4.335584
C	-9.851323	3.693133	-4.223732
C	-5.331816	2.426347	-3.582285
C	-4.271945	1.745264	-4.182461
C	-4.200242	1.839122	-5.550319
N	-5.038512	2.576374	-6.301513
C	-6.039963	3.256086	-5.726587
C	-10.143688	3.748000	-2.846823
C	-10.913502	3.279748	-5.080053
N	-11.368347	3.420607	-2.380748
N	-5.620488	4.789009	-1.926842
H	-10.547685	6.762065	-14.802509
H	-8.256296	6.224759	-18.392212

H	-6.139817	5.071278	-18.230841
H	-1.655933	2.024949	-14.881117
H	-11.527714	2.824613	-10.688950
H	-1.959072	5.673908	-10.224766
H	-11.668345	6.935785	-6.531061
H	-13.306973	2.896628	-2.760415
H	-6.262122	5.033195	1.844633
H	-3.779604	5.562980	-1.575878
H	-3.450227	1.336470	-6.138487
H	-9.227509	3.154219	-14.269634
H	-9.457808	6.097448	-10.088520
H	-10.715300	3.176832	-6.142392
C	-10.645269	7.632782	-18.748462
O	-9.807993	7.362638	-19.601252
N	-10.592444	7.320120	-17.402941
C	-11.731204	2.045095	-14.644970
O	-10.793776	1.992859	-15.432648
N	-11.678023	2.421463	-13.316441
C	-11.773476	7.603637	-10.502086
O	-11.149458	7.023835	-11.381798
N	-11.660818	7.422218	-9.137479
C	-13.292995	2.337660	-6.751862
O	-12.322084	2.257783	-7.494859
N	-13.289151	2.618060	-5.399218
C	-3.008885	6.971359	0.737011
O	-2.818223	7.342430	-0.421074
N	-3.949080	6.038156	1.136307
C	-2.401099	0.134993	-3.864965
O	-2.177509	-0.118412	-5.048334
N	-3.366554	1.003454	-3.394031
C	-0.696619	6.462709	-7.890812
O	-0.601954	7.026306	-8.979519
N	-1.777087	5.702855	-7.480233
C	-0.459440	0.740602	-12.701511
O	-0.149778	0.664236	-13.890577
N	-1.500143	1.477737	-12.167822
C	-11.885494	8.412907	-19.107212
H	-11.708408	9.476404	-18.927399
H	-12.744421	8.078103	-18.518523
H	-12.118963	8.256144	-20.164354
C	-13.115344	1.620537	-15.067596
H	-13.883382	2.171851	-14.517496
H	-13.234763	0.549087	-14.887290

H	-13.247648	1.829683	-16.133051
C	-12.841659	8.608347	-10.857274
H	-12.594096	9.076480	-11.814431
H	-12.907736	9.397655	-10.102856
H	-13.804797	8.097892	-10.938808
C	-14.681906	2.064356	-7.273711
H	-14.727260	2.325456	-8.335010
H	-15.427004	2.667070	-6.746249
H	-14.913110	1.003501	-7.147986
C	-2.178941	7.511413	1.871636
H	-2.819784	7.817782	2.701480
H	-1.623990	8.389786	1.529223
H	-1.467784	6.750648	2.202600
C	-1.547918	-0.472355	-2.784005
H	-2.168940	-0.954423	-2.029826
H	-0.920305	0.298214	-2.331032
H	-0.894737	-1.231983	-3.218912
C	0.394834	6.578525	-6.859252
H	1.305069	6.955698	-7.334716
H	0.620768	5.598933	-6.430152
H	0.089369	7.276344	-6.076100
C	0.329306	-0.001675	-11.649388
H	0.962431	-0.752355	-12.130591
H	-0.340741	-0.516279	-10.959309
H	0.967879	0.696362	-11.103169
C	-12.586802	6.538225	-0.282109
C	-12.735063	6.453208	1.084805
N	-11.927637	5.733190	1.887874
C	-10.909697	5.018299	1.368812
C	-10.630105	5.055106	-0.016465
C	-11.510350	5.842294	-0.817913
C	-9.440716	4.432572	-0.540851
C	-8.438982	4.217705	0.396742
C	-10.100224	4.262889	2.240738
C	-10.469060	3.959669	3.626309
C	-9.460716	3.626726	4.530264
C	-8.359652	2.927356	3.942352
C	-11.794390	3.868261	4.165241
C	-7.608672	1.959807	4.656536
C	-6.518054	1.301795	4.104980
C	-6.246156	1.591441	2.792404
N	-6.989386	2.431631	2.040989
C	-8.032105	3.090103	2.582626

C	-12.928485	3.528093	3.371534
C	-14.201613	3.389796	3.908024
C	-14.299683	3.566788	5.270190
N	-13.248691	3.825608	6.072702
C	-12.017008	4.002472	5.549446
C	-8.831193	3.892622	1.754852
C	-10.944027	4.251343	6.422548
C	-11.166832	4.639842	7.807345
C	-10.201355	4.287302	8.742249
C	-8.852691	4.234632	8.250873
C	-12.263621	5.417915	8.322576
C	-7.731398	4.420370	9.110003
C	-6.424848	4.236025	8.691008
C	-6.270150	4.002866	7.346851
N	-7.290656	4.002683	6.459610
C	-8.577279	4.063748	6.878821
C	-13.024714	6.307360	7.506261
C	-14.022639	7.124308	8.019481
C	-14.211351	7.052712	9.381487
N	-13.510115	6.244879	10.201809
C	-12.565661	5.417794	9.702527
C	-11.848720	4.585112	10.585910
C	-16.042003	4.161524	12.214969
C	-16.142724	4.388923	13.572150
C	-10.630398	4.052058	10.115641
C	-9.645829	3.951146	5.949566
C	-12.737735	4.682405	14.735379
C	-12.944157	4.983835	16.095783
C	-11.885341	5.205367	16.964207
C	-10.582484	5.175574	16.480792
C	-9.509984	5.502931	17.332524
C	-8.212206	5.556618	16.845130
C	-8.055762	5.264469	15.501783
C	-10.334581	4.848045	15.129790
C	-11.417733	4.513606	14.258299
C	-12.251578	4.366467	11.970078
C	-11.239056	4.080954	12.889083
C	-10.149514	3.298037	12.381753
C	-13.598749	4.380814	12.498117
C	-9.323243	2.483025	13.207541
C	-8.290934	1.701419	12.700620
C	-8.192456	1.660470	11.334487
N	-9.000720	2.355678	10.502554

C	-9.909779	3.224368	10.998657
C	-13.833397	4.531655	13.879715
C	-14.753477	4.153754	11.694359
N	-15.086869	4.558915	14.392760
N	-9.066871	4.914526	14.672076
H	-13.528296	6.972515	1.614526
H	-11.296929	5.950316	-1.877370
H	-5.427450	1.143655	2.252206
H	-15.245416	3.486364	5.797720
H	-5.295927	3.809733	6.907669
H	-14.953130	7.659524	9.891161
H	-17.103418	4.427688	14.076668
H	-12.094281	5.452273	18.003093
H	-9.723886	5.742395	18.369981
H	-7.084143	5.296432	15.015272
H	-7.472441	1.035576	10.825739
H	-12.784812	3.351781	2.310191
H	-12.781109	6.384879	6.450754
H	-14.614784	3.951704	10.636785
C	-13.583872	7.467128	-2.384043
O	-12.956744	6.843183	-3.231190
N	-13.487732	7.343121	-1.011442
C	-15.458659	2.754148	1.838172
O	-14.533081	2.594549	1.051518
N	-15.368307	3.080166	3.177620
C	-14.847526	8.170468	5.898681
O	-14.261763	7.489293	5.066026
N	-14.807747	8.027952	7.272431
C	-17.335316	3.640095	10.132664
O	-16.422017	3.424977	9.344999
N	-17.225571	3.957733	11.473257
C	-7.031089	6.193894	18.956361
O	-7.979961	6.212586	19.731875
N	-7.073665	5.895667	17.606655
C	-6.290396	0.330076	13.276076
O	-5.751778	0.270703	12.172098
N	-7.457733	0.999896	13.593533
C	-5.257915	4.566670	10.876632
O	-6.177538	5.045040	11.533453
N	-5.294476	4.221607	9.537735
C	-4.642334	-0.264640	4.562157
O	-4.112550	-0.264546	3.454049
N	-5.785466	0.417629	4.917354

C	-14.636591	8.467820	-2.793315
H	-14.382961	8.879575	-3.774503
H	-14.689205	9.297499	-2.082409
H	-15.607768	7.968991	-2.847262
C	-16.885985	2.544633	1.396522
H	-16.969436	2.760655	0.327468
H	-17.566090	3.214780	1.930278
H	-17.173179	1.506526	1.582758
C	-15.775381	9.279373	5.467195
H	-15.763745	10.104590	6.185120
H	-16.791250	8.885876	5.377364
H	-15.447969	9.669403	4.499126
C	-18.773342	3.528517	9.689395
H	-19.402946	4.256181	10.209608
H	-19.136567	2.517574	9.891646
H	-18.838467	3.731719	8.616561
C	-5.640273	6.541190	19.426034
H	-5.454805	7.604130	19.251257
H	-4.887726	5.945541	18.901384
H	-5.555070	6.327443	20.495395
C	-5.642067	-0.300648	14.479786
H	-4.964224	0.420333	14.943919
H	-5.075766	-1.179681	14.161209
H	-6.393276	-0.614112	15.208774
C	-3.917203	4.338051	11.525158
H	-3.167436	3.993264	10.813032
H	-4.029186	3.590667	12.314762
H	-3.567122	5.279443	11.957831
C	-4.000343	-0.983792	5.716105
H	-4.756563	-1.497646	6.310349
H	-3.462952	-0.263579	6.338270
H	-3.288168	-1.721903	5.339869
H	-18.090992	4.022536	11.995884
H	-15.444134	8.606643	7.806854
H	-16.229993	3.070083	3.709463
H	-14.170194	7.848973	-0.460444
H	-14.169479	2.519426	-4.909097
H	-12.344154	7.892832	-8.557653
H	-12.511558	2.270994	-12.762260
H	-11.360133	7.642001	-16.826794
H	-1.557706	1.505925	-11.152882
H	-1.797152	5.429662	-6.494317
H	-3.385662	1.168790	-2.390042

H	-6.160749	0.259192	5.846143
H	-4.449772	3.813823	9.136577
H	-7.699117	1.069807	14.575265
H	-6.194330	5.931373	17.103869
H	-7.894105	1.711193	5.677111
H	-7.913998	4.667715	10.150914
H	-9.498366	2.462516	14.279824
H	-4.141804	4.936447	-6.641041
H	-5.470068	2.303977	-2.510711
H	-3.659784	2.528733	-11.138219
H	-4.435889	3.985380	-16.888155
H	-13.956696	5.073835	16.490352
C	4.748807	4.948620	-19.114755
C	4.380656	4.810014	-17.787908
N	5.174378	4.289338	-16.826128
C	6.401653	3.826833	-17.134388
C	6.874644	3.980129	-18.455763
C	6.030853	4.530123	-19.440923
C	8.172924	3.598159	-18.777795
C	9.014952	3.117354	-17.784253
C	7.230769	3.233749	-16.136349
C	6.773291	2.938585	-14.799736
C	7.709312	2.943045	-13.762548
C	9.042084	2.510652	-14.124245
C	5.449827	2.528136	-14.433574
C	9.956984	1.950043	-13.188489
C	11.197930	1.451253	-13.565130
C	11.507870	1.539414	-14.905534
N	10.693298	2.072622	-15.838203
C	9.464460	2.507447	-15.470720
C	4.527236	1.918588	-15.341878
C	3.284167	1.457965	-14.935356
C	3.021778	1.577780	-13.595208
N	3.869809	2.089600	-12.691624
C	5.040929	2.609131	-13.089993
C	8.580463	2.964534	-16.452759
C	5.861664	3.196742	-12.119353
C	5.357600	3.549129	-10.797032
C	6.269302	3.565660	-9.755301
C	7.609161	3.961968	-10.109118
C	4.007680	3.903869	-10.444130
C	8.515511	4.538468	-9.169440
C	9.792316	4.946742	-9.524681

C	10.111800	4.821943	-10.859273
N	9.273372	4.345747	-11.799783
C	8.059678	3.873122	-11.444880
C	3.085200	4.420228	-11.402682
C	1.765390	4.688252	-11.091992
C	1.433188	4.529975	-9.769013
N	2.301099	4.191374	-8.792352
C	3.556160	3.822828	-9.103699
C	4.413933	3.364519	-8.079587
C	0.387222	1.819355	-6.427258
C	0.256596	2.020546	-5.076561
C	5.780436	3.237173	-8.428267
C	7.231053	3.310690	-12.432716
C	3.320321	3.379537	-3.887061
C	5.325838	3.963016	-2.056286
C	6.354948	4.483258	-1.213666
C	7.565767	4.942565	-1.716438
C	7.699768	4.909832	-3.087931
C	5.592299	3.967321	-3.438993
C	4.647001	3.458979	-4.342206
C	3.987050	3.105571	-6.691479
C	4.990502	3.107266	-5.712113
C	6.277051	2.617980	-6.137266
C	2.659621	2.779337	-6.191931
C	7.208154	1.966987	-5.264701
C	8.359232	1.337784	-5.724740
C	8.576540	1.404474	-7.081147
N	7.751116	2.031710	-7.938249
C	6.628014	2.636767	-7.493889
C	2.379109	2.902229	-4.804163
C	1.603547	2.200551	-6.981190
N	1.218311	2.498449	-4.281628
N	6.744180	4.468913	-3.925892
H	3.403527	5.116694	-17.424325
H	6.402894	4.619583	-20.456730
H	8.553065	3.697353	-19.791921
H	12.448923	1.168326	-15.298377
H	2.090752	1.247084	-13.170138
H	11.083357	5.110529	-11.246382
H	0.413948	4.655597	-9.428084
H	-0.669369	1.828635	-4.556701
H	6.175295	4.540805	-0.144856
H	8.598687	5.249535	-3.594014

H	9.434930	0.939250	-7.556437
H	4.809426	1.801224	-16.381322
H	3.425030	4.590638	-12.419561
H	1.779791	1.973917	-8.029861
C	4.002904	5.744358	-21.367151
O	5.035480	5.590552	-22.009556
N	3.815788	5.493040	-20.021914
C	2.385195	0.553893	-17.083669
O	3.381844	0.645111	-17.793581
N	2.290065	0.879793	-15.748716
C	0.878901	5.291633	-13.357526
O	1.903192	5.236733	-14.029083
N	0.762022	5.083405	-11.999210
C	-1.032773	1.232669	-8.407739
O	-0.470803	1.858678	-9.301029
N	-0.691370	1.194895	-7.078158
C	8.703406	5.598247	0.419963
O	7.812197	5.407942	1.238384
N	8.641230	5.437530	-0.949455
C	9.262510	0.464882	-3.558420
O	8.575857	1.084726	-2.753568
N	9.253360	0.588837	-4.932614
C	10.637454	5.720951	-7.294876
O	9.629830	5.623844	-6.605898
N	10.764808	5.468165	-8.645865
C	12.089962	0.853368	-11.314527
O	11.445358	1.622704	-10.609556
N	12.101731	0.810828	-12.693841
C	2.761889	6.298340	-22.022163
H	2.722508	7.379654	-21.867015
H	1.860924	5.834624	-21.610019
H	2.791410	6.085070	-23.094584
C	1.094939	-0.008496	-17.627003
H	1.038328	0.188257	-18.701393
H	0.228674	0.461075	-17.151495
H	1.068240	-1.087005	-17.450734
C	-0.437401	5.651386	-14.004637
H	-0.460890	6.728295	-14.190828
H	-1.290767	5.379660	-13.377111
H	-0.528935	5.111766	-14.951592
C	-2.302419	0.466122	-8.700732
H	-3.121366	1.176305	-8.843936
H	-2.564028	-0.212358	-7.887384

H	-2.173510	-0.130426	-9.605563
C	10.051457	6.108480	0.866505
H	10.080204	7.196140	0.761251
H	10.855483	5.659639	0.276023
H	10.212129	5.840269	1.914806
C	10.303973	-0.513278	-3.077057
H	9.960139	-0.967189	-2.144023
H	11.243431	0.016719	-2.900606
H	10.468155	-1.310257	-3.806504
C	11.930169	6.209755	-6.689474
H	11.919697	6.023826	-5.611509
H	12.033139	7.282627	-6.871865
H	12.786552	5.679181	-7.115714
C	13.058134	-0.119078	-10.692909
H	13.332062	-0.915499	-11.389212
H	12.588098	-0.572964	-9.816437
H	13.961490	0.416747	-10.390411
C	-0.405452	5.253743	-1.991428
C	-0.511447	4.997278	-0.646509
N	0.401312	4.346403	0.080964
C	1.497747	3.864488	-0.512652
C	1.733493	4.038103	-1.901146
C	0.719343	4.742636	-2.625706
C	2.997918	3.637613	-2.478608
C	4.041936	3.539073	-1.565385
C	2.432579	3.246787	0.318634
C	2.141338	2.891341	1.709702
C	3.231605	2.841983	2.572220
C	4.468360	2.381885	1.986935
C	0.867679	2.524141	2.309134
C	5.496881	1.725047	2.728667
C	6.629386	1.188690	2.126419
C	6.696827	1.308524	0.755693
N	5.755051	1.917763	0.011375
C	4.675855	2.471499	0.599490
C	-0.244462	1.992200	1.573427
C	-1.379474	1.497378	2.206905
C	-1.373578	1.548104	3.583344
N	-0.371329	2.044190	4.326093
C	0.713244	2.555519	3.719034
C	3.733005	3.116669	-0.206883
C	1.724704	3.054272	4.547406
C	1.525152	3.314893	5.975066

C	2.658475	3.250915	6.780216
C	3.882820	3.699364	6.172774
C	0.310796	3.704426	6.670538
C	4.966865	4.260490	6.913976
C	6.121701	4.733394	6.302928
C	6.143582	4.668408	4.926883
N	5.128532	4.191727	4.183404
C	4.028801	3.684969	4.775425
C	-0.810334	4.342320	6.037011
C	-1.873957	4.863493	6.767310
C	-1.774293	4.744947	8.136400
N	-0.784803	4.104232	8.781731
C	0.232508	3.570751	8.081371
C	1.250625	2.927844	8.803918
C	-2.407223	1.126282	10.955064
C	-2.245701	0.989754	12.315193
C	2.499199	2.819878	8.154191
C	3.008220	3.162966	3.970641
C	1.012099	2.306172	13.028621
C	0.965231	2.343605	14.438008
C	2.032977	2.805628	15.192476
C	3.165918	3.289715	14.549590
C	4.220581	3.838285	15.301149
C	5.337078	4.360185	14.665926
C	5.327478	4.319145	13.281467
C	3.254678	3.247055	13.140971
C	2.204509	2.661309	12.372501
C	1.104220	2.496255	10.196368
C	2.282969	2.448542	10.949802
C	3.473701	2.074542	10.245963
C	-0.105980	2.051397	10.889832
C	4.587281	1.455615	10.883088
C	5.699161	1.000741	10.186895
C	5.632385	1.122127	8.817189
N	4.589307	1.680937	8.165164
C	3.545368	2.193947	8.851186
C	-0.103059	1.901098	12.296984
C	-1.318965	1.642403	10.247141
N	-1.149402	1.389648	12.980125
N	4.319279	3.809731	12.536730
H	-1.375078	5.292138	-0.074804
H	0.871874	4.966907	-3.678214
H	7.525693	0.911503	0.176260

H	-2.210025	1.190078	4.170321
H	6.990371	5.012183	4.339859
H	-2.519837	5.171649	8.797726
H	-3.018454	0.558035	12.943879
H	1.952775	2.824528	16.276602
H	4.145512	3.840948	16.383691
H	6.151024	4.706809	12.688298
H	6.427914	0.776112	8.162889
H	-0.169383	1.895678	0.493323
H	-0.795685	4.487756	4.959714
H	-1.386264	1.688764	9.163994
C	-1.777350	6.101812	-3.887791
O	-1.396184	5.346811	-4.777192
N	-1.385238	6.078224	-2.573030
C	-2.880587	0.931589	0.275445
O	-2.351643	1.560365	-0.635521
N	-2.496436	0.898350	1.593721
C	-3.440698	5.638242	4.958519
O	-2.873153	5.234863	3.950319
N	-3.019911	5.516655	6.263688
C	-4.162814	1.105809	9.186300
O	-3.697510	1.955374	8.429869
N	-3.637546	0.724895	10.398048
C	6.610838	5.108895	16.684587
O	5.781243	4.890116	17.560568
N	6.453106	4.917619	15.325109
C	7.119787	0.340244	12.137019
O	6.423909	0.739534	13.064011
N	6.839391	0.420782	10.782812
C	7.401015	5.475198	8.325729
O	6.571932	5.292306	9.208665
N	7.242851	5.270230	6.969688
C	7.824934	0.340283	4.158511
O	7.096342	0.755162	5.052643
N	7.669602	0.514774	2.797201
C	-2.831940	7.144057	-4.169251
H	-2.612271	7.627416	-5.125243
H	-2.853503	7.918039	-3.397087
H	-3.810147	6.658293	-4.218524
C	-4.133395	0.126895	0.011263
H	-4.972696	0.811293	-0.137217
H	-4.361288	-0.548444	0.836338
H	-4.004829	-0.485391	-0.882443

C	-4.771935	6.347374	4.838765
H	-5.166784	6.665243	5.807266
H	-5.494168	5.667880	4.377527
C	-5.467421	0.420779	8.868839
H	-5.404532	-0.032002	7.882751
H	-6.272655	1.157071	8.873427
H	-5.700155	-0.371504	9.582803
C	7.958784	5.691170	17.030130
H	7.905135	6.781666	16.977424
H	8.732318	5.328540	16.347146
H	8.234598	5.387767	18.044244
C	8.461398	-0.289734	12.398580
H	8.685752	-0.242531	13.467499
H	9.237687	0.257579	11.857486
H	8.449029	-1.333793	12.083689
C	8.765811	6.024584	8.660067
H	9.534643	5.578322	8.022646
H	9.008449	5.785807	9.699668
H	8.761974	7.109625	8.527612
C	9.081339	-0.426816	4.479812
H	9.953804	0.103237	4.087551
H	9.030600	-1.426731	4.044622
H	9.183958	-0.510671	5.564471
H	-4.308783	0.303933	11.049845
H	-3.644183	5.895738	6.965707
H	-3.146320	0.437872	2.236620
H	-1.956598	6.611643	-1.913232
H	-1.357678	0.736233	-6.452472
H	-0.154384	5.230932	-11.594483
H	1.407373	0.667290	-15.276806
H	2.908252	5.721690	-19.633264
H	12.843235	0.280542	-13.133451
H	11.666621	5.657587	-9.063837
H	9.950321	0.059277	-5.443300
H	9.463947	5.696240	-1.479728
H	8.383392	0.106727	2.204902
H	8.015651	5.536424	6.371852
H	7.553367	0.081129	10.148983
H	7.214335	5.212643	14.725794
H	5.365417	1.592938	3.797812
H	4.875637	4.337365	7.992774
H	4.547343	1.286400	11.953794
H	8.190309	4.668618	-8.142091

H	6.958049	1.871060	-4.212523
H	9.641920	1.826536	-12.155150
H	10.038233	2.858609	-18.058136
H	0.066941	2.023957	14.965738
C	-5.294311	-0.397849	-14.620396
C	-5.360914	-0.690564	-13.268505
N	-4.440394	-1.421331	-12.593232
C	-3.370799	-1.927074	-13.244265
C	-3.187391	-1.598734	-14.605268
C	-4.166142	-0.852171	-15.285332
C	-2.034164	-1.993946	-15.271969
C	-1.038275	-2.663677	-14.578227
C	-2.404479	-2.732130	-12.565166
C	-2.583146	-3.224811	-11.217758
C	-1.450469	-3.456540	-10.433361
C	-0.254450	-3.860995	-11.158633
C	-3.833791	-3.597430	-10.626119
C	0.824246	-4.582871	-10.570767
C	1.951987	-4.987898	-11.291019
C	1.980217	-4.640719	-12.621466
N	0.969953	-3.994520	-13.241901
C	-0.136737	-3.640087	-12.546421
C	-4.957416	-4.008986	-11.398101
C	-6.142598	-4.440635	-10.818297
C	-6.160972	-4.471057	-9.440795
N	-5.111056	-4.135153	-8.666791
C	-3.969470	-3.685311	-9.229538
C	-1.184626	-3.009778	-13.220791
C	-2.896109	-3.330402	-8.400376
C	-3.074588	-3.103343	-6.973550
C	-1.984606	-3.296721	-6.133633
C	-0.717137	-2.963511	-6.732641
C	-4.246579	-2.570213	-6.331068
C	0.389245	-2.511435	-5.966142
C	1.558200	-2.024790	-6.543427
C	1.599042	-2.017128	-7.916235
N	0.583916	-2.458301	-8.690493
C	-0.555446	-2.919747	-8.137127
C	-5.203726	-1.763377	-7.012869
C	-6.273016	-1.160047	-6.363840
C	-6.343976	-1.367093	-5.003576
N	-5.462480	-2.111980	-4.307244
C	-4.438876	-2.721301	-4.944340

C	-3.527654	-3.471919	-4.185699
C	-7.433529	-5.070472	-2.684733
C	-7.563264	-5.050056	-1.313368
C	-2.253990	-3.688033	-4.745712
C	-1.611714	-3.285672	-8.987767
C	-4.388162	-3.894266	-0.039981
C	-2.369488	-3.514094	1.815608
C	-1.356891	-3.045475	2.693777
C	-0.161924	-2.513657	2.228867
C	0.023505	-2.538645	0.866311
C	-2.079832	-3.474246	0.430368
C	-3.055964	-3.854185	-0.511010
C	-3.828005	-3.897657	-2.829336
C	-2.770598	-4.063812	-1.935308
C	-1.534819	-4.503420	-2.544986
C	-5.131503	-4.246181	-2.331849
C	-0.519261	-5.207321	-1.832824
C	0.599251	-5.756238	-2.458699
C	0.670568	-5.580843	-3.819098
N	-0.251112	-4.912618	-4.539591
C	-1.324947	-4.366623	-3.940041
C	-5.388299	-4.271871	-0.948645
C	-6.196668	-4.678364	-3.179374
N	-6.579496	-4.688137	-0.467545
N	-0.892622	-3.011216	-0.009636
H	-6.187918	-0.354061	-12.649386
H	-4.011515	-0.622169	-16.334731
H	-1.876544	-1.743760	-16.318200
H	2.805647	-4.887125	-13.278631
H	-7.033235	-4.790753	-8.878966
H	2.448723	-1.654998	-8.482718
H	-7.127693	-0.928593	-4.393175
H	-8.483302	-5.341143	-0.815702
H	-1.549406	-2.994571	3.761969
H	0.936042	-2.191592	0.398181
H	1.479800	-5.961270	-4.424974
H	-4.870267	-4.009766	-12.479841
H	-5.062449	-1.571993	-8.071541
H	-6.021189	-4.729101	-4.249926
C	-6.498095	0.607764	-16.568349
O	-5.751495	0.272065	-17.481875
N	-6.345776	0.326867	-15.222140
C	-7.431119	-4.993803	-12.891200

O	-6.561249	-4.855044	-13.742145
N	-7.293136	-4.852162	-11.523703
C	-7.370674	-0.055357	-8.324895
O	-6.627197	-0.403081	-9.234079
N	-7.254242	-0.357602	-6.981312
C	-8.603559	-5.606529	-4.835613
O	-7.724273	-5.356245	-5.649101
N	-8.532923	-5.496143	-3.460077
C	1.750633	-1.070456	2.842694
O	2.015418	-0.643288	1.720238
N	0.743034	-1.960430	3.154247
C	2.613142	-7.192687	-2.185767
O	2.924494	-7.289115	-3.371040
N	1.546503	-6.468803	-1.691081
C	3.738246	-0.921136	-6.095913
O	4.103802	-0.714120	-7.251171
N	2.578660	-1.552370	-5.699501
C	4.104197	-6.250933	-11.193586
O	4.580662	-5.958307	-12.287620
N	2.936888	-5.760797	-10.639381
C	-7.771907	1.358987	-16.860113
H	-7.634695	1.972724	-17.754666
H	-8.036901	2.019554	-16.032811
H	-8.580618	0.643459	-17.029277
C	-8.825925	-5.412310	-13.284364
H	-9.571992	-4.985567	-12.607906
H	-8.894352	-6.503088	-13.264151
H	-9.040459	-5.051380	-14.294422
C	-8.601743	0.765666	-8.606050
H	-8.639801	1.000348	-9.672315
H	-8.574778	1.695201	-8.034046
H	-9.496330	0.194993	-8.341845
C	-9.962488	-6.073005	-5.295816
H	-9.882824	-6.485869	-6.305693
H	-10.652899	-5.225521	-5.305048
H	-10.348147	-6.857274	-4.638101
C	2.539366	-0.639460	4.050328
H	1.884201	-0.142704	4.764855
H	3.313880	0.067510	3.753687
H	3.028063	-1.505209	4.502946
C	3.426568	-7.879676	-1.121208
H	2.776809	-8.439183	-0.445562
H	4.010974	-7.142195	-0.568148

H	4.113749	-8.589069	-1.592255
C	4.585847	-0.505826	-4.925974
H	5.453323	0.033773	-5.294785
H	4.947845	-1.390004	-4.399486
H	4.029341	0.150047	-4.256858
C	4.803424	-7.240835	-10.299353
H	5.511649	-7.827859	-10.891814
H	4.082723	-7.932334	-9.855243
H	5.348081	-6.708973	-9.519463
C	-7.855390	-1.606943	1.676622
C	-8.059909	-1.803701	3.024416
N	-7.256946	-2.551403	3.810863
C	-6.198883	-3.200708	3.278960
C	-5.897082	-3.087674	1.904884
C	-6.765340	-2.268203	1.129762
C	-4.689494	-3.647138	1.362859
C	-3.680069	-3.846561	2.291319
C	-5.352034	-3.936508	4.130836
C	-5.705660	-4.263203	5.509762
C	-4.672712	-4.474512	6.423089
C	-3.501857	-5.075372	5.864993
C	-7.025474	-4.448591	6.039563
C	-2.640862	-5.903383	6.631356
C	-1.517495	-6.510630	6.088053
C	-1.314356	-6.303634	4.745949
N	-2.125211	-5.561573	3.963350
C	-3.205767	-4.953114	4.491070
C	-8.131626	-4.857767	5.237819
C	-9.392424	-5.092673	5.770474
C	-9.503675	-4.938277	7.135058
N	-8.481881	-4.584830	7.939792
C	-7.263127	-4.325545	7.421715
C	-4.058247	-4.221577	3.647072
C	-6.217756	-3.961791	8.290795
C	-6.470739	-3.552650	9.663293
C	-5.468616	-3.740628	10.598999
C	-4.138647	-3.596127	10.091530
C	-7.622160	-2.862777	10.161650
C	-3.075174	-3.152235	10.923458
C	-1.770135	-3.034867	10.476899
C	-1.579288	-3.308246	9.141458
N	-2.578831	-3.624033	8.281535
C	-3.848629	-3.787175	8.725174

C	-8.421579	-2.049302	9.317236
C	-9.486086	-1.303368	9.794800
C	-9.727647	-1.397952	11.147236
N	-8.965387	-2.112613	12.003697
C	-7.934103	-2.853777	11.538482
C	-7.130784	-3.579210	12.443001
C	-11.264292	-4.337608	14.122792
C	-11.368589	-4.117452	15.480916
C	-5.854727	-3.976553	11.981386
C	-4.894683	-4.116257	7.823000
C	-7.997616	-3.494505	16.603932
C	-8.214610	-3.196210	17.964114
C	-7.172140	-2.852725	18.814149
C	-5.880148	-2.749731	18.310955
C	-4.832850	-2.291653	19.134972
C	-3.557246	-2.104045	18.620077
C	-3.395666	-2.396690	17.277999
C	-5.621617	-3.072455	16.962276
C	-6.674497	-3.538144	16.114027
C	-7.505992	-3.804250	13.834153
C	-6.464997	-3.971842	14.750816
C	-5.294391	-4.632284	14.254030
C	-8.845847	-3.902689	14.375928
C	-4.403932	-5.352401	15.097565
C	-3.279632	-6.010725	14.615254
C	-3.130829	-6.019463	13.251412
N	-3.982445	-5.404090	12.401351
C	-5.023901	-4.679082	12.875751
C	-9.080529	-3.759276	15.759596
C	-9.985277	-4.237716	13.588222
N	-10.324082	-3.840561	16.287581
N	-4.380133	-2.865840	16.475764
H	-8.882497	-1.333656	3.554666
H	-6.534662	-2.112995	0.080764
H	-0.498794	-6.757452	4.206097
H	-10.439290	-5.102873	7.661546
H	-0.591656	-3.276435	8.688728
H	-10.542593	-0.859481	11.620730
H	-12.322991	-4.166551	15.996481
H	-7.391577	-2.615178	19.853275
H	-5.051622	-2.061723	20.173819
H	-2.442953	-2.265764	16.770206
H	-2.327830	-6.549854	12.758153

H	-7.969355	-5.024579	4.176845
H	-8.174820	-1.988037	8.262735
H	-9.840111	-4.451658	12.533465
C	-8.596503	-0.318830	-0.337936
O	-7.688356	-0.576562	-1.118796
N	-8.723485	-0.748745	0.970435
C	-10.632579	-5.676987	3.673371
O	-9.755836	-5.509649	2.835098
N	-10.531598	-5.490299	5.038980
C	-10.058805	-0.075155	7.698326
O	-9.044210	-0.276029	7.038230
N	-10.300117	-0.473095	9.000325
C	-12.548432	-4.869361	12.036501
O	-11.663724	-4.783993	11.193949
N	-12.432431	-4.656501	13.397461
C	-2.411885	-1.302064	20.694293
O	-3.348813	-1.352838	21.482750
N	-2.446591	-1.633530	19.352160
C	-1.149488	-7.126208	15.253285
O	-0.552888	-7.029303	14.182596
N	-2.409202	-6.633854	15.532121
C	-0.696671	-2.327929	12.625297
O	-1.696834	-2.193226	13.324943
N	-0.666982	-2.691767	11.289415
C	0.530087	-7.830114	6.576245
O	1.124172	-7.628753	5.519989
N	-0.699417	-7.303193	6.917330
C	-9.761049	0.538031	-0.761841
H	-9.561000	0.955170	-1.751937
H	-9.906767	1.357642	-0.054335
H	-10.666959	-0.071963	-0.812548
C	-12.019522	-6.102061	3.258788
H	-11.967980	-6.621764	2.297634
H	-12.654566	-5.217817	3.159875
H	-12.454280	-6.788028	3.991495
C	-11.234189	0.665551	7.116159
H	-11.514115	1.499468	7.763791
H	-12.081378	-0.017656	7.009917
H	-10.969599	1.051119	6.128429
C	-13.963090	-5.204356	11.632796
H	-14.536231	-4.280289	11.520934
H	-14.440613	-5.846681	12.378359
H	-13.951301	-5.742208	10.680358

C	-1.052756	-0.816827	21.133506
H	-0.965159	0.252228	20.923207
H	-0.256190	-1.359785	20.616646
H	-0.936524	-0.987620	22.207635
C	-0.497460	-7.773663	16.448718
H	0.089459	-7.025620	16.987850
H	0.158859	-8.579669	16.108320
H	-1.244027	-8.205698	17.121211
C	0.669286	-2.044015	13.195254
H	1.452056	-2.557819	12.637384
H	0.706254	-2.393666	14.230802
H	0.856923	-0.968754	13.165435
C	1.143674	-8.674038	7.663623
H	0.384631	-9.054262	8.353022
H	1.868312	-8.072646	8.218629
H	1.648080	-9.531728	7.209638
H	-13.286373	-4.718385	13.938639
H	-11.069768	-0.014229	9.473567
H	-11.373248	-5.640242	5.581611
H	-9.481288	-0.341124	1.504024
H	-9.366183	-5.742443	-2.941158
H	-7.949351	0.047208	-6.366845
H	-8.091872	-5.097789	-10.953431
H	-7.092401	0.626423	-14.606600
H	2.770115	-5.964859	-9.661686
H	2.451979	-1.680815	-4.699783
H	1.456332	-6.405915	-0.683341
H	-1.014081	-7.422258	7.872043
H	0.252794	-2.840323	10.872254
H	-2.692460	-6.604559	16.505116
H	-1.582202	-1.520867	18.833842
H	-2.874388	-6.090357	7.676707
H	-3.298728	-2.906206	11.956042
H	-4.604458	-5.398871	16.164666
H	0.304364	-2.462276	-4.884435
H	-0.644460	-5.390645	-0.769083
H	0.755478	-4.863995	-9.522656
H	-0.118571	-2.906851	-15.110628
H	-9.225843	-3.207651	18.372752
C	9.233518	-1.281520	-16.129139
C	9.063636	-1.534911	-14.779341
N	9.988083	-2.139681	-13.994995
C	11.146668	-2.575438	-14.523517

C	11.444827	-2.252538	-15.862589
C	10.469319	-1.623656	-16.661428
C	12.699719	-2.550817	-16.382710
C	13.661756	-3.131568	-15.566227
C	12.081025	-3.309398	-13.730334
C	11.770559	-3.830498	-12.419424
C	12.806049	-4.000266	-11.494168
C	14.104082	-4.318157	-12.066507
C	10.475352	-4.274802	-12.002478
C	15.143242	-4.979303	-11.346789
C	16.374505	-5.286977	-11.915588
C	16.531619	-4.941264	-13.241210
N	15.580663	-4.347626	-13.986467
C	14.377976	-4.067767	-13.430801
C	9.475038	-4.720627	-12.916816
C	8.211535	-5.120268	-12.507610
C	8.013737	-5.148862	-11.147170
N	8.964019	-4.846775	-10.236615
C	10.159862	-4.364375	-10.639238
C	13.384962	-3.495626	-14.231483
C	11.099680	-3.953165	-9.678104
C	10.714596	-3.715103	-8.299803
C	11.674057	-3.877135	-7.305549
C	13.013070	-3.534951	-7.705690
C	9.443569	-3.219561	-7.852174
C	14.000266	-3.104663	-6.770121
C	15.265951	-2.690655	-7.156278
C	15.495865	-2.665558	-8.513042
N	14.581460	-3.002918	-9.443830
C	13.372533	-3.477426	-9.074866
C	8.545703	-2.493103	-8.695041
C	7.338992	-1.994414	-8.235768
C	7.110361	-2.163897	-6.894227
N	7.953956	-2.762434	-6.035185
C	9.069717	-3.341602	-6.497932
C	9.879899	-4.023425	-5.581205
C	5.766221	-5.393568	-4.220620
C	5.555796	-5.234764	-2.870943
C	11.215720	-4.257220	-5.969034
C	12.451741	-3.853621	-10.077224
C	8.659566	-4.112512	-1.477613
C	10.633863	-3.569978	0.399491
C	11.637748	-3.025209	1.256339

C	12.887654	-2.641094	0.789385
C	13.097079	-2.796822	-0.563240
C	10.962056	-3.666284	-0.964007
C	10.013801	-4.133007	-1.884732
C	9.418808	-4.356063	-4.235347
C	10.385833	-4.486290	-3.247813
C	11.646329	-5.023876	-3.691323
C	8.066858	-4.616358	-3.800913
C	12.523556	-5.746192	-2.828003
C	13.698574	-6.330432	-3.281206
C	13.954446	-6.197552	-4.629070
N	13.151947	-5.544712	-5.491271
C	12.026420	-4.944301	-5.048801
C	7.722072	-4.523294	-2.431435
C	7.039635	-5.082889	-4.677193
N	6.501984	-4.863919	-1.994200
N	12.167385	-3.261664	-1.417814
H	8.148994	-1.272627	-14.252918
H	10.700089	-1.410771	-17.700221
H	12.948045	-2.308332	-17.413557
H	17.453880	-5.137579	-13.779092
H	7.061894	-5.429359	-10.713211
H	16.449444	-2.356173	-8.928196
H	6.199854	-1.823276	-6.435520
H	4.584596	-5.390362	-2.419554
H	11.404380	-2.882859	2.306670
H	14.035722	-2.532658	-1.041224
H	14.838403	-6.623743	-5.093179
H	9.704609	-4.728405	-13.977073
H	8.818573	-2.307035	-9.729202
H	7.272963	-5.241061	-5.726342
C	8.127052	-0.453352	-18.215402
O	9.029141	-0.638282	-19.026820
N	8.167497	-0.712302	-16.857765
C	7.199134	-5.614943	-14.740127
O	8.200489	-5.602417	-15.448370
N	7.149538	-5.467044	-13.368383
C	6.352511	-1.139577	-10.367294
O	7.222576	-1.468413	-11.166058
N	6.356991	-1.339669	-9.002340
C	4.638293	-6.057984	-6.353942
O	5.498564	-5.765753	-7.175761
N	4.700525	-5.890822	-4.991981

C	13.888766	-1.769147	2.915152
O	12.923104	-1.813949	3.668061
N	13.928056	-2.103677	1.576047
C	14.583521	-7.249924	-1.123791
O	13.804233	-6.756193	-0.318091
N	14.608680	-7.065454	-2.492083
C	16.263378	-2.192299	-4.910157
O	15.302454	-2.381513	-4.174655
N	16.302073	-2.287605	-6.287340
C	17.550151	-6.225899	-9.913795
O	16.795398	-5.869898	-9.017339
N	17.439143	-5.942207	-11.260745
C	6.781990	0.061052	-18.664403
H	6.308894	0.660675	-17.884380
H	6.137343	-0.786076	-18.912620
H	6.910654	0.692134	-19.548057
C	5.834489	-5.854242	-15.335085
H	5.895274	-5.796110	-16.425688
H	5.131636	-5.090885	-14.989521
H	5.477468	-6.847462	-15.049920
C	5.083118	-0.474956	-10.830194
H	4.942272	0.473075	-10.305901
H	4.230154	-1.132821	-10.641176
H	5.148407	-0.285530	-11.904364
C	3.314884	-6.630103	-6.800424
H	2.624681	-5.812426	-7.024183
H	2.877836	-7.280349	-6.039041
H	3.466808	-7.237782	-7.694158
C	15.219517	-1.258751	3.410349
H	15.294602	-0.187566	3.205898
H	16.047384	-1.784455	2.925787
H	15.295084	-1.430498	4.488022
C	15.706060	-8.141014	-0.651493
H	15.413896	-8.626498	0.284227
H	16.601351	-7.536552	-0.484351
H	15.922667	-8.923732	-1.384276
C	17.591065	-1.763881	-4.335083
H	17.663137	-2.103046	-3.297550
H	17.664535	-0.673741	-4.367253
H	18.421694	-2.205731	-4.893039
C	18.797212	-7.014353	-9.597961
H	19.053464	-7.692475	-10.417142
H	18.626697	-7.617312	-8.701436

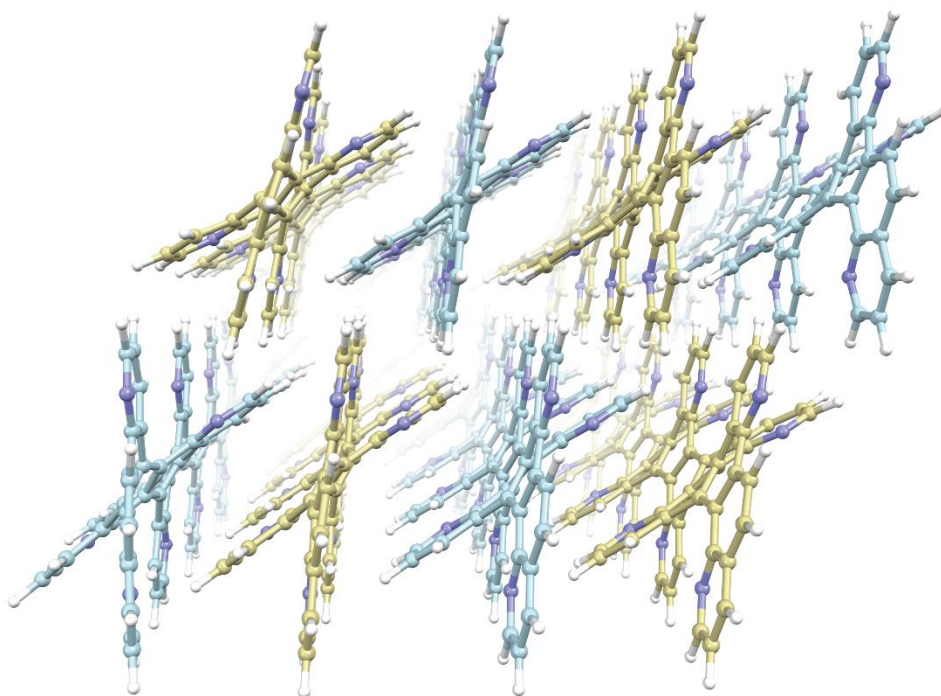
H	19.625992	-6.323683	-9.421864
C	4.933256	-1.978701	0.275099
C	4.945582	-1.978301	1.646755
N	5.822324	-2.635867	2.408128
C	6.809211	-3.315216	1.824886
C	7.035273	-3.279111	0.427301
C	6.010002	-2.612484	-0.336898
C	8.297498	-3.789402	-0.094204
C	9.314697	-3.900638	0.857114
C	7.660972	-3.981738	2.697721
C	7.289424	-4.254220	4.082080
C	8.325778	-4.330075	4.994848
C	9.558736	-4.887652	4.494637
C	5.974502	-4.511949	4.627087
C	10.509239	-5.553948	5.326770
C	11.650808	-6.157259	4.813092
C	11.802394	-6.093919	3.443861
N	10.932389	-5.485948	2.616647
C	9.837451	-4.874494	3.113212
C	4.901226	-5.034000	3.842508
C	3.704019	-5.443760	4.417480
C	3.596669	-5.276801	5.781852
N	4.570225	-4.777062	6.565632
C	5.733493	-4.381728	6.018731
C	8.958069	-4.231400	2.226950
C	6.733155	-3.889439	6.871556
C	6.494950	-3.510343	8.271329
C	7.616412	-3.521083	9.103494
C	8.868937	-3.156497	8.500983
C	5.269453	-3.048137	8.920137
C	9.940429	-2.537608	9.217505
C	11.129369	-2.158986	8.605556
C	11.208245	-2.394738	7.251870
N	10.212141	-2.937545	6.529994
C	9.066266	-3.333238	7.126860
C	4.132619	-2.477519	8.246742
C	3.097135	-1.851164	8.940480
C	3.234087	-1.790915	10.307268
N	4.217376	-2.390176	10.994859
C	5.197052	-3.034715	10.336481
C	6.186271	-3.649997	11.116006
C	2.464314	-5.248988	13.174244
C	2.511397	-5.225951	14.549406

C	7.434490	-3.841208	10.500535
C	8.049883	-3.891647	6.345952
C	5.791296	-3.963721	15.347319
C	5.683126	-3.804626	16.742485
C	6.744917	-3.349032	17.509164
C	7.939805	-3.000605	16.890171
C	8.989355	-2.457164	17.653521
C	10.167289	-2.056006	17.043233
C	10.227565	-2.218579	15.669475
C	8.095528	-3.169837	15.496154
C	7.036978	-3.735495	14.721855
C	6.004086	-3.981574	12.526344
C	7.155959	-4.040336	13.316675
C	8.343365	-4.506679	12.659477
C	4.762790	-4.339097	13.193447
C	9.410304	-5.158631	13.347017
C	10.510517	-5.693300	12.688443
C	10.497254	-5.584570	11.315954
N	9.503113	-5.000400	10.621249
C	8.449591	-4.448886	11.259680
C	4.682309	-4.371966	14.603007
C	3.599223	-4.794942	12.506358
N	3.565753	-4.780265	15.257462
N	9.234758	-2.739999	14.912352
H	4.219914	-1.428474	2.216823
H	6.108392	-2.521262	-1.415239
H	12.649536	-6.541088	2.931829
H	2.694251	-5.547964	6.320456
H	12.085991	-2.141142	6.664746
H	2.521387	-1.262012	10.929727
H	1.676380	-5.565596	15.152437
H	6.612488	-3.228920	18.581827
H	8.853043	-2.348544	18.724821
H	11.102033	-1.926406	15.095437
H	11.295611	-5.976005	10.691885
H	5.054402	-5.193016	2.778662
H	4.118922	-2.433406	7.160459
H	3.621606	-4.869159	11.422378
C	3.389626	-1.435812	-1.616915
O	3.775818	-2.233451	-2.464945
N	3.867761	-1.293878	-0.337916
C	2.470007	-6.277712	2.400535
O	3.164657	-5.839820	1.490959

N	2.637970	-6.079738	3.751147
C	1.350038	-1.514364	7.184286
O	1.624593	-2.421988	6.408091
N	1.968413	-1.220047	8.379484
C	0.942779	-5.557909	11.238209
O	1.350600	-4.655257	10.511514
N	1.327530	-5.785027	12.539033
C	11.393718	-1.239944	19.066296
O	10.543557	-1.427551	19.928909
N	11.276630	-1.497267	17.713221
C	11.801072	-6.546646	14.658298
O	11.068585	-6.201984	15.577425
N	11.597749	-6.346820	13.306240
C	12.330034	-1.177157	10.571297
O	11.450767	-1.226488	11.422579
N	12.232511	-1.560719	9.249370
C	12.709674	-6.966179	6.937029
O	11.991112	-6.439340	7.777545
N	12.626148	-6.846821	5.564067
C	2.206851	-0.533140	-1.886933
H	2.270680	-0.128570	-2.898024
H	2.170556	0.309510	-1.191307
H	1.288685	-1.116734	-1.789834
C	1.246514	-7.102705	2.087195
H	0.384128	-6.441342	1.972985
H	1.054860	-7.843836	2.866411
H	1.401700	-7.653517	1.158725
C	0.173856	-0.600617	6.928943
H	-0.534601	-0.654000	7.759962
H	-0.341102	-0.906473	6.014111
C	-0.128348	-6.502785	10.764836
H	0.099540	-6.799921	9.739977
H	-1.097137	-6.001787	10.791670
H	-0.169192	-7.407829	11.376724
C	12.729251	-0.634756	19.421043
H	12.679661	0.449679	19.292844
H	13.526206	-1.042829	18.792816
H	12.967364	-0.868112	20.462845
C	13.077596	-7.301626	14.934965
H	13.403080	-7.101891	15.960020
H	13.875233	-6.981735	14.258248
H	12.898502	-8.373090	14.813444
C	13.686965	-0.606933	10.901516

H	14.475964	-1.123983	10.347762
H	13.883878	-0.735603	11.969758
H	13.701669	0.457641	10.653797
C	13.874283	-7.827746	7.359737
H	14.779048	-7.216018	7.405294
H	14.024675	-8.655399	6.660451
H	13.671569	-8.252559	8.347162
H	0.655634	-6.279264	13.134440
H	1.468187	-0.586594	8.991672
H	1.932030	-6.491708	4.365274
H	3.247133	-0.795100	0.305797
H	3.893446	-6.215702	-4.456618
H	5.534616	-1.007123	-8.491758
H	6.235392	-5.611019	-12.932069
H	7.323000	-0.516277	-16.331090
H	18.232747	-6.183349	-11.841690
H	17.176443	-2.045740	-6.736419
H	15.386996	-7.475464	-2.993140
H	14.791363	-1.914194	1.081656
H	13.364379	-7.280355	5.022996
H	13.035864	-1.380304	8.659501
H	12.307166	-6.704967	12.678712
H	12.068777	-1.257661	17.130121
H	10.307021	-5.626849	6.391539
H	9.814141	-2.353054	10.278910
H	9.335196	-5.280263	14.422729
H	13.742268	-3.068816	-5.717004
H	12.237275	-5.880583	-1.789075
H	14.945637	-5.295436	-10.326236
H	14.649219	-3.314262	-15.989979
H	4.741737	-4.019355	17.247396
H	-4.028430	5.850342	2.131949
H	-4.650272	7.234861	4.212706
H	0.587430	-2.146679	4.141245
H	0.525031	0.426222	6.809979

4.3. Packing Structure Coordinates for fjord-edge N₂[8]GNR Parent System 1d



1456 atoms

C	0.699803	-6.627835	-18.144286
C	0.418636	-6.982801	-16.847168
N	1.328718	-7.511151	-16.001590
C	2.601660	-7.729961	-16.407692
C	2.964430	-7.352006	-17.721596
C	1.998807	-6.814131	-18.584827
C	4.275091	-7.487257	-18.160352
C	5.243733	-7.951114	-17.285500
C	3.586237	-8.286108	-15.526671
C	3.294254	-8.754940	-14.185122
C	4.317833	-8.751346	-13.233943
C	5.647805	-8.979418	-13.757766
C	2.058416	-9.322706	-13.740344
C	6.700364	-9.533233	-12.987785
C	7.973171	-9.739560	-13.504150

C	8.181923	-9.399262	-14.813561
N	7.208748	-8.919878	-15.609650
C	5.954937	-8.749489	-15.113417
C	1.162546	-9.979604	-14.621143
C	0.002137	-10.594607	-14.175810
C	-0.235748	-10.569206	-12.829419
N	0.599379	-9.987378	-11.949478
C	1.726757	-9.363296	-12.373967
C	4.933181	-8.320169	-15.963190
C	2.596153	-8.783604	-11.433277
C	2.177906	-8.529788	-10.064480
C	3.149901	-8.541421	-9.072981
C	4.442174	-8.081542	-9.489182
C	0.858704	-8.169324	-9.623254
C	5.365270	-7.525748	-8.571448
C	6.584094	-7.015515	-8.971413
C	6.851546	-7.017242	-10.308307
N	5.987980	-7.483535	-11.230080
C	4.811693	-8.042460	-10.853705
C	-0.096529	-7.587751	-10.496006
C	-1.341851	-7.166153	-10.057894
C	-1.609645	-7.306259	-8.725497
N	-0.730019	-7.828385	-7.850858
C	0.486396	-8.261366	-8.264003
C	1.403609	-8.774508	-7.319233
C	-2.407239	-10.577535	-5.750965
C	-2.628678	-10.409346	-4.412268
C	2.753035	-8.898458	-7.720941
C	3.934472	-8.557849	-11.833764
C	0.360214	-8.753507	-3.179439
C	2.315663	-7.945563	-1.371788
C	3.285817	-7.313078	-0.553967
C	4.483073	-6.837638	-1.062647
C	4.689506	-6.979892	-2.406475
C	2.624218	-8.044052	-2.742916
C	1.694623	-8.635813	-3.621906
C	1.038920	-9.058948	-5.936521
C	2.044575	-9.021131	-4.977215
C	3.324133	-9.477124	-5.429072
C	-0.243067	-9.461586	-5.435846
C	4.272065	-10.069852	-4.555897
C	5.472415	-10.591653	-5.012535
C	5.701137	-10.533865	-6.359592

N	4.827520	-9.995278	-7.231605
C	3.657175	-9.460720	-6.796429
C	-0.565772	-9.341736	-4.067258
C	-1.200524	-10.114065	-6.251427
N	-1.737157	-9.827984	-3.588655
N	3.791153	-7.550971	-3.230623
H	-0.573126	-6.850937	-16.426962
H	2.262039	-6.516812	-19.596602
H	4.565769	-7.201131	-19.167792
H	9.155014	-9.517137	-15.279924
H	-1.116698	-11.032703	-12.397464
H	7.790401	-6.622561	-10.680127
H	-2.554870	-6.986960	-8.300054
H	-3.543628	-10.750875	-3.939478
H	3.102263	-7.154352	0.506235
H	5.603020	-6.628006	-2.874536
H	6.609750	-10.934680	-6.796634
H	1.385830	-10.048130	-15.683517
H	0.138648	-7.415257	-11.543864
H	-0.994010	-10.300899	-7.303193
C	-3.467560	-6.967038	-1.680834
C	-3.684687	-7.097082	-0.337508
N	-2.774680	-7.614114	0.508923
C	-1.579738	-8.060037	0.049784
C	-1.266078	-7.989662	-1.323205
C	-2.244803	-7.403460	-2.164576
C	0.019126	-8.393917	-1.811858
C	1.040803	-8.368078	-0.868738
C	-0.628856	-8.578247	0.956123
C	-0.956016	-8.895492	2.338159
C	0.067961	-8.866245	3.278348
C	1.350792	-9.274086	2.785000
C	-2.220993	-9.340750	2.846561
C	2.329708	-9.869636	3.621622
C	3.537472	-10.340776	3.131191
C	3.742742	-10.232301	1.784379
N	2.836028	-9.699008	0.944042
C	1.656199	-9.215497	1.410786
C	-3.182607	-9.984533	2.027507
C	-4.369298	-10.490729	2.534605
C	-4.566095	-10.375680	3.882512
N	-3.670195	-9.803819	4.707825
C	-2.520505	-9.273173	4.222852

C	0.708596	-8.684481	0.510190
C	-1.594161	-8.688270	5.111057
C	-1.928321	-8.364912	6.489422
C	-0.902100	-8.351643	7.426171
C	0.371162	-7.916918	6.921171
C	-3.213221	-7.980793	6.992302
C	1.357965	-7.322677	7.747748
C	2.552132	-6.836684	7.241130
C	2.737179	-6.919544	5.889872
N	1.821645	-7.447488	5.057552
C	0.663818	-7.963549	5.542557
C	-4.204713	-7.404108	6.158629
C	-5.433346	-6.994452	6.647168
C	-5.637075	-7.127321	7.990620
N	-4.708528	-7.619210	8.831472
C	-3.515074	-8.059876	8.369138
C	-2.559674	-8.581246	9.273160
C	-6.429249	-10.172325	10.966673
C	-6.607868	-9.971571	12.308803
C	-1.228967	-8.690056	8.803388
C	-0.268496	-8.533619	4.653258
C	-3.407576	-8.752944	13.448653
C	-3.693440	-8.551021	14.812482
C	-2.727886	-8.122349	15.707427
C	-1.452869	-7.831044	15.245422
C	-0.494300	-7.340561	16.142288
C	0.759781	-6.980125	15.684316
C	1.001083	-7.109484	14.337487
C	-1.122162	-8.004543	13.879693
C	-2.087921	-8.557810	12.974737
C	-2.866818	-8.895876	10.669787
C	-1.809497	-8.883301	11.585782
C	-0.549048	-9.309817	11.057347
C	-4.153941	-9.269726	11.214974
C	0.434881	-9.941416	11.859817
C	1.616103	-10.440408	11.331940
C	1.793386	-10.317353	9.982963
N	0.880846	-9.749080	9.174587
C	-0.275198	-9.245086	9.677281
C	-4.424095	-9.185827	12.594931
C	-5.192311	-9.835103	10.433870
N	-5.636187	-9.509812	13.116138
N	0.098351	-7.596722	13.455491

H	-4.612655	-6.773758	0.120315
H	-2.046891	-7.249792	-3.223338
H	4.656584	-10.588747	1.320791
H	-5.463403	-10.753798	4.361137
H	3.644429	-6.550315	5.422747
H	-6.568976	-6.827558	8.454811
H	-7.551941	-10.195302	12.794552
H	-2.998279	-7.986870	16.751583
H	-0.724577	-7.215286	17.197603
H	1.958498	-6.818029	13.915244
H	2.685595	-10.690830	9.491646
H	-2.995686	-10.130709	0.965775
H	-4.016122	-7.240221	5.099815
H	-5.038916	-10.030463	9.374554
H	2.136415	-10.011904	4.682757
H	1.191947	-7.203228	8.815750
H	0.263932	-10.095447	12.922695
H	5.135261	-7.454163	-7.511542
H	4.058846	-10.168888	-3.493614
H	6.525191	-9.833693	-11.956883
H	6.271222	-8.012369	-17.645032
H	-4.700108	-8.729802	15.192468
H	8.769676	-10.150399	-12.897043
H	7.292791	-6.613232	-8.262510
H	6.186019	-11.045851	-4.336783
H	5.219107	-6.356515	-0.429678
H	4.274790	-10.794977	3.780721
H	3.301202	-6.387336	7.881121
H	2.355183	-10.927806	11.954013
H	1.517656	-6.579811	16.350006
H	-7.218192	-10.573556	10.343366
H	-6.187038	-6.566868	5.998075
H	-5.096375	-10.971688	1.893124
H	-4.208480	-6.523869	-2.334683
H	-3.131432	-11.066621	-6.389584
H	-2.058752	-6.721174	-10.736737
H	-0.674799	-11.087044	-14.862007
H	-0.059542	-6.197520	-18.785204
C	-2.978496	-1.469639	-18.730722
C	-3.270459	-1.828252	-17.437028
N	-2.379293	-2.403389	-16.601291
C	-1.110860	-2.650952	-17.007161
C	-0.737679	-2.271069	-18.318042

C	-1.687345	-1.696731	-19.175677
C	0.569090	-2.435524	-18.758779
C	1.527039	-2.926553	-17.887262
C	-0.139382	-3.234829	-16.129617
C	-0.440822	-3.696574	-14.786562
C	0.591052	-3.714099	-13.847994
C	1.911489	-3.980093	-14.370030
C	-1.688846	-4.222569	-14.316909
C	2.944341	-4.560847	-13.594118
C	4.210826	-4.803666	-14.109075
C	4.433375	-4.456875	-15.415523
N	3.478799	-3.942136	-16.214690
C	2.225031	-3.752307	-15.722794
C	-2.627040	-4.842197	-15.179416
C	-3.794233	-5.425162	-14.709357
C	-3.999292	-5.402205	-13.357104
N	-3.129358	-4.848288	-12.492535
C	-1.994972	-4.253266	-12.942126
C	1.208849	-3.296303	-16.567606
C	-1.097971	-3.680785	-12.021763
C	-1.476853	-3.386510	-10.646802
C	-0.481302	-3.358633	-9.678035
C	0.789055	-2.892683	-10.148841
C	-2.776794	-2.998247	-10.183017
C	1.714476	-2.257076	-9.288846
C	2.895314	-1.711283	-9.756323
C	3.130375	-1.780495	-11.098861
N	2.268863	-2.341649	-11.964621
C	1.123784	-2.914157	-11.522260
C	-3.736864	-2.409006	-11.045027
C	-4.957205	-1.940648	-10.584625
C	-5.198847	-2.056078	-9.244376
N	-4.318389	-2.599856	-8.382744
C	-3.118351	-3.060318	-8.814018
C	-2.186822	-3.566403	-7.879865
C	-5.979780	-5.384389	-6.287925
C	-6.186004	-5.240874	-4.944132
C	-0.840993	-3.675756	-8.302072
C	0.231124	-3.479399	-12.454201
C	-3.199438	-3.580058	-3.718499
C	-1.251832	-2.730756	-1.926255
C	-0.301276	-2.055828	-1.122280
C	0.864522	-1.526734	-1.649997

C	1.061407	-1.665345	-2.995695
C	-0.954577	-2.823033	-3.301972
C	-1.872964	-3.438233	-4.174952
C	-2.537322	-3.852386	-6.491436
C	-1.525123	-3.806241	-5.539127
C	-0.236416	-4.223056	-6.005199
C	-3.813850	-4.266907	-5.980786
C	0.728452	-4.794464	-5.138309
C	1.944148	-5.270405	-5.600720
C	2.172369	-5.186261	-6.945026
N	1.286270	-4.664409	-7.813637
C	0.092660	-4.189181	-7.376123
C	-4.124740	-4.167929	-4.606192
C	-4.780395	-4.908428	-6.794546
N	-5.287495	-4.670301	-4.120809
N	0.183081	-2.283691	-3.805704
H	-4.255181	-1.659492	-17.012115
H	-1.415729	-1.396953	-20.184611
H	0.867905	-2.145908	-19.762832
H	5.403675	-4.603750	-15.878867
H	-4.884253	-5.842240	-12.908882
H	4.041619	-1.374389	-11.522653
H	-6.122444	-1.697647	-8.802180
H	-7.095159	-5.592882	-4.467810
H	-0.478122	-1.901930	-0.060260
H	1.952694	-1.275431	-3.475800
H	3.091217	-5.559855	-7.381460
H	-2.434529	-4.904980	-16.247540
H	-3.519789	-2.258094	-12.100276
H	-4.589518	-5.074366	-7.852152
C	-7.022972	-1.824371	-2.168013
C	-7.227715	-1.962310	-0.823526
N	-6.311721	-2.489747	0.010111
C	-5.117087	-2.924618	-0.460590
C	-4.816512	-2.845245	-1.837042
C	-5.805345	-2.259287	-2.666346
C	-3.533224	-3.232999	-2.341434
C	-2.505710	-3.187930	-1.405782
C	-4.154866	-3.435804	0.437401
C	-4.473003	-3.773406	1.820525
C	-3.444735	-3.741942	2.755716
C	-2.154785	-4.109826	2.250314
C	-5.731315	-4.237524	2.335765

C	-1.165306	-4.698799	3.077550
C	0.055481	-5.124502	2.580069
C	0.265428	-4.968491	1.239369
N	-0.647819	-4.436227	0.405880
C	-1.849157	-4.014848	0.876536
C	-6.700756	-4.868554	1.515781
C	-7.877396	-5.395611	2.025251
C	-8.058974	-5.312572	3.377219
N	-7.158336	-4.751431	4.204367
C	-6.018476	-4.199240	3.717388
C	-2.817355	-3.509516	-0.019238
C	-5.094100	-3.615979	4.608888
C	-5.419096	-3.312770	5.997987
C	-4.386314	-3.270568	6.929104
C	-3.141205	-2.787202	6.404192
C	-6.706737	-2.963823	6.516087
C	-2.182887	-2.130827	7.213208
C	-1.021858	-1.589131	6.688029
C	-0.841659	-1.683651	5.336757
N	-1.731823	-2.272313	4.519687
C	-2.858749	-2.834384	5.021420
C	-7.715034	-2.385887	5.704177
C	-8.944396	-2.002939	6.214516
C	-9.136443	-2.168855	7.556429
N	-8.197736	-2.677714	8.374678
C	-6.998879	-3.082238	7.890492
C	-6.029675	-3.605021	8.772776
C	-9.828701	-5.545239	10.263628
C	-10.029836	-5.489279	11.615866
C	-4.689021	-3.628897	8.314039
C	-3.778281	-3.431703	4.138620
C	-6.959684	-4.099484	12.906950
C	-7.282800	-4.019217	14.275140
C	-6.388898	-3.525870	15.210294
C	-5.157845	-3.046784	14.786968
C	-4.291215	-2.463772	15.720909
C	-3.092953	-1.916221	15.301562
C	-2.811687	-1.964644	13.956516
C	-4.785576	-3.117066	13.422606
C	-5.662211	-3.737369	12.474526
C	-6.350878	-4.006349	10.140214
C	-5.320539	-3.962437	11.082525
C	-4.013195	-4.253057	10.575576

C	-7.623622	-4.506837	10.626063
C	-3.002602	-4.819813	11.392753
C	-1.763486	-5.186438	10.894245
C	-1.557958	-4.999519	9.557832
N	-2.490574	-4.490557	8.731635
C	-3.707758	-4.119310	9.204428
C	-7.917914	-4.560693	12.003050
C	-8.618449	-5.064087	9.782059
N	-9.106485	-5.023250	12.474261
N	-3.622047	-2.538145	13.039590
H	-8.149803	-1.637257	-0.354473
H	-5.616774	-2.097377	-3.725670
H	1.189779	-5.290550	0.774636
H	-8.948544	-5.707234	3.857005
H	0.042747	-1.276759	4.857938
H	-10.065844	-1.885024	8.036646
H	-10.954099	-5.834666	12.066912
H	-6.688348	-3.479645	16.254038
H	-4.555002	-2.411305	16.774299
H	-1.896082	-1.531032	13.563635
H	-0.621368	-5.277762	9.090399
H	-6.529989	-4.985710	0.448426
H	-7.537188	-2.194506	4.648058
H	-8.451145	-5.140717	8.710676
H	-1.359129	-4.875942	4.133196
H	-2.347688	-2.002708	8.280320
H	-3.195192	-5.035598	12.440981
H	1.510199	-2.141888	-8.227326
H	0.519324	-4.916972	-4.077873
H	2.755189	-4.856540	-12.564376
H	2.553950	-3.007705	-18.244579
H	-8.265518	-4.337774	14.624235
H	4.989637	-5.252536	-13.504176
H	3.603279	-1.234518	-9.093729
H	2.671630	-5.715370	-4.933965
H	1.587012	-1.011634	-1.029302
H	0.799607	-5.585216	3.217825
H	-0.292579	-1.091350	7.314559
H	-1.004463	-5.631641	11.525543
H	-2.408841	-1.439287	15.995866
H	-10.580348	-5.945022	9.595189
H	-9.709091	-1.569060	5.583329
H	-8.609527	-5.865982	1.381662

H	-7.769495	-1.375060	-2.811116
H	-6.711036	-5.862657	-6.926729
H	-5.674644	-1.479271	-11.252246
H	-4.503539	-5.890435	-15.381846
H	-3.722900	-1.004079	-19.364621
C	-6.486908	3.778124	-19.133303
C	-6.834417	3.483286	-17.837293
N	-5.998582	2.900999	-16.951376
C	-4.735298	2.569827	-17.309471
C	-4.305744	2.877109	-18.622569
C	-5.197232	3.468745	-19.529130
C	-3.000602	2.619621	-19.020378
C	-2.098427	2.102977	-18.105946
C	-3.824685	1.963882	-16.383923
C	-4.187071	1.567194	-15.034411
C	-3.181380	1.526669	-14.069571
C	-1.867739	1.163684	-14.545256
C	-5.470887	1.117502	-14.580473
C	-0.893035	0.552470	-13.720303
C	0.360593	0.194298	-14.198825
C	0.632302	0.461930	-15.514350
N	-0.265395	1.011732	-16.354149
C	-1.510362	1.312244	-15.898151
C	-6.420099	0.527496	-15.451289
C	-7.624558	0.011777	-14.996621
C	-7.852982	0.063109	-13.649292
N	-6.971469	0.583309	-12.776082
C	-5.804822	1.123994	-13.210489
C	-2.473184	1.802359	-16.783760
C	-4.900634	1.667702	-12.278318
C	-5.280047	1.973467	-10.904193
C	-4.283968	1.983268	-9.936212
C	-3.008334	2.432702	-10.399185
C	-6.576203	2.367599	-10.432681
C	-2.087547	3.073321	-9.537276
C	-0.898378	3.601974	-9.999733
C	-0.643875	3.491199	-11.336511
N	-1.495885	2.915347	-12.203353
C	-2.660612	2.379164	-11.765886
C	-7.536870	2.968297	-11.284683
C	-8.766513	3.406312	-10.820617
C	-9.011332	3.258750	-9.484434
N	-8.121135	2.727538	-8.626600

C	-6.916190	2.287122	-9.064146
C	-5.981717	1.772649	-8.138454
C	-9.737617	-0.123346	-6.563864
C	-9.934657	-0.014913	-5.215601
C	-4.639512	1.659897	-8.565041
C	-3.558730	1.816423	-12.693791
C	-6.969564	1.683283	-3.978267
C	-5.013400	2.522643	-2.185501
C	-4.059640	3.184568	-1.373478
C	-2.905649	3.742054	-1.897813
C	-2.719266	3.637514	-3.248025
C	-4.729078	2.468884	-3.564411
C	-5.649187	1.858638	-4.439491
C	-6.323897	1.469579	-6.754638
C	-5.307867	1.512561	-5.806844
C	-4.023108	1.095040	-6.278907
C	-7.591347	1.028639	-6.247293
C	-3.056304	0.515786	-5.420286
C	-1.855912	0.016424	-5.896777
C	-1.644508	0.085210	-7.244596
N	-2.525851	0.631743	-8.102728
C	-3.705283	1.128525	-7.651247
C	-7.892259	1.094099	-4.869620
C	-8.551589	0.386683	-7.068359
N	-9.040275	0.554805	-4.387253
N	-3.598791	3.029016	-4.064719
H	-7.822768	3.710543	-17.451747
H	-4.881500	3.713233	-20.540165
H	-2.658613	2.851461	-20.025520
H	1.595935	0.217851	-15.949895
H	-8.766785	-0.325625	-13.211697
H	0.282588	3.868086	-11.753783
H	-9.946820	3.578761	-9.041049
H	-10.831832	-0.396640	-4.739476
H	-4.223459	3.301906	-0.304832
H	-1.835550	4.047385	-3.726548
H	-0.743346	-0.320426	-7.690148
H	-6.208050	0.431915	-16.513148
H	-7.316809	3.141634	-12.335930
H	-8.363913	0.244483	-8.130039
C	-10.836143	3.255374	-2.361723
C	-11.027619	3.059163	-1.023213
N	-10.079523	2.563703	-0.207737

C	-8.869840	2.196908	-0.696339
C	-8.583932	2.327191	-2.070288
C	-9.600080	2.900146	-2.875542
C	-7.294255	1.991182	-2.593191
C	-6.256600	2.032552	-1.667690
C	-7.881840	1.697872	0.178753
C	-8.175455	1.324355	1.554156
C	-7.141162	1.376546	2.479518
C	-5.848973	1.042207	1.959359
C	-9.412036	0.808261	2.068760
C	-4.840521	0.462487	2.768455
C	-3.624134	0.051857	2.247527
C	-3.435959	0.217619	0.904351
N	-4.363377	0.756008	0.090395
C	-5.561652	1.159707	0.583716
C	-10.358587	0.144539	1.247701
C	-11.514624	-0.425660	1.757687
C	-11.697158	-0.350838	3.110175
N	-10.814847	0.239127	3.937111
C	-9.693685	0.829801	3.449950
C	-6.547400	1.668212	-0.289689
C	-8.780801	1.436609	4.337942
C	-9.106636	1.718651	5.726886
C	-8.071495	1.774215	6.654072
C	-6.838846	2.294171	6.132678
C	-10.402140	2.029036	6.247085
C	-5.890830	2.959264	6.947055
C	-4.751885	3.546963	6.422785
C	-4.578070	3.478244	5.068823
N	-5.454845	2.874975	4.247604
C	-6.563405	2.275550	4.748153
C	-11.424057	2.592089	5.441167
C	-12.665120	2.931084	5.953256
C	-12.855214	2.732036	7.291109
N	-11.902502	2.243011	8.105128
C	-10.690468	1.883861	7.618306
C	-9.704300	1.380888	8.492640
C	-13.425539	-0.699049	9.985372
C	-13.629482	-0.642946	11.337812
C	-8.363801	1.393224	8.032647
C	-7.472638	1.663844	3.865041
C	-10.612986	0.869654	12.620427
C	-10.936916	0.956526	13.987674

C	-10.058792	1.494552	14.913456
C	-8.844602	2.009113	14.481022
C	-7.999302	2.639647	15.403155
C	-6.820931	3.220671	14.972554
C	-6.536106	3.154956	13.628855
C	-8.469040	1.927691	13.118150
C	-9.327824	1.268165	12.182638
C	-10.010765	0.965044	9.856229
C	-8.981543	1.040768	10.794494
C	-7.671248	0.771693	10.285855
C	-11.263680	0.422628	10.343898
C	-6.654244	0.214582	11.099830
C	-5.418474	-0.152047	10.591727
C	-5.217903	0.040133	9.254442
N	-6.152902	0.556089	8.434388
C	-7.371582	0.912654	8.914619
C	-11.555427	0.367057	11.721375
C	-12.234766	-0.174592	9.500726
N	-12.726671	-0.135126	12.194834
N	-7.324046	2.534704	12.722929
H	-11.966122	3.305303	-0.542077
H	-9.423370	3.105468	-3.929405
H	-2.517862	-0.099949	0.423360
H	-12.571628	-0.777061	3.590788
H	-3.707280	3.915353	4.591481
H	-13.795590	2.969917	7.773271
H	-14.539413	-1.022865	11.790298
H	-10.358173	1.546990	15.956962
H	-8.264258	2.702566	16.455639
H	-5.633033	3.608029	13.229380
H	-4.283017	-0.238573	8.782615
H	-10.183728	0.038089	0.179710
H	-11.248279	2.804161	4.388577
H	-12.061591	-0.249869	8.430127
H	-5.016427	0.274901	3.825537
H	-6.047394	3.057515	8.018486
H	-6.841048	-0.000978	12.149207
H	-2.301121	3.200338	-8.479077
H	-3.253181	0.398700	-4.356935
H	-1.120766	0.318107	-12.682706
H	-1.070027	1.943336	-18.431259
H	-11.907946	0.609797	14.342316
H	1.095233	-0.283485	-13.561567

H	-0.192859	4.083903	-9.337209
H	-1.127094	-0.437382	-5.238335
H	-2.181440	4.246206	-1.269639
H	-2.865348	-0.407403	2.868611
H	-4.030575	4.051707	7.053215
H	-4.656614	-0.602065	11.216555
H	-6.152824	3.731797	15.658163
H	-14.160529	-1.132917	9.319737
H	-13.439192	3.354247	5.325714
H	-12.230501	-0.920863	1.114534
H	-11.606299	3.685602	-2.989570
H	-10.463820	-0.602785	-7.207414
H	-9.489322	3.865625	-11.483528
H	-8.343074	-0.427937	-15.676347
H	-7.188004	4.254045	-19.807577
C	-9.458043	9.003455	-19.726022
C	-9.862834	8.768501	-18.433728
N	-9.082485	8.182473	-17.498814
C	-7.822340	7.788709	-17.801131
C	-7.333512	8.039867	-19.106539
C	-8.168634	8.631447	-20.064711
C	-6.023301	7.732276	-19.445673
C	-5.171787	7.228368	-18.477913
C	-6.969483	7.182522	-16.822781
C	-7.399271	6.817012	-15.481710
C	-6.437859	6.792692	-14.472318
C	-5.104078	6.421611	-14.878988
C	-8.703736	6.372185	-15.077308
C	-4.166691	5.847209	-13.990435
C	-2.902521	5.450767	-14.407489
C	-2.578723	5.646274	-15.723370
N	-3.434436	6.172616	-16.618284
C	-4.690017	6.510847	-16.221620
C	-9.627960	5.808144	-15.990695
C	-10.842127	5.273925	-15.586525
C	-11.110178	5.285711	-14.246117
N	-10.261948	5.794188	-13.333751
C	-9.084714	6.349889	-13.716656
C	-5.606785	6.982645	-17.162310
C	-8.222152	6.890023	-12.741416
C	-8.642863	7.161412	-11.366132
C	-7.665148	7.216186	-10.381092
C	-6.417718	7.757283	-10.820298

C	-9.959420	7.497037	-10.901960
C	-5.580080	8.490962	-9.946857
C	-4.444084	9.141076	-10.397955
C	-4.144084	9.026754	-11.727746
N	-4.903228	8.334243	-12.598716
C	-6.034098	7.715336	-12.177952
C	-10.933307	8.072732	-11.755150
C	-12.175996	8.481139	-11.295467
C	-12.423778	8.329390	-9.959712
N	-11.524918	7.812539	-9.101606
C	-10.309502	7.397162	-9.535608
C	-9.372065	6.905026	-8.603356
C	-13.089207	4.849259	-7.108795
C	-13.297658	4.905868	-5.759141
C	-8.021518	6.862400	-9.015804
C	-6.875002	7.088249	-13.117707
C	-10.414995	6.685039	-4.450119
C	-8.538932	7.624634	-2.625659
C	-7.654432	8.362729	-1.801311
C	-6.543603	9.015224	-2.312987
C	-6.326334	8.915853	-3.660424
C	-8.229377	7.588558	-4.001994
C	-9.099653	6.931880	-4.894227
C	-9.724426	6.561847	-7.227644
C	-8.723361	6.629886	-6.264984
C	-7.412160	6.289523	-6.729011
C	-10.983407	6.063424	-6.742997
C	-6.432561	5.740106	-5.867034
C	-5.205655	5.303099	-6.336549
C	-4.977060	5.412205	-7.679058
N	-5.864609	5.942109	-8.541120
C	-7.073850	6.368135	-8.096705
C	-11.300166	6.077727	-5.365671
C	-11.921216	5.421006	-7.589559
N	-12.434087	5.489053	-4.907667
N	-7.133864	8.224625	-4.486018
H	-10.854940	9.050724	-18.093743
H	-7.809303	8.827593	-21.072242
H	-5.635600	7.919468	-20.443764
H	-1.604994	5.364958	-16.110835
H	-12.032075	4.875250	-13.847084
H	-3.255026	9.493141	-12.138820
H	-13.366829	8.634704	-9.517949

H	-14.181091	4.471979	-5.302207
H	-7.853231	8.477383	-0.737669
H	-5.472950	9.392934	-4.132110
H	-4.052656	5.054967	-8.117076
H	-9.387453	5.739987	-17.048048
H	-10.710120	8.257351	-12.803774
H	-11.728961	5.324720	-8.655026
C	-14.335173	8.193161	-2.872480
C	-14.525040	8.049651	-1.526226
N	-13.580380	7.562034	-0.700508
C	-12.373491	7.170089	-1.179895
C	-12.079839	7.272757	-2.556255
C	-13.101412	7.812525	-3.377377
C	-10.775086	6.967977	-3.064425
C	-9.756331	7.061723	-2.122594
C	-11.391156	6.691860	-0.288483
C	-11.697961	6.312755	1.086133
C	-10.675327	6.387321	2.026062
C	-9.363507	6.106304	1.521762
C	-12.935019	5.771973	1.585396
C	-8.344891	5.567349	2.345940
C	-7.103396	5.211401	1.844741
C	-6.898615	5.395840	0.506072
N	-7.835378	5.891378	-0.324684
C	-9.058908	6.236891	0.149282
C	-13.879949	5.123697	0.749183
C	-15.024005	4.516822	1.245061
C	-15.203019	4.545222	2.599468
N	-14.333426	5.134191	3.440057
C	-13.223169	5.755990	2.966563
C	-10.051509	6.698131	-0.742383
C	-12.335625	6.373351	3.870346
C	-12.688188	6.638149	5.258041
C	-11.666436	6.738065	6.194473
C	-10.460393	7.322684	5.686900
C	-13.993841	6.926059	5.771034
C	-9.580788	8.064820	6.512512
C	-8.483745	8.739424	6.000379
C	-8.276231	8.659139	4.650429
N	-9.080826	7.967536	3.822550
C	-10.162174	7.308152	4.306442
C	-15.021592	7.468662	4.958364
C	-16.255624	7.837072	5.471443

C	-16.439266	7.678052	6.816675
N	-15.488917	7.184823	7.632222
C	-14.282722	6.802001	7.145081
C	-13.295389	6.316443	8.025763
C	-16.984519	4.169411	9.520309
C	-17.173455	4.191390	10.875958
C	-11.954015	6.357479	7.572243
C	-11.031654	6.652333	3.411366
C	-14.196391	5.785399	12.155170
C	-14.516348	5.865470	13.523843
C	-13.670922	6.476352	14.436036
C	-12.496672	7.067394	13.987979
C	-11.708689	7.802933	14.884718
C	-10.584499	8.470738	14.431379
C	-10.281102	8.360867	13.094689
C	-12.111004	6.966161	12.629809
C	-12.932427	6.238440	11.711648
C	-13.600792	5.901620	9.389447
C	-12.576114	6.008954	10.329439
C	-11.255587	5.775274	9.831733
C	-14.843869	5.333124	9.877271
C	-10.229800	5.255285	10.658297
C	-8.981868	4.917726	10.159954
C	-8.776292	5.111903	8.822705
N	-9.717888	5.597931	7.991090
C	-10.950933	5.915748	8.461710
C	-15.124326	5.252962	11.257081
C	-15.813955	4.735534	9.032696
N	-16.278363	4.713902	11.732315
N	-11.002806	7.628404	12.217775
H	-15.457049	8.335712	-1.049484
H	-12.924774	7.987346	-4.436638
H	-5.956115	5.131205	0.041886
H	-16.065690	4.084054	3.069154
H	-7.433285	9.153083	4.178103
H	-17.370059	7.954206	7.301639
H	-18.065236	3.773127	11.331094
H	-13.973088	6.535146	15.478495
H	-11.984946	7.883945	15.933235
H	-9.412669	8.859908	12.675010
H	-7.829224	4.861344	8.360099
H	-13.712148	5.052405	-0.322432
H	-14.849436	7.655363	3.900425

H	-15.653541	4.685578	7.958876
H	-8.529597	5.369066	3.399571
H	-9.773391	8.165491	7.578615
H	-10.418443	5.045036	11.708538
H	-5.836309	8.609474	-8.896151
H	-6.639360	5.591544	-4.809571
H	-4.434874	5.661210	-12.953224
H	-4.135004	7.034997	-18.755291
H	-15.464421	5.468656	13.887624
H	-2.199405	4.994050	-13.722681
H	-3.809276	9.704736	-9.727217
H	-4.469376	4.863129	-5.677103
H	-5.875108	9.579442	-1.674957
H	-6.337726	4.781189	2.478679
H	-7.818249	9.305406	6.639870
H	-8.213846	4.490377	10.793832
H	-9.964725	9.059063	15.098295
H	-17.714779	3.726768	8.854911
H	-17.028369	8.254414	4.838545
H	-15.732502	4.027074	0.589295
H	-15.103254	8.607533	-3.513128
H	-13.793364	4.362999	-7.771825
H	-12.906318	8.922803	-11.961559
H	-11.536281	4.847029	-16.299114
H	-10.117225	9.477494	-20.443160
C	9.524031	-5.861896	-14.489834
C	9.305304	-6.004126	-13.141990
N	10.282816	-6.240659	-12.233027
C	11.575635	-6.355168	-12.630847
C	11.864549	-6.177134	-14.007750
C	10.831365	-5.945948	-14.926101
C	13.169710	-6.213053	-14.472916
C	14.207386	-6.390154	-13.576297
C	12.650431	-6.615323	-11.704983
C	12.475375	-6.878325	-10.281091
C	13.537435	-6.618621	-9.403237
C	14.857906	-6.778506	-9.964919
C	11.345717	-7.505883	-9.667844
C	15.999385	-7.093140	-9.184816
C	17.266920	-7.230638	-9.734298
C	17.379963	-7.074166	-11.088809
N	16.324486	-6.837524	-11.888056
C	15.078421	-6.729977	-11.353627

C	10.513459	-8.409722	-10.371916
C	9.498348	-9.119577	-9.748093
C	9.319388	-8.905572	-8.409240
N	10.082509	-8.060130	-7.693935
C	11.094793	-7.379121	-8.285507
C	13.979180	-6.565388	-12.199520
C	11.950752	-6.597874	-7.489536
C	11.612950	-6.250681	-6.122121
C	12.645656	-6.086340	-5.210758
C	13.842071	-5.520500	-5.769816
C	10.305031	-5.968776	-5.614892
C	14.770822	-4.788350	-4.987468
C	15.862185	-4.137926	-5.541086
C	16.001377	-4.211354	-6.897795
N	15.146636	-4.889388	-7.686212
C	14.090536	-5.558684	-7.160697
C	9.267115	-5.495254	-6.451178
C	8.030586	-5.126304	-5.945188
C	7.847965	-5.240092	-4.595702
N	8.803447	-5.679859	-3.757357
C	10.024171	-6.030307	-4.234888
C	11.038737	-6.429290	-3.342046
C	7.593794	-8.759115	-1.670985
C	7.379702	-8.612186	-0.327956
C	12.372464	-6.402520	-3.813192
C	13.221189	-6.280687	-8.012735
C	10.118704	-6.524382	0.816892
C	11.993830	-5.520016	2.604374
C	12.915081	-4.808017	3.414321
C	14.052577	-4.206572	2.897890
C	14.247679	-4.306290	1.549043
C	12.293792	-5.571883	1.226785
C	11.422658	-6.249734	0.349711
C	10.763944	-6.743398	-1.949577
C	11.785880	-6.577241	-1.023123
C	13.099720	-6.852995	-1.529618
C	9.563237	-7.317856	-1.418563
C	14.147858	-7.316716	-0.695645
C	15.393318	-7.666512	-1.194443
C	15.569370	-7.569817	-2.546891
N	14.600267	-7.156798	-3.384751
C	13.380934	-6.797464	-2.911564
C	9.259227	-7.235844	-0.043947

C	8.691689	-8.107229	-2.209896
N	8.179900	-7.878721	0.467580
N	13.398411	-4.956223	0.732365
H	8.298250	-5.935333	-12.744541
H	11.042380	-5.816723	-15.984596
H	13.400709	-6.072273	-15.525639
H	18.338464	-7.154012	-11.591313
H	8.535549	-9.411123	-7.854841
H	16.818572	-3.717692	-7.413775
H	6.910148	-4.960803	-4.130776
H	6.536401	-9.084401	0.165499
H	12.732862	-4.683548	4.478780
H	15.107428	-3.855559	1.064042
H	16.509973	-7.835692	-3.018458
H	10.686364	-8.613322	-11.425928
H	9.433650	-5.356256	-7.517221
H	8.889517	-8.259205	-3.269056
C	6.157749	-5.069356	2.300317
C	5.949494	-5.203367	3.644855
N	6.890211	-5.650953	4.498507
C	8.121665	-5.994439	4.043371
C	8.430863	-5.912667	2.668891
C	7.407591	-5.424123	1.818917
C	9.749633	-6.196001	2.183533
C	10.765644	-6.060905	3.121961
C	9.117601	-6.407914	4.954404
C	8.821284	-6.736647	6.342000
C	9.832425	-6.589786	7.284434
C	11.149030	-6.875381	6.796181
C	7.615345	-7.315714	6.855130
C	12.179627	-7.356766	7.642741
C	13.426917	-7.720762	7.159642
C	13.621812	-7.620599	5.810307
N	12.669837	-7.188737	4.962210
C	11.449345	-6.812295	5.419871
C	6.744499	-8.083522	6.042765
C	5.646182	-8.749337	6.563502
C	5.432946	-8.641179	7.910200
N	6.231700	-7.927992	8.725342
C	7.307655	-7.266252	8.232045
C	10.460046	-6.390314	4.507917
C	8.157752	-6.565330	9.113578
C	7.782011	-6.257882	10.485064

C	8.787426	-6.101399	11.437594
C	10.012728	-5.544713	10.919408
C	6.451723	-6.013024	10.954936
C	10.925778	-4.814084	11.722274
C	12.061264	-4.206295	11.207830
C	12.264953	-4.315963	9.862180
N	11.423321	-4.977801	9.048522
C	10.320754	-5.597741	9.542196
C	5.433294	-5.514667	10.104013
C	4.163444	-5.215769	10.570752
C	3.927177	-5.422693	11.900876
N	4.864622	-5.872479	12.756939
C	6.118919	-6.148130	12.316955
C	7.114967	-6.529000	13.237417
C	3.771739	-9.169451	14.654230
C	3.468546	-9.078427	15.987624
C	8.466707	-6.424687	12.831438
C	9.459519	-6.276294	8.657149
C	5.995110	-6.858238	17.303643
C	5.522825	-6.668741	18.614886
C	6.219815	-5.905084	19.536237
C	7.408675	-5.293622	19.159853
C	8.073552	-4.463834	20.073795
C	9.235794	-3.811206	19.701375
C	9.688975	-4.014101	18.420002
C	7.940730	-5.482725	17.861758
C	7.250447	-6.320718	16.926237
C	6.789480	-6.893878	14.602693
C	7.731046	-6.612375	15.588573
C	9.091032	-6.764132	15.170400
C	5.614263	-7.599828	15.046745
C	10.122627	-7.091876	16.083669
C	11.414703	-7.374923	15.669156
C	11.653330	-7.350893	14.323140
N	10.705573	-7.051038	13.414789
C	9.443336	-6.742570	13.801981
C	5.234191	-7.606411	16.402391
C	4.841793	-8.418429	14.187785
N	4.158694	-8.307322	16.847427
N	9.070777	-4.814270	17.523087
H	5.000697	-4.935338	4.094738
H	7.594679	-5.266834	0.758763
H	14.564467	-7.897875	5.349605

H	4.590725	-9.128655	8.390605
H	13.123485	-3.862895	9.377488
H	2.957008	-5.213643	12.336151
H	2.635954	-9.628403	16.415567
H	5.803503	-5.771073	20.531516
H	7.678992	-4.312204	21.075707
H	10.589805	-3.527167	18.058359
H	12.632122	-7.581868	13.915040
H	6.943550	-8.209986	4.980612
H	5.638546	-5.308447	9.055559
H	5.095291	-8.508950	13.133549
H	12.000876	-7.497053	8.706782
H	10.738755	-4.675714	12.783561
H	9.912060	-7.187680	17.146324
H	14.627289	-4.679908	-3.915653
H	13.985309	-7.454222	0.371376
H	15.904560	-7.258342	-8.113611
H	15.224588	-6.378890	-13.968593
H	4.573038	-7.107056	18.923097
H	18.126933	-7.453597	-9.115493
H	16.557679	-3.578948	-4.928283
H	16.181948	-8.021866	-0.543351
H	14.744198	-3.668062	3.533122
H	14.201731	-8.090014	7.819496
H	12.743890	-3.656077	11.842700
H	12.190626	-7.635005	16.378079
H	9.763644	-3.159900	20.387154
H	3.190794	-9.794768	13.989624
H	3.397154	-4.816954	9.917161
H	4.986452	-9.331909	5.933091
H	5.388006	-4.677898	1.646397
H	6.933738	-9.358241	-2.285534
H	7.247039	-4.740523	-6.585116
H	8.868652	-9.809990	-10.294446
H	8.703747	-5.684747	-15.174786
C	6.169403	-0.869902	-15.203549
C	5.920621	-0.914043	-13.853811
N	6.837931	-1.274653	-12.925376
C	8.097599	-1.613036	-13.299585
C	8.424866	-1.539422	-14.676839
C	7.448569	-1.186329	-15.618610
C	9.714753	-1.793507	-15.116197
C	10.704616	-2.070949	-14.191141

C	9.105627	-1.999487	-12.345778
C	8.865127	-2.198836	-10.922324
C	9.929003	-2.043929	-10.023197
C	11.240424	-2.356813	-10.547904
C	7.645871	-2.657129	-10.331021
C	12.334770	-2.732656	-9.730460
C	13.591173	-3.013953	-10.250132
C	13.740398	-2.949631	-11.607868
N	12.727230	-2.653046	-12.439362
C	11.491238	-2.397645	-11.933759
C	6.700378	-3.420030	-11.056075
C	5.570170	-3.954744	-10.461764
C	5.406744	-3.736440	-9.123783
N	6.284626	-3.039045	-8.382629
C	7.387721	-2.496425	-8.952758
C	10.435855	-2.145786	-12.811855
C	8.311088	-1.815600	-8.139634
C	7.992196	-1.418490	-6.772851
C	9.029958	-1.297068	-5.860550
C	10.254533	-0.817394	-6.433857
C	6.707708	-1.041922	-6.263530
C	11.198272	-0.078473	-5.678368
C	12.338863	0.468210	-6.245113
C	12.506816	0.292497	-7.589953
N	11.626065	-0.375508	-8.357651
C	10.522551	-0.946959	-7.815966
C	5.700369	-0.507244	-7.100187
C	4.489736	-0.059301	-6.595912
C	4.304500	-0.140886	-5.245022
N	5.231493	-0.635552	-4.405777
C	6.424603	-1.072538	-4.881657
C	7.407752	-1.530355	-3.982868
C	3.774842	-3.583537	-2.333727
C	3.582308	-3.457997	-0.986733
C	8.742395	-1.570140	-4.451833
C	9.614079	-1.634851	-8.650841
C	6.467437	-1.604098	0.186950
C	8.374417	-0.667463	1.974282
C	9.303525	0.039167	2.779130
C	10.465397	0.592057	2.263938
C	10.671980	0.455603	0.920158
C	8.678843	-0.748944	0.599168
C	7.782958	-1.394385	-0.279409

C	7.105876	-1.832941	-2.588902
C	8.131932	-1.731863	-1.655658
C	9.437104	-2.049215	-2.157842
C	5.864370	-2.326320	-2.062470
C	10.463334	-2.548904	-1.318204
C	11.714382	-2.891055	-1.803156
C	11.912096	-2.762634	-3.148216
N	10.956731	-2.342157	-3.995833
C	9.733441	-1.984995	-3.537561
C	5.569558	-2.249881	-0.684272
C	4.923916	-3.020949	-2.863312
N	4.451488	-2.826655	-0.177658
N	9.806408	-0.176119	0.106305
H	4.941552	-0.651356	-13.467834
H	7.686133	-1.135857	-16.678032
H	9.977328	-1.738369	-16.169207
H	14.693237	-3.160060	-12.082602
H	4.543520	-4.122279	-8.594007
H	13.367197	0.695752	-8.114134
H	3.389757	0.211105	-4.781856
H	2.701314	-3.867836	-0.504635
H	9.109994	0.192537	3.838048
H	11.552769	0.863676	0.435407
H	12.861932	-3.006499	-3.607426
H	6.858256	-3.651431	-12.106004
H	5.872827	-0.387776	-8.167617
H	5.093298	-3.161064	-3.928222
C	2.584966	0.041340	1.704311
C	2.377988	-0.084762	3.049233
N	3.298691	-0.587399	3.893062
C	4.507019	-0.994496	3.428964
C	4.813603	-0.919733	2.053155
C	3.812174	-0.373988	1.212825
C	6.114156	-1.264214	1.558701
C	7.134296	-1.168660	2.495828
C	5.481828	-1.466928	4.333195
C	5.169955	-1.806246	5.717013
C	6.189039	-1.724555	6.659180
C	7.497946	-2.034240	6.165402
C	3.931912	-2.322524	6.225884
C	8.517542	-2.550979	7.003998
C	9.769211	-2.893513	6.520373
C	9.974867	-2.746868	5.178346

N	9.027826	-2.302185	4.333831
C	7.804232	-1.942260	4.791912
C	3.002317	-2.999783	5.400140
C	1.860369	-3.592803	5.909274
C	1.666017	-3.516285	7.259679
N	2.526052	-2.903627	8.092342
C	3.635402	-2.294856	7.607344
C	6.823279	-1.492188	3.883520
C	4.523670	-1.665473	8.501577
C	4.174963	-1.377904	9.888990
C	5.203183	-1.275339	10.816479
C	6.423686	-0.722453	10.295683
C	2.871280	-1.094189	10.413256
C	7.342064	-0.004025	11.104176
C	8.487213	0.585463	10.592035
C	8.690438	0.471618	9.246221
N	7.835133	-0.168970	8.429041
C	6.721265	-0.771866	8.919616
C	1.841736	-0.552525	9.604781
C	0.604415	-0.202577	10.121522
C	0.417306	-0.387245	11.462253
N	1.370587	-0.874192	12.279141
C	2.586177	-1.225218	11.790772
C	3.588018	-1.689223	12.671732
C	-0.042232	-3.824059	14.287138
C	-0.221078	-3.752298	15.642096
C	4.921621	-1.647459	12.198769
C	5.833348	-1.419667	8.037242
C	2.760111	-2.118718	16.834641
C	2.449193	-1.999840	18.202880
C	3.328588	-1.425913	19.105586
C	4.534870	-0.912898	18.649802
C	5.386978	-0.255554	19.549167
C	6.565687	0.313753	19.099349
C	6.841799	0.211246	17.757154
C	4.897028	-1.024576	17.286495
C	4.032046	-1.711336	16.370673
C	3.313723	-2.077477	14.054595
C	4.356698	-1.961526	14.978720
C	5.666705	-2.195098	14.455214
C	2.082696	-2.629809	14.576511
C	6.729995	-2.673229	15.260896
C	7.986129	-2.944759	14.744880

C	8.151602	-2.772339	13.401852
N	7.161704	-2.373554	12.586859
C	5.935905	-2.073200	13.078115
C	1.817605	-2.663308	15.961662
C	1.112800	-3.263781	13.763425
N	0.674361	-3.190751	16.471028
N	6.039578	-0.421837	16.873752
H	1.448687	0.235892	3.506201
H	4.001383	-0.219001	0.152783
H	10.924161	-2.993849	4.719073
H	0.789278	-3.953597	7.724258
H	9.559102	0.906047	8.762712
H	-0.521453	-0.130160	11.939686
H	-1.105737	-4.167549	16.114265
H	3.036191	-1.347651	20.149445
H	5.125193	-0.170278	20.601033
H	7.741230	0.645618	17.331088
H	9.101407	-2.955004	12.916664
H	3.175522	-3.105028	4.332000
H	2.014065	-0.350738	8.549778
H	1.257988	-3.351343	12.689642
H	8.332342	-2.725021	8.061914
H	7.154833	0.124146	12.167280
H	6.578163	-2.877617	16.318222
H	11.030939	0.111267	-4.620925
H	10.285059	-2.708381	-0.256819
H	12.214700	-2.832355	-8.653935
H	11.718837	-2.224624	-14.560720
H	1.485563	-2.347450	18.576677
H	14.418046	-3.286192	-9.607114
H	13.052632	1.024816	-5.651831
H	12.491923	-3.262044	-1.147971
H	11.167387	1.120163	2.896065
H	10.540489	-3.280504	7.173537
H	9.179827	1.120436	11.228570
H	8.791046	-3.301155	15.373872
H	7.236699	0.833093	19.772301
H	-0.775629	-4.304583	13.651934
H	-0.172503	0.223509	9.498659
H	1.151857	-4.105117	5.270507
H	1.830372	0.473050	1.059147
H	3.059685	-4.107612	-2.955423
H	3.727350	0.362309	-7.236895

H	4.852690	-4.538821	-11.024033
H	5.395788	-0.583862	-15.905504
C	2.308459	4.460355	-15.412510
C	2.084967	4.407773	-14.058836
N	3.001426	3.991144	-13.153643
C	4.234145	3.593024	-13.560055
C	4.538033	3.672294	-14.941967
C	3.563022	4.091425	-15.857687
C	5.803066	3.353333	-15.411942
C	6.791898	2.992033	-14.514407
C	5.238530	3.131852	-12.636668
C	5.016551	2.926272	-11.213851
C	6.104554	2.993007	-10.334905
C	7.373183	2.578897	-10.889321
C	3.788522	2.527070	-10.604615
C	8.439542	2.085357	-10.098182
C	9.651242	1.689151	-10.647938
C	9.782303	1.768058	-12.007657
N	8.792503	2.189230	-12.814539
C	7.594476	2.545939	-12.280701
C	2.799825	1.805655	-11.313080
C	1.670269	1.301871	-10.690511
C	1.540809	1.526854	-9.349792
N	2.453706	2.198852	-8.629758
C	3.565987	2.695504	-9.222689
C	6.546206	2.903066	-13.132109
C	4.539133	3.322080	-8.426892
C	4.261020	3.733853	-7.058619
C	5.315480	3.780313	-6.158161
C	6.569155	4.164452	-6.740827
C	3.002438	4.177466	-6.536031
C	7.573014	4.826634	-5.991709
C	8.758734	5.262730	-6.560964
C	8.912153	5.057192	-7.902995
N	7.971866	4.470678	-8.666988
C	6.819148	4.008945	-8.123227
C	2.011035	4.764198	-7.358832
C	0.823667	5.256502	-6.840197
C	0.647821	5.169777	-5.488138
N	1.563386	4.634416	-4.661865
C	2.731666	4.148185	-5.153103
C	3.699158	3.629463	-4.270453
C	0.001515	1.698064	-2.647977

C	-0.185879	1.813872	-1.299414
C	5.025253	3.516564	-4.749873
C	5.845618	3.409615	-8.953769
C	2.770256	3.542421	-0.104093
C	4.721766	4.373801	1.686935
C	5.682185	5.029932	2.497692
C	6.863140	5.543364	1.984813
C	7.058802	5.417965	0.638166
C	5.017048	4.301668	0.309767
C	4.092587	3.706032	-0.572459
C	3.392687	3.324514	-2.881527
C	4.425070	3.369656	-1.952279
C	5.710634	2.988558	-2.464498
C	2.134950	2.873480	-2.360859
C	6.718298	2.436994	-1.635444
C	7.935859	2.007362	-2.138132
C	8.119341	2.099579	-3.488261
N	7.184787	2.588752	-4.321903
C	5.996166	3.033275	-3.847471
C	1.844904	2.945457	-0.982353
C	1.172085	2.221179	-3.170580
N	0.705693	2.404054	-0.483652
N	6.165241	4.833785	-0.181031
H	1.129206	4.715711	-13.649595
H	3.782537	4.148197	-16.920514
H	6.047785	3.414477	-16.468870
H	10.700675	1.472136	-12.504311
H	0.676655	1.171647	-8.801449
H	9.808123	5.370908	-8.428733
H	-0.249151	5.548047	-5.011597
H	-1.079480	1.428098	-0.820612
H	5.500919	5.171872	3.560347
H	7.953994	5.796428	0.155737
H	9.037331	1.767994	-3.960508
H	2.924594	1.574956	-12.368069
H	2.178017	4.881873	-8.427380
H	1.341326	2.080634	-4.235679
C	-1.066674	5.239309	1.473489
C	-1.269977	5.078653	2.815030
N	-0.350342	4.548364	3.642347
C	0.849148	4.131587	3.163217
C	1.148984	4.234995	1.788146
C	0.154069	4.822618	0.967792

C	2.436892	3.868013	1.276105
C	3.466397	3.908618	2.207085
C	1.818191	3.613443	4.049789
C	1.507686	3.246719	5.427318
C	2.534009	3.275422	6.365968
C	3.830941	2.943220	5.851464
C	0.257442	2.753988	5.928654
C	4.844706	2.378886	6.665122
C	6.078033	2.005240	6.156335
C	6.271667	2.164645	4.814288
N	5.331730	2.665166	3.994273
C	4.127590	3.057963	4.476977
C	-0.695200	2.124507	5.091268
C	-1.861452	1.569716	5.588693
C	-2.051101	1.622365	6.940739
N	-1.164472	2.179529	7.783949
C	-0.034705	2.759656	7.310421
C	3.155060	3.561908	3.589002
C	0.875772	3.336620	8.217327
C	0.534972	3.601416	9.610820
C	1.562695	3.638134	10.545518
C	2.807280	4.150976	10.039638
C	-0.762208	3.919187	10.132333
C	3.756011	4.802340	10.869018
C	4.921402	5.365257	10.372554
C	5.114414	5.295400	9.022080
N	4.232939	4.718037	8.185859
C	3.101377	4.136845	8.660110
C	-1.768321	4.512645	9.328755
C	-3.003076	4.879160	9.839364
C	-3.209153	4.663861	11.172149
N	-2.274408	4.139040	11.986496
C	-1.062955	3.766071	11.503113
C	-0.082097	3.256849	12.383315
C	-3.802819	1.252840	13.966467
C	-3.992025	1.334623	15.319709
C	1.257387	3.253186	11.921076
C	2.192316	3.550109	7.756766
C	-0.966815	2.870667	16.531669
C	-1.291799	3.025927	17.892799
C	-0.403319	3.583247	18.796877
C	0.829771	4.037149	18.349579
C	1.694382	4.678297	19.248502

C	2.902969	5.187540	18.805905
C	3.196675	5.041830	17.471292
C	1.208122	3.881700	16.994732
C	0.326402	3.219460	16.078058
C	-0.382225	2.868994	13.760119
C	0.655482	2.944449	14.691893
C	1.957763	2.653227	14.179154
C	-1.636806	2.363123	14.272151
C	2.986048	2.118779	14.994972
C	4.228338	1.775256	14.488154
C	4.415383	1.941967	13.145828
N	3.461090	2.412872	12.323670
C	2.245751	2.771746	12.804687
C	-1.918123	2.349840	15.654285
C	-2.619653	1.760657	13.451709
N	-3.084267	1.865753	16.155167
N	2.382916	4.422700	16.588611
H	-2.194722	5.392392	3.284937
H	0.342604	5.002571	-0.088417
H	7.203047	1.882307	4.337087
H	-2.941218	1.204728	7.398758
H	5.996683	5.715545	8.550704
H	-4.147563	4.926819	11.646680
H	-4.894744	0.951052	15.784646
H	-0.706919	3.694190	19.834534
H	1.420444	4.797076	20.293960
H	4.121193	5.426897	17.051492
H	5.354493	1.690667	12.668329
H	-0.523044	2.028111	4.022038
H	-1.579319	4.736105	8.280960
H	-2.463769	1.658081	12.380439
H	4.668812	2.187410	7.721482
H	3.576988	4.903557	11.936544
H	2.811416	1.921028	16.050055
H	7.423430	5.037105	-4.935385
H	6.548587	2.297682	-0.569939
H	8.327625	1.977824	-9.021475
H	7.786465	2.779794	-14.907719
H	-2.273648	2.723832	18.258516
H	10.459059	1.323104	-10.027282
H	9.520301	5.755546	-5.970800
H	8.698794	1.592599	-1.493016
H	7.589071	6.032361	2.621356

H	6.845538	1.581581	6.789715
H	5.636265	5.849865	11.024732
H	5.005396	1.367374	15.121666
H	3.585216	5.692419	19.478562
H	-4.547684	0.797997	13.325936
H	-3.762257	5.337948	9.218114
H	-2.587094	1.095137	4.940356
H	-1.817552	5.696153	0.841414
H	-0.730329	1.205044	-3.275339
H	0.072814	5.711735	-7.471791
H	0.921799	0.741248	-11.234886
H	1.537734	4.797661	-16.094125
C	-0.932045	9.132418	-15.899484
C	-1.206771	9.200979	-14.556352
N	-0.300404	8.934574	-13.585551
C	0.953131	8.525421	-13.901700
C	1.307923	8.493848	-15.274061
C	0.356290	8.788250	-16.260763
C	2.598409	8.180847	-15.666968
C	3.552304	7.908218	-14.704226
C	1.921092	8.155201	-12.897113
C	1.648192	8.057265	-11.462069
C	2.716575	8.162546	-10.557861
C	3.997016	7.705982	-11.047055
C	0.383783	7.759093	-10.854391
C	5.034553	7.250873	-10.196159
C	6.261251	6.820654	-10.682573
C	6.436675	6.817697	-12.038666
N	5.475257	7.195793	-12.898267
C	4.264173	7.592570	-12.424768
C	-0.637924	7.060166	-11.537842
C	-1.834717	6.712808	-10.933698
C	-1.987319	7.038475	-9.619321
N	-1.030778	7.652775	-8.908242
C	0.135320	8.020149	-9.489472
C	3.247982	7.901709	-13.329905
C	1.126695	8.610277	-8.689713
C	0.855894	9.039760	-7.323964
C	1.907744	9.045063	-6.421443
C	3.164589	9.417366	-6.998736
C	-0.382588	9.532356	-6.798089
C	4.174415	10.063057	-6.244765
C	5.354167	10.507229	-6.820818

C	5.493655	10.324881	-8.169488
N	4.551842	9.740240	-8.934119
C	3.408961	9.266553	-8.382390
C	-1.345410	10.176698	-7.614007
C	-2.507230	10.723469	-7.090528
C	-2.686065	10.626633	-5.737628
N	-1.794570	10.037905	-4.918179
C	-0.654598	9.493697	-5.416595
C	0.284842	8.909064	-4.544326
C	-3.501941	7.067008	-3.035851
C	-3.696703	7.127712	-1.684742
C	1.608468	8.764710	-5.019686
C	2.439721	8.648252	-9.203020
C	-0.687949	8.694965	-0.391903
C	1.261644	9.451361	1.436551
C	2.219347	10.084279	2.268119
C	3.410073	10.594270	1.774098
C	3.615623	10.493413	0.426258
C	1.567591	9.404934	0.059409
C	0.641755	8.846961	-0.844934
C	-0.044007	8.566760	-3.170223
C	0.981589	8.550081	-2.231834
C	2.265180	8.157938	-2.746347
C	-1.322509	8.137645	-2.679292
C	3.256201	7.562767	-1.927538
C	4.466717	7.119291	-2.436340
C	4.661091	7.244674	-3.782737
N	3.743486	7.778235	-4.608268
C	2.563106	8.235982	-4.125944
C	-1.621914	8.157818	-1.300754
C	-2.305278	7.563640	-3.524714
N	-2.786857	7.639702	-0.835800
N	2.723096	9.934805	-0.411699
H	-2.199293	9.470714	-14.209831
H	0.614375	8.749780	-17.316128
H	2.887375	8.163203	-16.714326
H	7.368199	6.488319	-12.487256
H	-2.901600	6.814448	-9.084539
H	6.382386	10.653148	-8.698801
H	-3.565638	11.037223	-5.252517
H	-4.609418	6.755487	-1.231074
H	2.025962	10.217244	3.330586
H	4.517354	10.874059	-0.042071

H	5.574817	6.905287	-4.258066
H	-0.496983	6.732287	-12.563954
H	-1.171617	10.290674	-8.682127
H	-2.133269	7.464476	-4.593919
C	-4.491514	10.467773	1.197946
C	-4.694891	10.320386	2.541532
N	-3.789590	9.757744	3.363889
C	-2.612968	9.284578	2.880286
C	-2.307692	9.384759	1.506310
C	-3.286377	10.007280	0.690456
C	-1.026198	8.990907	0.995902
C	-0.003823	8.995974	1.937608
C	-1.670266	8.722521	3.765753
C	-2.008584	8.344981	5.133835
C	-0.990559	8.337888	6.081555
C	0.306407	7.986212	5.578960
C	-3.282317	7.885917	5.616206
C	1.300390	7.397967	6.400546
C	2.533129	7.007180	5.903692
C	2.745395	7.176685	4.565083
N	1.823309	7.697309	3.736018
C	0.621428	8.108171	4.208493
C	-4.255965	7.314277	4.759872
C	-5.451646	6.803226	5.235903
C	-5.650680	6.837082	6.587167
N	-4.745458	7.335791	7.448539
C	-3.586483	7.875503	6.995820
C	-0.332105	8.640694	3.315020
C	-2.664109	8.407173	7.919470
C	-3.002310	8.658422	9.317278
C	-1.975450	8.668747	10.251711
C	-0.732624	9.186076	9.757976
C	-4.284127	9.010703	9.851571
C	0.201606	9.837846	10.601498
C	1.370201	10.403103	10.115488
C	1.576051	10.336137	8.765301
N	0.703652	9.760318	7.917574
C	-0.430895	9.179063	8.380449
C	-5.282281	9.635192	9.061895
C	-6.485906	10.067615	9.595735
C	-6.664193	9.894968	10.939698
N	-5.739318	9.327251	11.737131
C	-4.569113	8.868343	11.227140

C	-3.614183	8.278523	12.084361
C	-7.475583	6.452963	13.511081
C	-7.673337	6.455296	14.864272
C	-2.279423	8.252251	11.616737
C	-1.340414	8.604217	7.469689
C	-4.523286	7.590719	16.204271
C	-4.832780	7.590220	17.578583
C	-3.911895	7.977974	18.536970
C	-2.664299	8.432813	18.135663
C	-1.765538	8.915005	19.098516
C	-0.543375	9.434246	18.708139
C	-0.272416	9.460194	17.361138
C	-2.305333	8.439307	16.766447
C	-3.213061	7.921357	15.782988
C	-3.932844	7.820287	13.436976
C	-2.891153	7.775249	14.372326
C	-1.590528	7.498870	13.834333
C	-5.221849	7.353233	13.907567
C	-0.566584	6.885598	14.600308
C	0.678936	6.584213	14.073364
C	0.871419	6.857482	12.750326
N	-0.083424	7.385358	11.965801
C	-1.296473	7.714509	12.472159
C	-5.511228	7.240812	15.282990
C	-6.244399	6.888739	13.043472
N	-6.722846	6.825015	15.739226
N	-1.118659	8.991744	16.418319
H	-5.606217	10.667341	3.017241
H	-3.094766	10.181062	-0.366255
H	3.679060	6.885951	4.097720
H	-6.562242	6.449454	7.029824
H	2.461609	10.759805	8.303158
H	-7.571611	10.221828	11.436412
H	-8.612766	6.125648	15.296662
H	-4.202088	7.965854	19.584269
H	-2.023792	8.905544	20.154571
H	0.658734	9.868010	16.979308
H	1.812480	6.647468	12.256685
H	-4.079252	7.231382	3.690244
H	-5.108066	9.833847	8.006679
H	-6.086741	6.843901	11.968522
H	1.109445	7.203887	7.453837
H	0.005400	9.938048	11.667097

H	-0.743481	6.597542	15.633835
H	4.030960	10.260144	-5.184368
H	3.078687	7.400212	-0.866658
H	4.890308	7.203464	-9.119106
H	4.565363	7.684553	-15.040955
H	-5.825842	7.292459	17.916418
H	7.046704	6.491380	-10.015628
H	6.120591	10.993871	-6.231349
H	5.216228	6.669676	-1.798599
H	4.134117	11.066388	2.425550
H	3.285515	6.564910	6.543472
H	2.077509	10.889389	10.774773
H	1.453359	6.128211	14.676649
H	0.164607	9.821847	19.430076
H	-8.248600	6.113624	12.833594
H	-7.240231	10.543839	8.982511
H	-6.191627	6.373898	4.571702
H	-5.228650	10.944075	0.564195
H	-4.247066	6.632605	-3.691018
H	-3.233310	11.221717	-7.720271
H	-2.614680	6.189904	-11.469560
H	-1.696019	9.343881	-16.638027

4.4. Computational Methods for the Preferred Hopf Cyclization Pathway

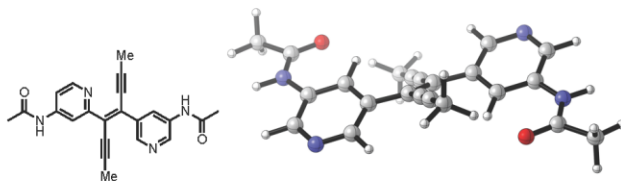
All geometries were fully optimized at the B3LYP/6-31G(d) level of theory in the gas phase.⁵⁻⁸ All optimized geometries were verified by frequency computations as minima (no imaginary frequencies) or transition structures (one imaginary frequency). Frequency analysis was performed at 298.15 K. Free energy corrections were calculated using Truhlar's quasiharmonic approximation, which sets all the real vibrational frequencies that are lower than 100 cm⁻¹ to 100 cm⁻¹ to correct entropies for the breakdown of the harmonic oscillator approximation.^{9,10} Subsequent single-point energy calculations on the optimized structures were performed using the M06-2X/6-311+G(d,p) level of theory.¹¹ The thermal corrections calculated from vibrational frequencies at the B3LYP/6-31G(d) level on the optimized geometries were then added to the M06-2X/6-311+G(d,p) electronic energies to obtain the Gibbs free energies. All quantum chemical computations were performed using Gaussian 09.¹² All graphics on optimized structures were generated with CYLview.¹³

Conformational searches were carried out with *MacroModel* from *Schrödinger* using OPLS_2005 and an energy window of 10.0 kcal mol⁻¹.¹⁴ A redundant conformer elimination was performed using an energy window of 10.0 kcal mol⁻¹ and a maximum atom deviation cutoff of 0.5 Å. The lowest energy conformers were optimized with B3LYP/6-31G(d) to locate the global minimum for each reaction.

4.5. XYZ Geometries

Substituted Model System

7

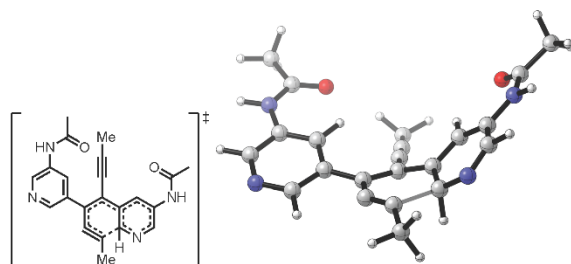


C	-3.85523	-2.54708	-0.66947
C	-3.89940	-1.26503	-0.10302
C	-2.69048	-0.61468	0.16611
C	-1.57334	-2.55580	-0.70740
C	-0.14679	-0.66674	0.09873
C	0.14686	0.66670	-0.09904
C	1.49099	1.26511	0.15479
C	0.83569	-1.58997	0.56037
C	1.62780	-2.42573	0.94597
C	1.57319	2.55609	0.70661
C	3.85509	2.54738	0.66933
C	3.89944	1.26510	0.10339
C	2.69060	0.61464	-0.16581
H	2.69787	-0.36501	-0.61714
H	0.66376	3.09740	0.95291
H	-0.66397	-3.09702	-0.95412
H	-2.69759	0.36480	0.61780
C	-1.49098	-1.26505	-0.15509
C	-1.62799	2.42533	-0.94653
C	-0.83571	1.58978	-0.56082
C	-2.64500	3.37910	-1.37144
H	-2.37428	4.40286	-1.08607
H	-2.77498	3.36232	-2.46116
H	-3.60589	3.13178	-0.90476
C	2.64466	-3.37973	1.37071
H	2.37428	-4.40327	1.08421
H	3.60584	-3.13188	0.90490
H	2.77389	-3.36396	2.46054
H	4.78320	3.07976	0.88767
H	-4.78340	-3.07936	-0.88778
N	5.17023	0.72266	-0.18174

N	-5.17010	-0.72268	0.18277
C	5.49977	-0.57637	-0.49528
C	-5.49957	0.57644	0.49606
H	5.94111	1.36633	-0.06459
H	-5.94101	-1.36637	0.06591
O	4.67979	-1.47632	-0.62181
O	-4.67954	1.47642	0.62206
C	6.98977	-0.82605	-0.68162
H	7.14909	-1.26088	-1.67284
H	7.61193	0.06896	-0.58337
H	7.31661	-1.56660	0.05525
C	-6.98953	0.82617	0.68257
H	-7.31704	1.56469	-0.05606
H	-7.14841	1.26354	1.67273
H	-7.61145	-0.06928	0.58697
N	2.72348	3.18465	0.96219
N	-2.72370	-3.18425	-0.96290

SCF energy: -1219.57775 hartree
zero-point correction: +0.377081 hartree
enthalpy correction: +0.406316 hartree
free energy correction: +0.314572 hartree
quasiharmonic free energy correction: +0.322698 hartree

8

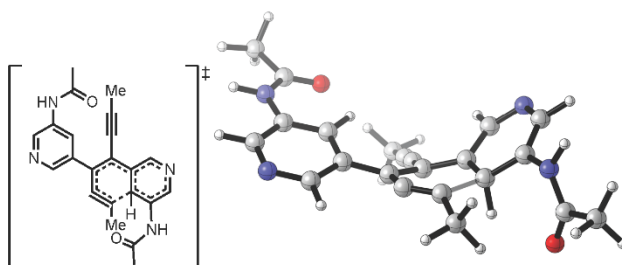


C	-4.62537	1.46473	-0.59909
C	-3.92755	0.24961	-0.50853
C	-2.56846	0.28256	-0.18551
C	-2.77531	2.67509	-0.04145
C	-0.53361	1.72213	0.38097
C	0.33145	0.71619	0.89156

C	1.74069	0.81559	0.59301
C	0.05629	2.94554	0.02475
C	1.25114	3.40277	-0.06328
C	2.41253	2.10939	0.42141
C	4.19628	1.06522	-0.63766
C	3.70411	-0.24126	-0.33358
C	2.47422	-0.35104	0.29888
H	2.04707	-1.32556	0.48875
H	2.46850	2.68012	1.35744
H	-2.34075	3.64972	0.16599
H	-2.00807	-0.63776	-0.14626
C	-1.96398	1.52649	0.06416
C	-0.52405	-1.55523	1.92298
C	-0.15319	-0.51001	1.42835
C	-1.02180	-2.81548	2.45994
H	-1.85728	-3.17641	1.84879
H	-0.23621	-3.58032	2.46096
H	-1.37436	-2.69576	3.49233
C	1.93199	4.72667	-0.10448
H	2.41405	4.96101	0.85494
H	2.72033	4.73506	-0.86521
H	1.20457	5.51127	-0.32788
H	-5.68447	1.46699	-0.86565
H	5.12600	1.15934	-1.20705
N	3.64304	2.17262	-0.23331
N	-4.06671	2.65160	-0.37740
N	4.49941	-1.32384	-0.73833
C	4.29667	-2.67158	-0.51023
H	5.33001	-1.07423	-1.25837
O	3.34397	-3.11514	0.11244
C	5.35637	-3.58135	-1.11480
H	4.89464	-4.18515	-1.90316
H	6.21581	-3.05114	-1.53725
H	5.70759	-4.26922	-0.34063
N	-4.64878	-0.93161	-0.78248
C	-4.26973	-2.23928	-0.58121
H	-5.59260	-0.78347	-1.11326
O	-3.17296	-2.57316	-0.15324
C	-5.33115	-3.27018	-0.93904
H	-5.56370	-3.85831	-0.04573
H	-6.25889	-2.84351	-1.33283
H	-4.91433	-3.95717	-1.68171

SCF energy: -1219.500711 hartree
 zero-point correction: +0.375505 hartree
 enthalpy correction: +0.403605 hartree
 free energy correction: +0.314993 hartree
 quasiharmonic free energy correction: +0.322328 hartree

8[‡]

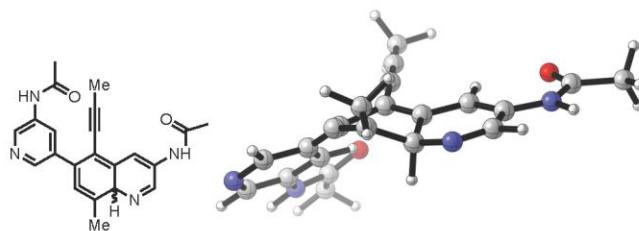


C	-3.92383	2.71154	-0.15660
C	-1.71504	2.79471	0.40432
C	-2.67641	0.65245	-0.11924
C	-0.19689	0.84528	0.45988
C	0.09222	-0.53184	0.71489
C	1.33447	-1.05116	0.21589
C	0.90704	1.65545	0.22370
C	2.17868	1.49271	0.09030
C	2.56407	-0.23867	0.12693
C	3.33169	-1.97375	-1.44731
C	1.30774	-2.31295	-0.43639
H	0.43087	-2.94266	-0.30794
H	2.99768	-0.03941	1.11648
H	-2.62227	-0.41372	-0.27133
H	-0.86676	3.39781	0.71884
C	-1.54993	1.40258	0.25681
C	-1.74706	-2.21288	1.57682
C	-0.90922	-1.44167	1.15443
C	-2.80369	-3.10533	2.03574
H	-3.71100	-2.94066	1.44260
H	-2.50869	-4.15677	1.93776
H	-3.04332	-2.92242	3.09085
C	3.33633	2.41441	0.31550
H	3.99854	2.04294	1.10397

H	3.93937	2.52625	-0.59368
H	2.96224	3.40191	0.59724
H	-4.85561	3.25682	-0.32047
H	4.07881	-2.36789	-2.13691
C	-3.89011	1.31668	-0.31448
N	-2.86398	3.43803	0.18891
N	2.22764	-2.74644	-1.27528
C	3.57889	-0.80835	-0.76770
N	4.84140	-0.18159	-0.93971
C	5.70187	0.14592	0.08906
H	5.25770	-0.28927	-1.85597
O	5.37744	0.09825	1.26823
N	-5.08853	0.67748	-0.69515
C	-5.38260	-0.66727	-0.69560
H	-5.84579	1.30342	-0.93393
O	-4.58142	-1.53948	-0.38664
C	-6.80087	-1.00757	-1.13006
H	-7.26834	-1.61683	-0.35088
H	-7.43288	-0.13607	-1.32705
H	-6.75204	-1.61929	-2.03654
C	7.06688	0.64746	-0.35435
H	7.34568	0.32868	-1.36385
H	7.07106	1.74347	-0.32680
H	7.81563	0.29469	0.35823

SCF energy: -1219.496056 hartree
 zero-point correction: +0.375626 hartree
 enthalpy correction: +0.403462 hartree
 free energy correction: +0.316340 hartree
 quasiharmonic free energy correction: +0.322810 hartree

9



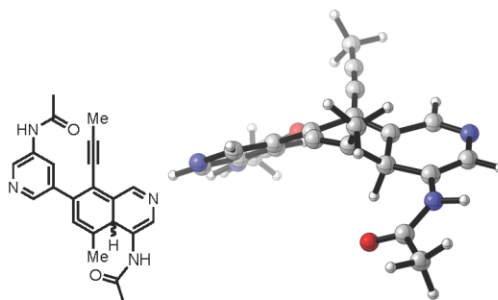
C	4.77247	1.39817	-0.44706
---	---------	---------	----------

C	4.03514	0.20477	-0.47404
C	2.67026	0.25716	-0.17401
C	2.95894	2.61764	0.20780
C	0.68322	1.68094	0.48763
C	-0.28041	0.54009	0.69026
C	-1.51638	0.70591	0.09971
C	0.06385	2.87645	0.44526
C	-1.20960	3.17014	0.17836
C	-1.85802	2.04886	-0.61684
C	-4.06580	1.29781	-0.88635
C	-3.73243	-0.07453	-0.47973
C	-2.49860	-0.33539	0.04315
H	-2.26383	-1.31320	0.44185
H	-1.37622	1.98651	-1.60988
H	2.55785	3.57711	0.52685
H	2.07816	-0.64440	-0.20947
C	2.10624	1.49519	0.17365
C	0.48569	-1.69003	1.85946
C	0.10658	-0.66294	1.33233
C	0.98759	-2.93078	2.43650
H	1.82814	-3.30422	1.83864
H	1.33875	-2.78048	3.46527
H	0.20956	-3.70311	2.45691
C	-2.01553	4.29686	0.74664
H	-2.38800	4.95863	-0.04602
H	-2.90396	3.91443	1.26554
H	-1.41406	4.88676	1.44363
H	-5.10986	1.49622	-1.15869
H	5.83440	1.38976	-0.70090
N	-4.78209	-0.99199	-0.61262
N	4.72007	-0.97638	-0.82801
C	-4.79474	-2.33344	-0.27163
C	4.30603	-2.28299	-0.68983
C	-6.08862	-3.05935	-0.60864
H	-6.91572	-2.39563	-0.88020
H	-5.90393	-3.74608	-1.44198
H	-6.38354	-3.66303	0.25362
C	5.33593	-3.32192	-1.11170
H	4.88934	-3.96552	-1.87552
H	6.26667	-2.90031	-1.50394
H	5.56989	-3.95417	-0.24934
H	-5.62804	-0.62516	-1.02810
H	5.66977	-0.83686	-1.14548

O	-3.84503	-2.90306	0.24147
O	3.20496	-2.60861	-0.26795
N	4.25127	2.57922	-0.11477
N	-3.26720	2.29263	-0.91269

SCF energy: -1219.516951 hartree
 zero-point correction: +0.377093 hartree
 enthalpy correction: +0.405364 hartree
 free energy correction: +0.316222 hartree
 quasiharmonic free energy correction: +0.323958 hartree

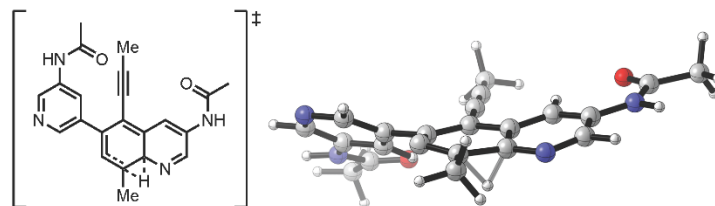
9'



C	-3.53923	-2.83580	0.28388
C	-3.56652	-1.52075	-0.20289
C	-2.44291	-0.71224	-0.00327
C	-1.45704	-2.56491	1.17684
C	-0.11227	-0.49641	0.95965
C	0.05194	0.98691	0.69099
C	1.25838	1.35715	0.13996
C	1.05832	-1.05946	1.29553
C	2.31012	-0.63399	1.11369
C	2.35182	0.27297	-0.10987
C	3.66665	0.90048	-0.51775
C	3.76286	2.23371	-0.73577
C	1.57462	2.72771	-0.18472
H	0.79036	3.47175	-0.04807
H	2.01544	-0.32539	-0.97115
H	-0.64499	-2.98254	1.76824
H	-2.43373	0.29585	-0.38789
C	-1.35152	-1.24146	0.70262

C	-1.96544	2.61822	1.12584
C	-1.00665	1.89740	0.93064
C	-3.15147	3.44449	1.30956
H	-3.49024	3.42942	2.35308
H	-2.96037	4.48815	1.03306
H	-3.96432	3.06314	0.67934
C	3.42715	-0.76033	2.10210
H	4.22593	-1.41204	1.72580
H	3.88002	0.21964	2.30210
H	3.05960	-1.18466	3.04113
H	4.70805	2.66380	-1.06657
H	-4.39375	-3.49593	0.12316
N	-4.72539	-1.10120	-0.88917
C	-5.11677	0.18086	-1.20487
C	-6.46828	0.27896	-1.89827
H	-6.32807	0.76219	-2.87010
H	-6.96684	-0.68320	-2.05129
H	-7.12071	0.92608	-1.30401
H	-5.38770	-1.83712	-1.09397
O	-4.44739	1.17663	-0.96417
N	4.80386	0.09140	-0.69943
C	4.84496	-1.24152	-1.06962
H	5.67653	0.60034	-0.76715
C	6.24591	-1.77898	-1.32447
H	7.04023	-1.16523	-0.88807
H	6.30816	-2.79104	-0.91758
H	6.41136	-1.84851	-2.40577
O	3.84879	-1.93731	-1.19364
N	2.73428	3.15616	-0.57965
N	-2.51046	-3.35026	0.95817

SCF energy: -1219.511649 hartree
 zero-point correction: +0.377185 hartree
 enthalpy correction: +0.405382 hartree
 free energy correction: +0.316992 hartree
 quasiharmonic free energy correction: +0.324138 hartree

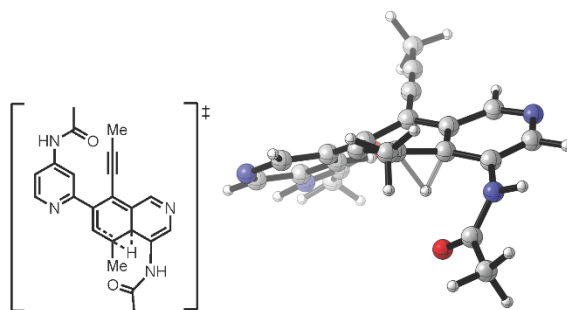


C	-4.76415	1.28642	-0.86972
C	-4.18288	0.25617	-0.11900
C	-2.80767	0.31247	0.13310
C	-2.77075	2.37656	-1.10126
C	-0.60855	1.56705	-0.15039
C	0.27408	0.45555	-0.12502
C	1.67178	0.67193	0.12143
C	-0.14956	2.91351	-0.08753
C	1.14761	3.15741	0.33240
C	2.12674	1.98779	0.50494
C	4.31843	1.29268	0.56627
C	3.97674	-0.04238	0.19650
C	2.63928	-0.34666	-0.00489
H	2.33785	-1.34823	-0.27615
H	1.47333	2.37964	1.46049
H	-2.22448	3.23750	-1.47284
H	-2.34108	-0.46484	0.71843
C	-2.07235	1.38958	-0.37998
C	-0.45175	-2.01683	-0.69494
C	-0.13642	-0.87088	-0.44854
C	-0.91301	-3.37550	-0.94968
H	-1.86235	-3.54722	-0.42848
H	-1.07080	-3.54506	-2.02226
H	-0.18230	-4.11418	-0.59989
C	1.70785	4.53461	0.59769
H	2.31549	4.58091	1.50685
H	2.36421	4.83106	-0.23123
H	0.88107	5.24339	0.66526
H	-5.83543	1.26781	-1.08235
H	5.36930	1.54398	0.73187
N	3.45832	2.26102	0.74281
N	-4.08203	2.32561	-1.35091
N	5.03686	-0.95286	0.07792
C	4.98141	-2.29663	-0.24586

H	5.95531	-0.56920	0.25594
O	3.94132	-2.88954	-0.48422
C	6.33217	-2.99592	-0.27603
H	7.18648	-2.33251	-0.10863
H	6.33917	-3.77934	0.48859
H	6.44877	-3.48692	-1.24653
N	-5.03341	-0.76738	0.35476
C	-4.69436	-2.00270	0.85397
H	-6.02066	-0.58729	0.23180
O	-3.54097	-2.38437	1.00983
C	-5.87860	-2.88604	1.22227
H	-6.85581	-2.44176	1.00889
H	-5.79225	-3.83158	0.67839
H	-5.82296	-3.11878	2.29035

SCF energy: -1219.488153 hartree
 zero-point correction: +0.373174 hartree
 enthalpy correction: +0.401112 hartree
 free energy correction: +0.312864 hartree
 quasiharmonic free energy correction: +0.320207 hartree

10'



C	3.70653	-2.67984	-0.40376
C	3.61907	-1.41653	0.19990
C	2.43171	-0.69361	0.05618
C	1.60497	-2.51691	-1.28111
C	0.08464	-0.58132	-0.89036
C	-0.09113	0.84616	-0.76296
C	-1.37416	1.30920	-0.38555
C	-1.01862	-1.34351	-1.26375

C	-2.32103	-0.88042	-1.08183
C	-2.50217	0.38422	-0.29946
C	-3.84485	2.30751	0.38290
C	-1.64264	2.68550	-0.12708
H	-0.82165	3.39235	-0.22143
H	-2.35674	-0.62815	0.39549
H	0.81821	-2.95318	-1.88882
H	2.33097	0.27111	0.52981
C	1.39328	-1.24863	-0.70727
C	1.89401	2.53436	-1.16618
C	0.96500	1.77133	-0.99342
C	3.06188	3.39495	-1.31089
H	2.81838	4.44174	-1.09302
H	3.84867	3.07161	-0.61837
H	3.46269	3.35203	-2.33153
C	-3.49470	-1.61446	-1.66948
H	-3.97867	-2.28555	-0.95086
H	-4.25344	-0.93135	-2.06741
H	-3.09555	-2.24159	-2.47281
H	4.61682	-3.27336	-0.29370
H	-4.78801	2.74326	0.71175
N	2.72755	-3.22424	-1.12488
C	-3.74698	0.95025	0.19212
N	-2.81652	3.17830	0.20457
N	-4.83788	0.11551	0.53727
C	-4.70383	-1.01317	1.31444
H	-5.75494	0.54175	0.49918
O	-3.60582	-1.49238	1.59096
C	-5.99640	-1.63682	1.80206
H	-6.05405	-1.52378	2.88993
H	-6.89060	-1.19703	1.35077
H	-5.97527	-2.70840	1.58618
N	4.73353	-0.96609	0.94040
C	5.02511	0.31495	1.35058
H	5.44741	-1.66381	1.10080
O	4.28678	1.27421	1.16928
C	6.36000	0.46323	2.06776
H	6.90690	-0.47559	2.19987
H	6.98637	1.16118	1.50321
H	6.18020	0.90913	3.05057

SCF energy: -1219.479542 hartree
zero-point correction: +0.372581 hartree

enthalpy correction: +0.400220 hartree
free energy correction: +0.313616 hartree
quasiharmonic free energy correction: +0.320069 hartree

Results for Unsubstituted N-PDA Model System

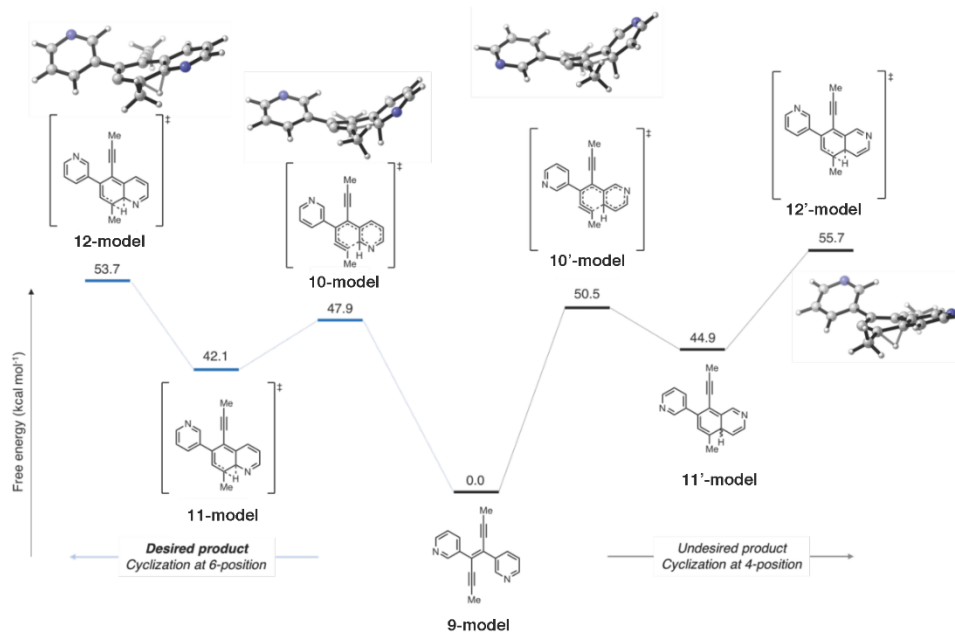
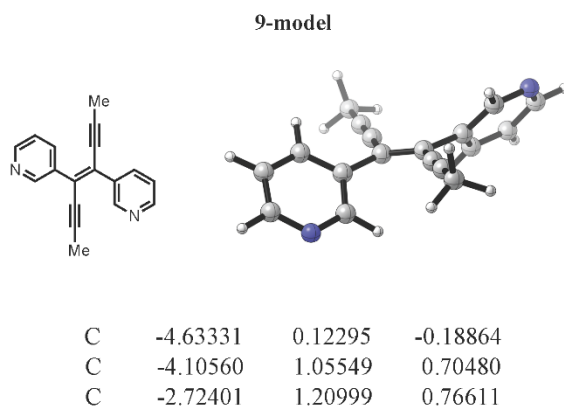


Figure S35. Potential energy surface of cyclization at 6- and 4-positions starting from structure 9-model. Energies are free energies in kcal mol⁻¹.

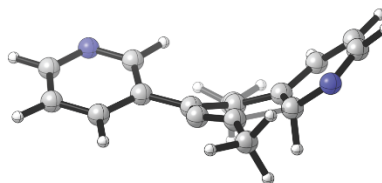
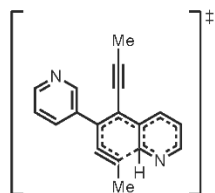
Unsubstituted Model System



C	-2.55269	-0.47185	-0.92773
C	-0.42290	0.55457	-0.02023
C	0.43047	-0.52753	-0.02218
C	1.91331	-0.39353	-0.06831
C	0.05817	1.89387	0.07139
C	0.37728	3.06115	0.17559
C	2.72537	-1.21927	0.72990
C	4.65886	-0.29494	-0.06171
C	3.95936	0.56105	-0.91225
C	2.56814	0.51148	-0.91529
H	1.99454	1.15312	-1.57520
H	2.25989	-1.95041	1.38744
H	-1.96419	-1.07568	-1.61313
H	-2.27320	1.93305	1.43948
C	-1.90610	0.41633	-0.05034
C	-0.39609	-3.02499	0.18056
C	-0.05844	-1.86323	0.07680
C	-0.84603	-4.40755	0.29056
H	-0.14703	-5.09452	-0.20137
H	-0.93164	-4.71504	1.34040
H	-1.83117	-4.53750	-0.17392
C	0.80232	4.45156	0.29163
H	0.83624	4.77118	1.34087
H	1.80561	4.59450	-0.12829
H	0.11740	5.12424	-0.23870
H	5.74734	-0.28071	-0.03407
H	-5.70949	-0.02564	-0.26287
N	4.05989	-1.17598	0.75003
N	-3.87757	-0.62433	-1.00342
H	4.49389	1.24463	-1.56566
H	-4.76306	1.64764	1.33477

SCF energy: -803.5793978 hartree
 zero-point correction: +0.266766 hartree
 enthalpy correction: +0.286689 hartree
 free energy correction: +0.215898 hartree
 quasiharmonic free energy correction: +0.222057 hartree

10-model

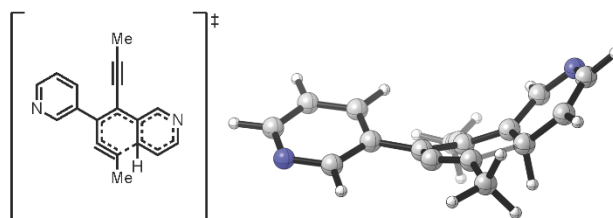


C	-4.49511	-0.97200	0.80217
C	-3.91628	-1.99716	0.05550
C	-2.58600	-1.86627	-0.33064
C	-2.54915	0.23265	0.82007
C	-0.43635	-0.61455	-0.32829
C	0.26733	0.61597	-0.46421
C	1.67332	0.62536	-0.16911
C	0.32908	-1.78208	-0.38848
C	1.57714	-2.07669	-0.42155
C	2.53527	-0.55085	-0.36889
C	4.14095	0.33003	1.06262
C	3.47328	1.58360	1.13155
C	2.25715	1.72663	0.49980
H	1.69186	2.64988	0.58102
H	3.90736	2.39824	1.70295
H	2.68986	-0.76655	-1.43503
H	-2.02906	1.11476	1.18007
H	-4.49697	-2.87157	-0.22372
H	-2.09302	-2.63089	-0.92360
C	-1.86273	-0.72032	0.04078
C	-0.96115	2.92624	-0.81235
C	-0.41065	1.86221	-0.61381
C	-1.66128	4.18586	-1.03762
H	-2.64784	4.17908	-0.55834
H	-1.09651	5.03460	-0.63360
H	-1.81323	4.36857	-2.10882
C	2.44036	-3.22386	-0.81639
H	2.90382	-3.05862	-1.79918
H	3.25624	-3.35633	-0.09774
H	1.84469	-4.13895	-0.86417
H	-5.53835	-1.02980	1.10878
H	5.06088	0.17978	1.62899

N	3.75093	-0.66346	0.30583
N	-3.82679	0.12333	1.19066

SCF energy: -803.5030161 hartree
 zero-point correction: +0.265611 hartree
 enthalpy correction: +0.284074 hartree
 free energy correction: +0.218649 hartree
 quasiharmonic free energy correction: +0.222060 hartree

10³-model

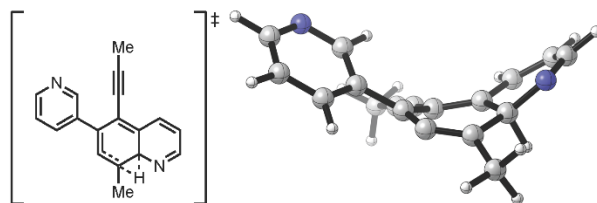


C	-4.51332	-1.08602	0.72826
C	-2.60963	-1.81841	-0.29537
C	-2.56186	0.30989	0.80164
C	-0.45162	-0.59947	-0.31801
C	0.28768	0.61295	-0.46121
C	1.69190	0.58281	-0.16554
C	0.28387	-1.78340	-0.37689
C	1.52819	-2.10128	-0.42222
C	2.51701	-0.61185	-0.37755
C	4.14251	0.44364	1.11620
C	2.28176	1.66292	0.54805
H	1.72033	2.59118	0.63462
H	2.65711	-0.81927	-1.44829
H	-2.05145	1.20637	1.13301
H	-2.11595	-2.59157	-0.88110
C	-1.87657	-0.66682	0.05937
C	-0.87839	2.95219	-0.82247
C	-0.35740	1.87526	-0.61348
C	-1.52283	4.23704	-1.06912
H	-2.56632	4.22987	-0.73157
H	-1.00421	5.04866	-0.54480

H	-1.52158	4.48006	-2.13906
C	2.31858	-3.29023	-0.85461
H	2.94405	-3.06546	-1.72962
H	2.99011	-3.63457	-0.05736
H	1.64622	-4.11246	-1.11212
H	-5.55400	-1.28278	0.98133
H	5.07030	0.42845	1.68449
C	-3.89569	0.09876	1.13379
H	-4.44618	0.83218	1.71601
N	-3.88576	-2.03734	0.02640
H	4.43369	-1.48393	0.25236
N	3.43520	1.61166	1.18459
C	3.77445	-0.62579	0.35002

SCF energy: -803.4989812 hartree
 zero-point correction: +0.265520 hartree
 enthalpy correction: +0.283912 hartree
 free energy correction: +0.218703 hartree
 quasiharmonic free energy correction: +0.222143 hartree

11-model

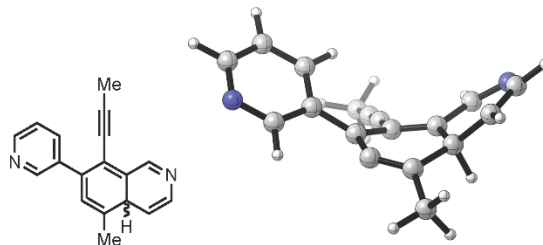


C	-4.32142	-1.05991	1.05513
C	-3.81853	-2.05342	0.21482
C	-2.54908	-1.88440	-0.32710
C	-2.41082	0.18612	0.87149
C	-0.43451	-0.59362	-0.52276
C	0.28835	0.68819	-0.53721
C	1.61311	0.61064	-0.11404
C	0.33741	-1.68271	-0.78751
C	1.64601	-1.87349	-0.75231
C	2.42750	-0.66892	-0.17281

C	3.79888	0.13588	1.53602
C	3.38074	1.50067	1.27158
C	2.30670	1.72761	0.46555
H	1.88851	2.72342	0.34901
H	3.88053	2.31622	1.78584
H	3.07690	-0.50668	-1.06077
H	-1.87168	1.07650	1.18180
H	-4.41618	-2.92917	-0.02131
H	-2.12038	-2.61222	-1.00961
C	-1.80245	-0.73924	-0.00059
C	-0.93568	2.99057	-0.93559
C	-0.36613	1.93860	-0.72590
C	-1.64213	4.24280	-1.18100
H	-2.68016	4.18254	-0.83232
H	-1.16183	5.08102	-0.66214
H	-1.66407	4.48267	-2.25147
C	2.43743	-2.95076	-1.43707
H	3.11388	-2.53565	-2.19780
H	3.06611	-3.48421	-0.71217
H	1.77167	-3.67160	-1.91854
H	-5.32012	-1.14676	1.48029
H	4.52939	-0.03696	2.32936
N	3.37348	-0.90316	0.90175
N	-3.63416	0.04219	1.38620

SCF energy: -803.5135248 hartree
 zero-point correction: +0.266927 hartree
 enthalpy correction: +0.285545 hartree
 free energy correction: +0.219460 hartree
 quasiharmonic free energy correction: +0.223345 hartree

11'-model

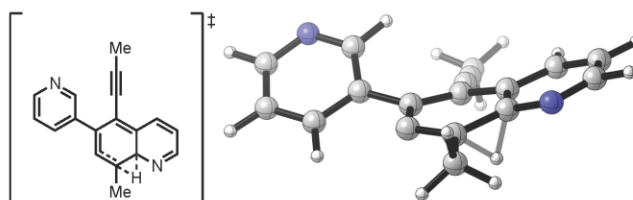


C	-4.33832	-1.16422	0.98995
C	-3.67994	0.00678	1.37078
C	-2.39707	0.24365	0.88871
C	-2.57929	-1.82699	-0.30597
C	-0.43998	-0.59213	-0.50673
C	0.31439	0.67397	-0.54235
C	1.63863	0.57196	-0.12626
C	0.30222	-1.69964	-0.77515
C	1.60989	-1.90640	-0.76511
C	2.42138	-0.72014	-0.18332
C	3.77463	0.26851	1.61515
C	2.33414	1.68251	0.48710
H	1.94620	2.68780	0.32417
H	3.07116	-0.57469	-1.07655
H	-2.16638	-2.55335	-1.00318
H	-4.15871	0.70951	2.04668
H	-1.85297	1.13106	1.19233
C	-1.80559	-0.69530	0.02661
C	-0.86880	2.99276	-0.96417
C	-0.31674	1.93443	-0.74064
C	-1.53640	4.26135	-1.23240
H	-2.62569	4.13821	-1.25385
H	-1.29664	5.00684	-0.46443
H	-1.22785	4.67318	-2.20132
C	2.35311	-2.98831	-1.50006
H	3.07743	-2.57171	-2.21396
H	2.92064	-3.62776	-0.80905
H	1.65310	-3.62697	-2.04515
H	-5.33763	-1.38363	1.36262
H	4.50112	0.21905	2.42234
N	-3.80326	-2.07222	0.16390

C	3.38035	-0.84287	0.95905
H	3.81127	-1.81375	1.18707
N	3.32974	1.55604	1.31625

SCF energy: -803.5089787 hartree
 zero-point correction: +0.266651 hartree
 enthalpy correction: +0.285277 hartree
 free energy correction: +0.219136 hartree
 quasiharmonic free energy correction: +0.223144 hartree

12-model

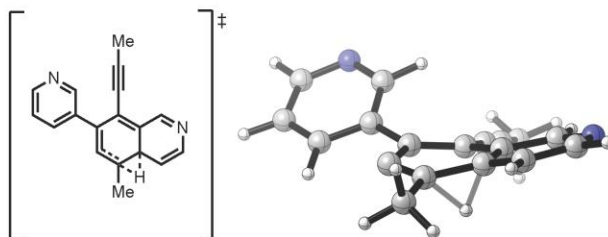


C	-4.60823	-0.83132	0.59379
C	-4.03948	-1.81511	-0.21530
C	-2.67544	-1.75081	-0.48132
C	-2.59490	0.21222	0.88437
C	-0.45001	-0.65756	-0.17992
C	0.26463	0.57983	-0.21489
C	1.66810	0.55763	0.01976
C	0.21261	-1.89242	-0.32854
C	1.60553	-1.95407	-0.34099
C	2.39787	-0.69818	-0.02655
C	4.35817	0.31325	0.63311
C	3.75814	1.60617	0.63303
C	2.42994	1.72130	0.31100
H	1.92964	2.68445	0.30906
H	4.35430	2.47627	0.88891
H	2.20512	-1.08739	-1.19375
H	-2.04741	1.01183	1.37616
H	-4.65402	-2.61034	-0.62787
H	-2.17012	-2.49319	-1.09032
C	-1.91414	-0.70337	0.06036
C	-0.90380	2.91082	-0.63492
C	-0.37223	1.84092	-0.41769

C	-1.59405	4.17186	-0.87955
H	-2.63425	4.12146	-0.53613
H	-1.10520	5.00299	-0.35734
H	-1.60665	4.41371	-1.94962
C	2.37211	-3.23791	-0.57122
H	3.35192	-3.07455	-1.03006
H	2.54837	-3.73637	0.38989
H	1.77074	-3.90783	-1.19107
H	-5.67504	-0.84235	0.81324
H	5.40854	0.21207	0.90691
N	3.73720	-0.79359	0.31045
N	-3.90362	0.16355	1.15024

SCF energy: -803.4921869 hartree
 zero-point correction: +0.263518 hartree
 enthalpy correction: +0.281690 hartree
 free energy correction: +0.217089 hartree
 quasiharmonic free energy correction: +0.220366 hartree

12'-model



C	-4.60313	-0.80148	0.63445
C	-4.05477	-1.78778	-0.18564
C	-2.69366	-1.73624	-0.46860
C	-2.57573	0.22033	0.90450
C	-0.45410	-0.66599	-0.18672
C	0.27523	0.56593	-0.22471
C	1.67551	0.52755	0.01991
C	0.20024	-1.89842	-0.35928
C	1.59257	-1.97003	-0.37612
C	2.39248	-0.73291	-0.02125
C	4.35295	0.48403	0.71073
C	2.44213	1.69387	0.32331

H	1.93785	2.65801	0.30503
H	2.22017	-1.08775	-1.20091
H	-2.01390	1.01288	1.39175
H	-4.68276	-2.57434	-0.59448
H	-2.20512	-2.47975	-1.08981
C	-1.91493	-0.69967	0.06905
C	-0.88749	2.89165	-0.67817
C	-0.35262	1.82687	-0.44486
C	-1.56924	4.15321	-0.94018
H	-1.63851	4.34521	-2.01813
H	-2.58902	4.13886	-0.53731
H	-1.03552	4.99641	-0.48608
C	2.33310	-3.26347	-0.64859
H	3.28207	-3.11961	-1.18019
H	2.55899	-3.77837	0.29438
H	1.68922	-3.92169	-1.23664
H	-5.66719	-0.80216	0.86686
H	5.39222	0.52475	1.02900
N	-3.88143	0.18406	1.18560
C	3.77534	-0.71130	0.38739
H	4.34335	-1.63413	0.43312
N	3.71070	1.69378	0.66186

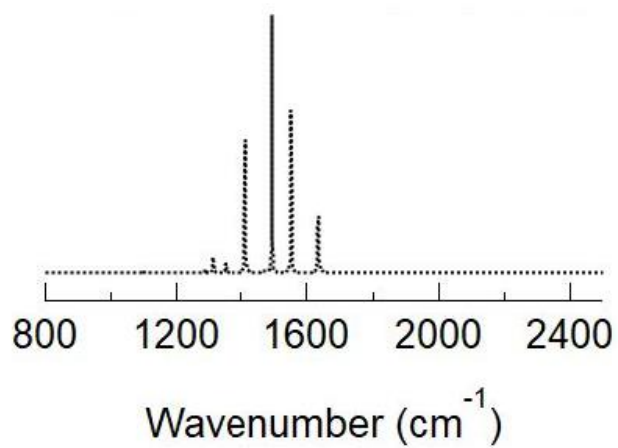
SCF energy: -803.4890542 hartree
 zero-point correction: +0.263436 hartree
 enthalpy correction: +0.281605 hartree
 free energy correction: +0.216966 hartree
 quasiharmonic free energy correction: +0.220398 hartree

4.6. Calculation of the Electronic Density of States (DOS) for **1d** and **2d**

Periodic DFT calculations were applied for the calculation of the electronic Density of States (DOS) of **1d** and **2d** and the Raman and IR spectra of **2d**. The calculations were performed using the Vienna ab initio simulation package (VASP).^{15,16,17} We applied the Projector Augmented Wave (PAW) method to represent interactions between core and valence electrons,^{18,19} in conjunction with the PBE functional.²⁰ For geometry optimizations, we applied large kinetic energy cutoffs of 500.0 eV and kpoint meshes of 8 points in the non-periodic direction for both systems. A convergence criterion of 10^{-6} eV was set for the energy, while the geometric relaxation was performed until forces on atoms were smaller than 0.01 eV/Å. Additionally, we selected a Gaussian smearing with the sigma value of 0.1 eV. Dispersion interactions were accounted for by means of the D3BJ dispersion correction scheme.^{21,22} The optimized cell parameters in the non-periodic direction for **1d** and **2d** were 8.598 Å and 4.256 Å, respectively, and we added a vacuum of ~ 10 Å in the non-periodic directions (the geometrical coordinates of both structures are provided below).

For DOS calculations, we increased the kinetic energy cutoffs to 600 eV and the kpoint meshes to 16 and 30 points in the non-periodic direction for **1d** and **2d**, respectively. In this regard, note that the unit cell of **1d** is bigger than that of **2d**, and that is why it requires a lower number of kpoints. Note also that we obtained the same results (geometry and DOS) when spin polarized calculations at the PBE level were performed for **2d**. In order to obtain more reliable band gaps, we also applied the HSE06 hybrid functional for **2d**.²³ The Raman and IR spectra of **2d** (Figure S36) were modelled by means of an interplay between VASP,^{24,25,26} and Phonopy-Spectroscopy.^{27,28} For these calculations, we increased the energy cutoff up to 800 eV and the energy convergence criteria to 10^{-8} eV. We used a $4 \times 1 \times 1$ supercell for the calculation of the second order force constants and a $2 \times 1 \times 1$ one for the third order force constants.

a)



b)

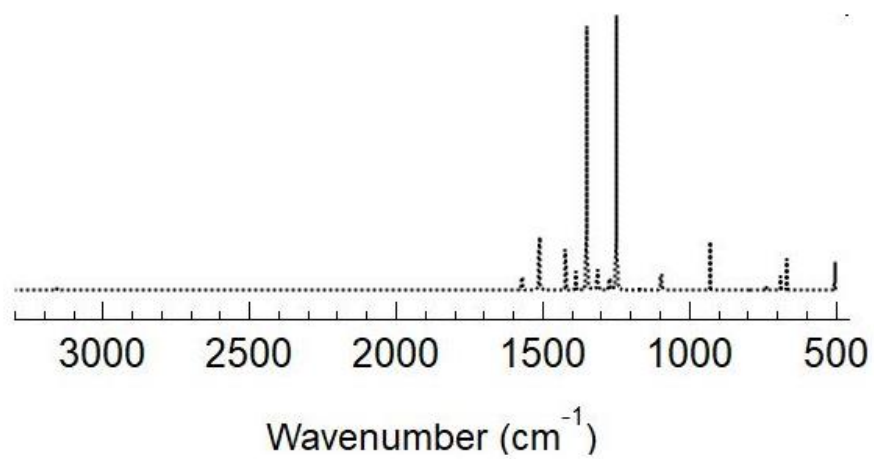


Figure S36. Calculated (a) Raman and (b) IR spectra for fully cyclized GNR **2d**.

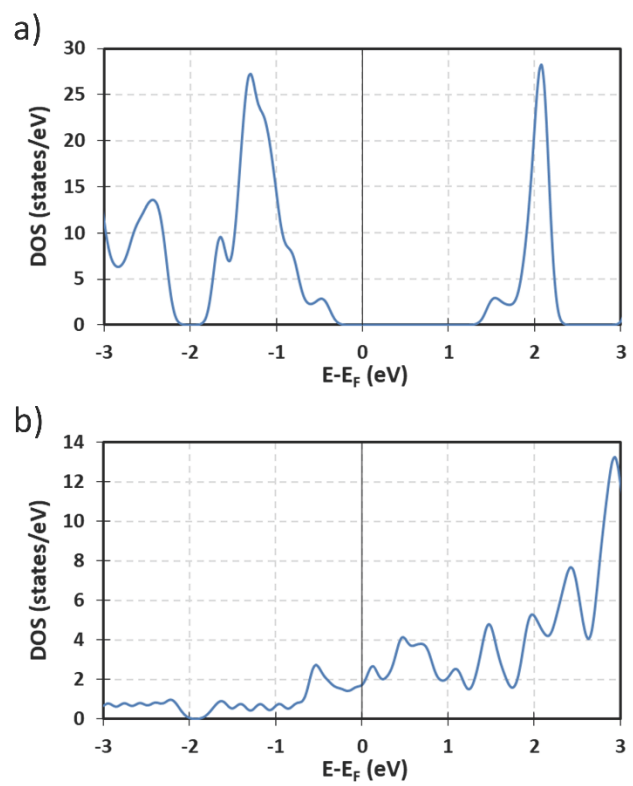


Figure S37. DOS plot calculated at the PBE level for (a) **1d** and (b) **2d**.

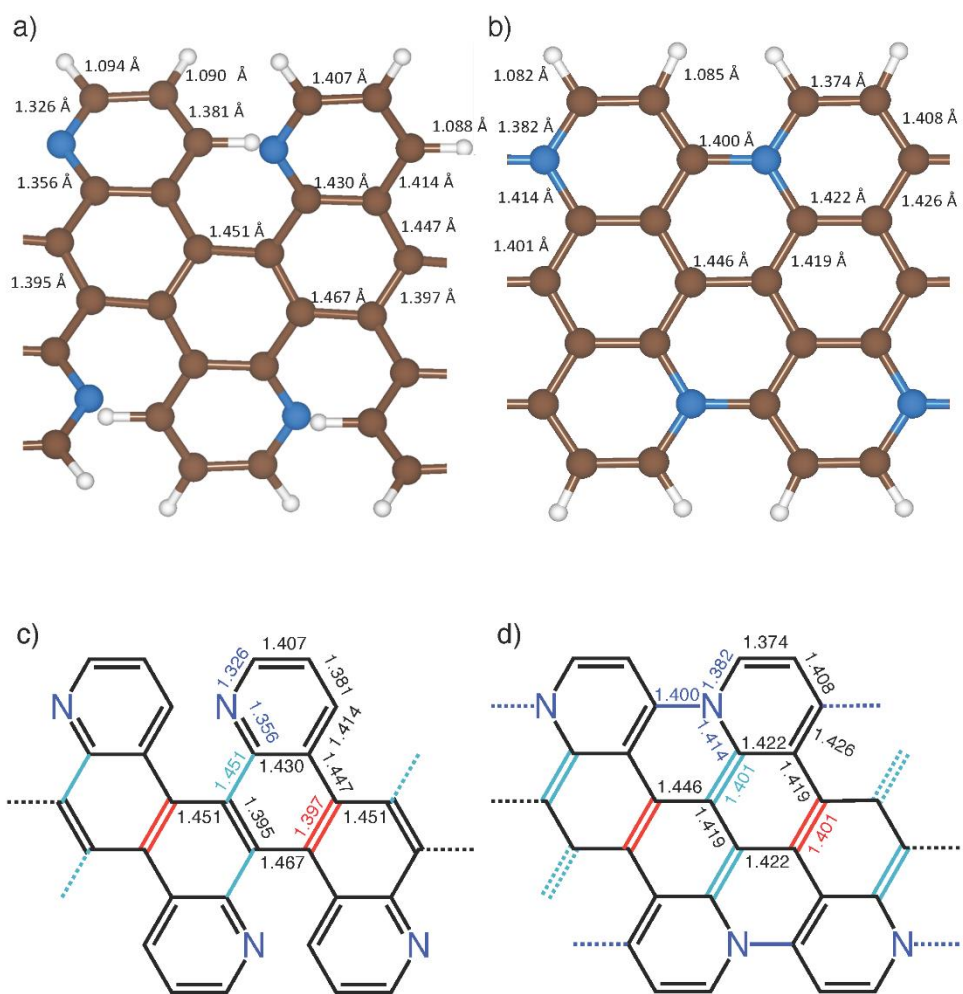


Figure S38. Calculated bond lengths for the optimized geometries of (a) **1d** and (b) **2d** using the PBE functional. The corresponding ChemDraw representations of these structures and bond lengths are shown in (c) for **1d** and (d) for **2d**.

Geometrical coordinates for fjord-edge N₂[8]GNR 1d and N₂[8]_AGNR 2d

Fjord-edge N₂[8]GNR **1d**:

	1.0	
25.1674594879	0.0000000000	0.0000000000
-0.0028574353	24.5526026017	0.0000000000
-1.7866384065	-0.0516127589	8.4106562111
	C N H	
	28 4 12	
	Direct	
0.432421744	0.436544806	0.130215734
0.565167308	0.496146679	0.280019552
0.663246393	0.516243160	0.502604544
0.710398793	0.536950111	0.604733646
0.752062082	0.558341265	0.541746259
0.746341586	0.559094608	0.375510603
0.661757231	0.518798232	0.335679650
0.613547027	0.499666303	0.225604787
0.516956151	0.476942003	0.170065403
0.438716829	0.583686471	0.473754197
0.444150418	0.582994580	0.639923036
0.521173477	0.529792130	0.673527002
0.519459307	0.527819097	0.506353080
0.476467758	0.556063473	0.407515168
0.565308511	0.506441236	0.448381752
0.614526510	0.502428114	0.558216035
0.565725267	0.503263175	0.780471802
0.564242005	0.493083298	0.947504759
0.613479912	0.489301831	0.057345603
0.659342349	0.467901230	0.999367714
0.515579462	0.479170710	0.003164211
0.702486575	0.439847261	0.098194897
0.740034699	0.411928475	0.031829599
0.734260976	0.412122637	0.865548432
0.657443404	0.465617180	0.832127154
0.468641460	0.458040476	0.901049972
0.426945090	0.436736971	0.964022040
0.612932801	0.492181331	0.725111425
0.475797296	0.455692947	0.230321556
0.702827394	0.540223718	0.275404453
0.483916402	0.557231724	0.737037122

0.694449008	0.437844485	0.768487275
0.400389344	0.419359416	0.183844656
0.712967098	0.537856281	0.732580066
0.778301001	0.576445520	0.321878791
0.415272862	0.605017364	0.696182787
0.474360406	0.557705700	0.279991388
0.762899697	0.389768362	0.809185982
0.466330081	0.456737578	0.773338914
0.704898059	0.438544333	0.225896746
0.405848235	0.606435001	0.399512589
0.390852809	0.419504434	0.887089968
0.772990406	0.389300257	0.106117986
0.788323104	0.575224161	0.618747175

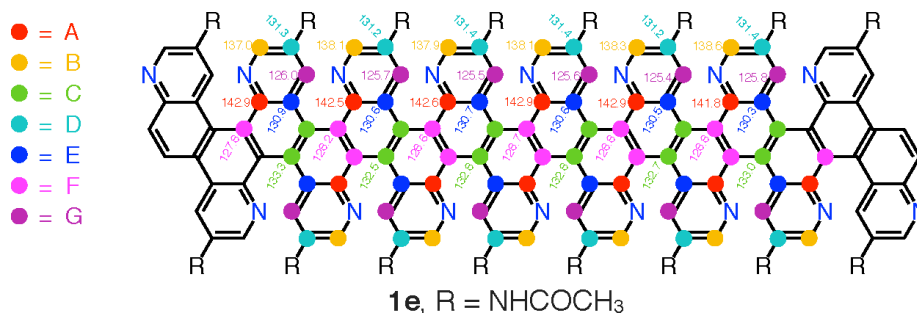
Fully fused N₂[8]_AGNR **2d**:

	1.0	
4.2563900948	0.0000000000	0.0000000000
0.0000000000	25.0000000000	0.0000000000
0.0000000000	0.0000000000	16.0000000000
	C N H	
	12 2 4	
	Direct	
0.377191484	0.354893386	0.498291105
0.876676321	0.401480317	0.498291105
0.372411758	0.450767487	0.498291105
0.699971199	0.353889853	0.498291105
0.706492066	0.450597167	0.498291105
0.207790837	0.499313742	0.498291105
0.369450420	0.548967838	0.498291105
0.199266180	0.598084629	0.498291105
0.703530669	0.548797488	0.498291105
0.868151665	0.500251234	0.498291105
0.375971317	0.645675123	0.498291105
0.698750973	0.644671619	0.498291105

4.7. Calculation of the ^{13}C NMR Chemical Shifts on Model Compounds

1) Fjord edge $\text{N}_2[8]_{\text{A}}\text{GNR}$ **1e** (Hopf cyclizations at the 6-pyridyl position, zigzag conformer (see page S145 for calculations on the conformational preference for **1e**, and also Table 1 and the main text for chemical shift analysis):

a)



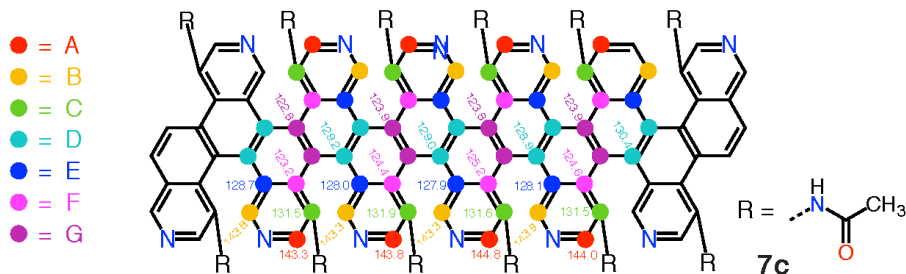
b)

δ (ppm)	C=O	A	B	C	D	E	F	G
	165.3	142.9	137.0	133.3	131.3	130.9	127.8	126.0
	165.1	142.5	138.1	132.5	131.2	130.6	128.2	125.7
	165.1	142.6	137.9	132.8	131.4	130.7	128.6	125.5
	165.2	142.9	138.1	132.8	131.4	130.6	128.7	125.6
	165.0	142.9	138.3	132.7	131.2	130.5	128.8	125.4
	165.1	141.8	138.6	133.0	131.4	130.3	128.8	125.8
Average:	165.1	142.6	138.0	132.9	131.3	130.6	128.5	125.7

Figure S39. a) Chemical structure of the model compound **1e** used for the calculation of fjord edge $\text{N}_2[8]_{\text{A}}\text{GNR}$ **1** ($\text{R} = \text{NHCOCH}_3$), showing ^{13}C NMR chemical shifts (δ , ppm) calculated at the B3LYP//6-31G* level of theory. NMR shielding tensors were computed with the Gauge-Independent Atomic Orbital (GIAO) method. The calculation was performed with the Spartan 18 package.²⁹ b) Table of the ^{13}C NMR chemical shifts obtained in the B3LYP//6-31G(d) calculation for model compound **1e**, showing also the average values reported in Table 1. However, only those aryl and carbonyl carbons chemical shifts for the six-inner chrysene-type units are included (the chemical shifts of carbons in the two chrysene units at each end of the model structure have some outliers that we did not deem representative of an infinite structure).

2) Fjord edge $N_2[8]_A$ GNR model **7c** (Hopf cyclizations at the 4-pyridyl position, ^{13}C NMR chemical shifts calculated on zigzag conformer):

a)



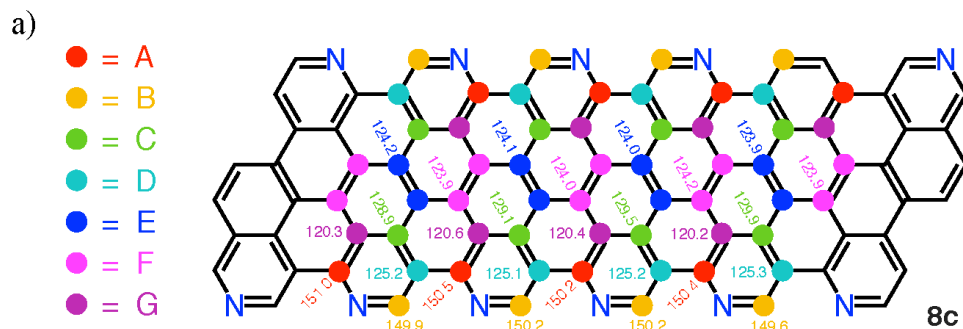
b)

δ (ppm)	C=O	A	B	C	E	E	F	G
		143.3	143.8	131.5	129.2	128.7	123.2	122.8
		143.8	143.3	131.9	129.0	128.0	124.4	123.9
		144.8	143.3	131.6	128.9	127.9	125.2	123.8
		144.0	143.9	131.5	130.4	128.1	124.6	123.9
Average:		144.0	143.6	131.6	129.4	128.2	124.4	123.6

Figure S40. a) Chemical structure of the model compound (**7c**) used for the calculation of fjord edge $N_2[8]_A$ GNR **7** (Figure 5), showing most of the ^{13}C NMR chemical shifts (δ , ppm) calculated at the B3LYP//6-31G* level of theory. NMR shielding tensors were computed with the Gauge-Independent Atomic Orbital (GIAO) method. The calculation was performed with the Spartan 18 package.²⁹ b) Table of the ^{13}C NMR chemical shifts obtained in the B3LYP//6-31G(d) calculation for model compound **7c**, showing also the average values reported in Table 1. However, only those aryl and carbonyl carbons chemical shifts for the four-inner chrysene-type units are included (the chemical shifts of carbons in the two chrysene units at each end of the model structure have some outliers that we did not deem representative of an infinite structure).

Note that the ranges of chemical shifts are very different from those for the model of fjord edge $N_2[8]_A$ GNR **1** in Figure S39 and Table 1, with a group of two peaks at ~ 144 ppm, and a group of five peaks between 132–124 ppm (see also Figure S42).

3) Fully graphitized $N_2[8]_A$ GNR model **8** (aromatization after Hopf cyclizations at the 4-position):



b)

δ (ppm)	A	B	C	D	E	F	G
	151.0	149.9	128.9	125.2	124.2	123.9	120.3
	150.5	150.2	129.1	125.1	124.1	124.0	120.6
	150.2	150.2	129.5	125.2	124.0	124.2	120.4
	149.6	149.6	129.9	125.3	123.9	123.9	120.2
Average:	150.3	150.0	129.4	125.2	124.1	124.0	120.4

Figure S41. a) Chemical structure of the model compound (**8c**) used for the calculation of fjord edge $N_2[8]_A$ GNR **8** (Figure 5), showing most of the ^{13}C NMR chemical shifts (δ , ppm) calculated at the B3LYP//6-31G* level of theory. NMR shielding tensors were computed with the Gauge-Independent Atomic Orbital (GIAO) method. The calculation was performed with the Spartan 18 package.²⁹ b) Table of the ^{13}C NMR chemical shifts obtained in the B3LYP//6-31G(d) calculation for model compound **8c**, showing also the average values reported in Table 1. However, only those aryl and carbonyl carbons chemical shifts for the four-inner chrysene-type units are included (the chemical shifts of carbons in the two chrysene units at each end of the model structure have some outliers that we did not deem representative of an infinite structure).

Note that the ranges of chemical shifts are very different from those for the model of fjord edge $N_2[8]_A$ GNR **1** in Figure S39 and Table 1, with a group of two peaks at ~ 150 ppm, and a group of five averaged peaks between 130–120 ppm (see also Figure S42).

4) *Peak Fitting Methodology for the Curve-Fitted Experimental CP/MAS ^{13}C NMR Spectrum of fjord-edge $\text{N}_2[8]\text{GNR}$ **1b** (see Table 1 in main text):*

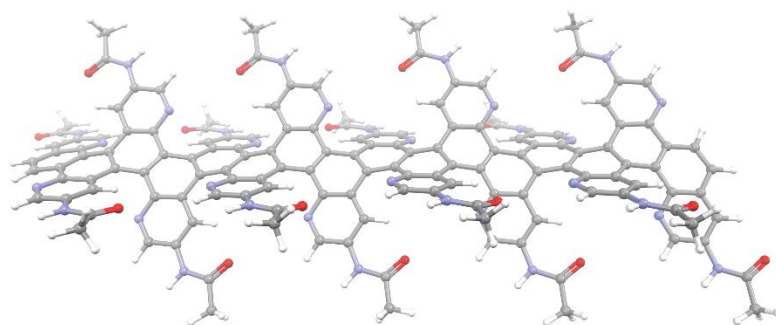
Peak fitting was performed using the Multippeak Fitting package in Igor Pro 8.04. The experimental CP/MAS ^{13}C NMR spectrum of fjord edge $\text{N}_2[8]\text{GNR}$ **1b** (generated from PDA **4b** heating at 400 °C for 1 h) was fit to 7 peaks, which correspond to the expected number of non-symmetrically related carbons for an infinite structure of fjord edge $\text{N}_2[8]\text{GNR}$ **1b**, over the aromatic region using a Gaussian line shape. As the fjord edge $\text{N}_2[8]\text{GNR}$ **1b** seven distinct carbons should have equal weighting in the aromatic region, the amplitude and width of each peak fit were both held constant for the fitting process. The results show seven distinct functions whose centers are taken as the chemical shifts of the carbons in question.

4.8. Conformational Preference Calculations on Model Compound **1e**.

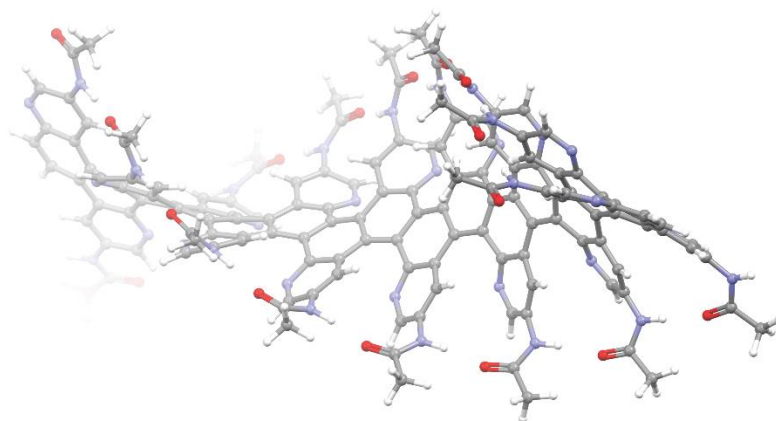
The conformational isomerism of model compound **1e** (model system for the fjord-edge N₂[8]GNR **1**, R = -NHCOMe) was investigated for its two extreme cases, zigzag and helical. Since **1e** has a fjord edge structure, there is strong repulsion between the nitrogen lone pairs and the C-H bonds of adjacent diazachrysene units. Thus, the two possible key conformations, zigzag (alternating up-down pyridyl units) and helical (all pyridyl units staggered in a non-alternating fashion) were calculated. Because the differences in overall energy between both conformers is very high, it is likely that only the zigzag conformation exists in the fjord-edge N₂[8]GNRs **1**.

Both structures were fully optimized at the PM3 and B3LYP/6-31G(d) levels of theory within the Spartan '18 Parallel Suite (Wavefunction, Inc).²⁹ The semiempirical PM3 level was carried out in the gas phase, while the DFT level (B3LYP) was carried out with a polar solvent parameter (DMSO) to counteract significant energetic contributions from intramolecular hydrogen bonding on the overall conformational energy.

The calculation results on the next page show that the zigzag conformation (**1e** zigzag) of model compound **1e** should be strongly preferred over the helical structure (**1e** helical). However, we note that any intra- and intermolecular π - π stacking and H-bonding contributions that must be operating in these strongly aggregated fjord-edge N₂[8]GNRs (**1a,b**, see TEM imaging section) are not accounted for by these calculations.



1e (zigzag)



1e (helical)

Method	Rel. Energy (kcal/mol)
PM3	
(gas phase)	
1e (zigzag)	0.0
1e (helical)	35.9
B3LYP//6-31G(d)	
(polar solvent)	
1e (zigzag)	0.0
1e (helical)	60.2

Section 5: References

1. Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
2. Sheldrick, G. M. (2008). *Acta Cryst.* **A64**, 112–122.
3. Sheldrick, G. M. (2015). *Acta Cryst.* **A71**, 3–8.
4. Sheldrick, G. M. (2015). *Acta Cryst.* **C71**, 3–8.
5. Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
6. Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
7. Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.
8. Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200–1211.
9. Zhao, Y.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2813–2818.
10. Ribeiro, R. F.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2011**, *115*, 14556–14562.
11. Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215.
12. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision B.01*; Gaussian, Inc.: Wallingford, CT, 2009.
13. Legault, C. Y. *CYLview, 1.0 b*; Université de Sherbrooke (<http://www.cylview.org>), 2009.
14. Watts, K. S.; Dalal, P.; Tebben, A. J.; Cheney, D. L.; Shelley, J. C. *J. Chem. Inf. Model.* **2014**, *54*, 2680.

15. Kresse, G.; Hafner, J. Ab Initio Molecular Dynamics for Liquid Metals. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1993**, *47*, 558–561. <https://doi.org/10.1103/PhysRevB.47.558>
16. Kresse, G.; Hafner, J. Ab Initio Molecular-Dynamics Simulation of the Liquid-Metal-Amorphous-Semiconductor Transition in Germanium. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1994**, *49*, 14251–14269. <https://doi.org/10.1103/PhysRevB.49.14251>
17. Kresse, G.; Furthmüller, J. Efficient Iterative Schemes for ab initio Total-Energy Calculations Using a Plane-Wave Basis Set. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 11169–11186. <https://doi.org/10.1103/PhysRevB.54.11169>
18. Blöchl, P. E. Projector Augmented-Wave Method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1994**, *50*, 17953–17979. <https://doi.org/10.1103/PhysRevB.50.17953>
19. Kresse, G.; Joubert, D. From Ultrasoft Pseudopotentials to the Projector Augmented-Wave Method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999**, *59*, 1758–1775. <https://doi.org/10.1103/PhysRevB.59.1758>
20. Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868. <https://doi.org/10.1103/PhysRevLett.77.3865>
21. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate ab initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104–154119. <https://doi.org/10.1063/1.3382344>
22. Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465. <https://doi.org/10.1002/jcc.21759>
23. Krukau, A. V.; Vydrov, O. A.; Izmaylov, A. F.; Scuseria, G. E. Influence of the exchange screening parameter on the performance of screened hybrid functionals. *J. Chem. Phys.* **2006**, *125*, 224106. <https://doi.org/10.1063/1.2404663>
24. Togo, A.; Tanaka, I. First principles phonon calculations in materials science. *Scr. Mater.* **2015**, *108*, 1–5. <https://doi.org/10.1016/j.scriptamat.2015.07.021>
25. Togo, A.; Chaput, L.; Tanaka, I. Distributions of phonon lifetimes in Brillouin zones. *Phys. Rev. B* **2015**, *91*, 094306. <https://doi.org/10.1103/PhysRevB.91.094306>
26. <https://github.com/phonopy>

27. Skelton, J. M.; Burton, L. A.; Jackson, A. J.; Oba, F.; Parker, S. C.; Walsh, A. Lattice dynamics of the tin sulphides SnS₂, SnS and Sn₂S₃: vibrational spectra and thermal transport. *Phys. Chem. Chem. Phys.* **2017**, *19*, 12452–12465. <https://doi.org/10.1039/C7CP01680H>
28. <https://github.com/JMSkelton/Phonopy-Spectroscopy>
29. Spartan, Wavefunction, Inc. Irvine, CA. Except for molecular mechanics and semi-empirical models, the calculation methods used in Spartan have been documented in: Y. Shao, L.F. Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S.T. Brown, A.T.B. Gilbert, L.V. Slipchenko, S.V. Levchenko, D.P. O'Neill, R.A. DiStasio Jr., R.C. Lochan, T. Wang, G.J.O. Beran, N.A. Besley, J.M. Herbert, C.Y. Lin, T. Van Voorhis, S.H. Chien, A. Sodt, R.P. Steele, V.A. Rassolov, P.E. Maslen, P.P. Korambath, R.D. Adamson, B. Austin, J. Baker, E.F.C. Byrd, H. Dachsel, R.J. Doerksen, A. Dreuw, B.D. Dunietz, A.D. Dutoi, T.R. Furlani, S.R. Gwaltney, A. Heyden, S. Hirata, C-P. Hsu, G. Kedziora, R.Z. Khalliulin, P. Klunzinger, A.M. Lee, M.S. Lee, W.Z. Liang, I. Lotan, N. Nair, B. Peters, E.I. Proynov, P.A. Pieniazek, Y.M. Rhee, J. Ritchie, E. Rosta, C.D. Sherrill, A.C. Simmonett, J.E. Subotnik, H.L. Woodcock III, W. Zhang, A.T. Bell, A.K. Chakraborty, D.M. Chipman, F.J. Keil, A. Warshel, W.J. Hehre, H.F. Schaefer, J. Kong, A.I. Krylov, P.M.W. Gill and M. Head-Gordon, *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172–3191.

Appendix C Preface

Benzylidenebutyrolactone crystals are known to undergo single-crystal-to-single-crystal photodimerization when exposed to UV irradiation. Moreover, the induced topotactic reaction produces mechanical responses or photosalience in these crystals. As stated in Chapter 1, the field of crystallography exploits the signal amplification that results from the physical alignment of molecules into ordered lattices. As a result, crystallographers traditionally strive to grow crystals as large as possible in order to maximize the benefit of this physical alignment. In this work, we forgo these tendencies in order to investigate the crystal size threshold necessary for benzylidenebutyrolactone crystals to demonstrate photosalient effects when irradiated. Our experiments successfully demonstrated the utility of electron diffraction for studying crystalline behavior at scales X-ray diffraction cannot.

Appendix C

Scale-Dependent Photosaliency and Topotactic Reaction of Microcrystalline Benzylidenebutyrolactone Determined by Electron Microscopy and Electron Diffraction

The following is a reprint of a research article from

Crystal Growth and Design

22(3), 1533-1537 (2022)

DOI: 10.1021/acs.cgd.1c01378

Reprinted by permission from American Chemical Society.

© 2022 American Chemical Society

Scale-Dependent Photosalience and Topotactic Reaction of Microcrystalline Benzylidenebutyrolactone Determined by Electron Microscopy and Electron Diffraction

Vince M. Hipwell, Chih-Te Zee, Jose Rodriguez,* and Miguel A. Garcia-Garibay*

Cite This: *Cryst. Growth Des.* 2022, 22, 1533–1537

Read Online

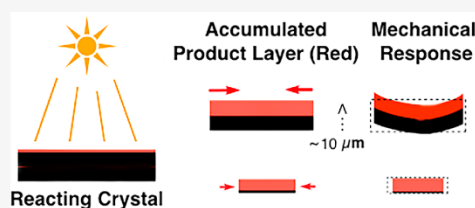
ACCESS |

Metrics & More

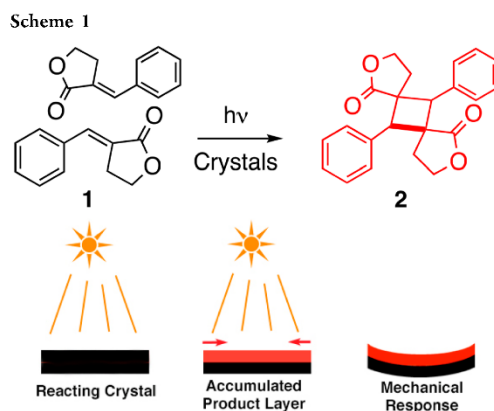
Article Recommendations

Supporting Information

ABSTRACT: The macroscopic motion of crystals induced by a solid-state photochemical reaction, also known as photosalience, is of interest for the development of micromechanical actuators and crystalline molecular machines. Using microcrystal electron diffraction, we report evidence for a minimum crystal size threshold below which photosalience is not observed for benzylidenebutyrolactone. We confirm our observations by solving the crystal structure of micron-scale specimens before and after their topotactic single crystal-to-single crystal reaction and by collecting transmission electron microscopy images that reveal the absence of photosalient effects for crystals that are below ca. 10 μm in size. These results indicate that photosalience depends not only on the ability of crystals to support a photochemical reaction but also on a size threshold where the accumulated product phase can form a layer that is separate from that of the reactant, such that their different packing dimensions are able to transduce the collective strain accumulated at the molecular-level boundaries into the macroscopic motion that propels the entire crystal specimen.



Photosalience is an interesting, emergent phenomenon that involves the macroscopic motion of crystals.¹ It occurs as a result of photochemical reactions in their crystalline phase,^{2,3} although recent results suggest that photothermal effects can also play a small role under some circumstances.⁴ Photosalience is thought to be the result of the collective or sudden release of strain that is built up between the grain boundaries of segregated reactant and product crystalline domains, along with more complex and subtle processes.^{1–7} As shown in Scheme 1, photosalience can be illustrated in a simple manner by considering a single crystal exposed to light that causes a chemical reaction up to a certain sample depth. Photosalience occurs when changes in volume between the reactant and product are manifested in the form of a relative expansion or contraction that leads to generation of mechanical stress that is eventually released in the form of macroscopic motion. In this paper, we investigate the well-established photosalience of crystalline benzylidenebutyrolactone **1** (Scheme 1) in order to determine whether this phenomenon has a size threshold in the nanometer or submicron scale through a combination of microelectron diffraction (microED) and TEM imaging.⁸ Our results indicate that crystals of all sizes facilitate the SCSC photodimerization reaction shown in Scheme 1, but only crystals beyond a certain size are able to exhibit photosalience. We conclude that photochemical reactions in nanocrystals with sizes of the same order of magnitude as wavelength of light occur in a relatively homogeneous manner,⁹ and we discuss the



Received: November 23, 2021

Revised: February 1, 2022

Published: February 7, 2022



size limit at which photosalience occurs as well as a hypothesis for the change in behavior.

Photosalience in crystals of **1** has been studied in great detail by Naumov and co-workers using optical microscopy and high-speed cameras to catalogue different types of motion and their relation to the size and aspect ratios of the crystals.¹⁰ The crystals analyzed by this method had sizes of 20–240 μm along their shortest dimension and 0.19–2.8 mm along the longest one. While it may be possible to investigate even smaller crystals with optical microscopy, the lower limit is fundamentally restricted by the maximum magnification achievable with this method. Thus, in order to establish the existence or lack of photosalience in nanocrystals and small microcrystals of **1**, we turned to the use of electron microscopy.

Compound **1** was synthesized as reported in the literature and photochemical experiments carried out with polycrystalline samples showed that it undergoes the expected quantitative dimerization reaction to the trans-head-to-tail cyclobutane **2** (Scheme 1 and Figures S3–S4). Powder X-ray diffraction (PXRD) patterns obtained before and after irradiation and comparison with those simulated from the single-crystal structure data of **1** and **2** confirmed the identity of the crystal phases as those of the previously analyzed crystal forms (Figure S5).¹⁰

Samples for microED experiments were prepared as aqueous suspensions using the reprecipitation method.¹¹ A stock solution of 30 mg/mL (40 μL) of **1** in acetonitrile was added to 4.0 mL of rapidly vortexing Millipore water by micropipette. Multiple droplets of these suspensions were drop-cast onto a TEM grid and then wicked away with filter paper held parallel to the surface of the grid on the outer edge of the suspension droplet. TEM imaging of grids prepared by this method revealed relatively sparse loading of crystalline material. Despite this, a sufficient number of crystals were found to enable structure solution.

Diffraction data was collected on a Tecnai TF30 microscope from multiple microcrystals before and after UV-irradiation using a 450 W Hanovia medium-pressure mercury lamp. Following data collection on the crystals of **1**, the sample holder was removed and transferred to the photochemical reactor without removing the grid. The crystals were irradiated for 10 min in the Hanovia reactor. The sample holder was then replaced in the TEM system and post-irradiation diffraction data was collected from the same crystals based on positions saved on the TEM computer system.

Crystal structures of **1** and **2** were determined from the same set of five crystals before and after UV irradiation, as shown in Figure 1 along with images of a subset of these crystals. Approximate sizes can be estimated for these crystals in relation to the red dotted circles, which denote the size and position of the $\sim 5 \mu\text{m}$ selected area aperture. Notably, these images show that each crystal remained in the same position and orientation before and after irradiation, suggesting that the photosalient effects previously observed using an optical microscope with larger crystals of this compound were not active in crystals of this size.

Notably, NMR analysis of similar aqueous suspensions of **1** before and after UV irradiation revealed that the topotactic reaction to form **2** constituted a very small portion (<1%) of the observed photochemical reactivity. The dominant reaction pathway in this case was the (E) to (Z) isomerization to **1**, as determined by ¹H NMR analysis following extraction of the

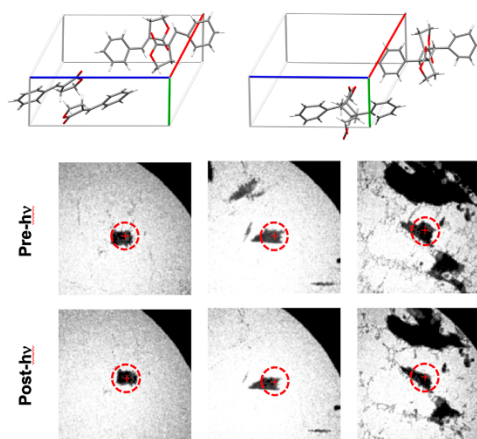


Figure 1. (Top) Crystal structures of **1** and **2** determined by microED from a set of five crystals before and after UV irradiation. (Bottom) TEM images of three of these crystals. The red dotted circles indicate the position and size of the $\sim 5 \mu\text{m}$ selected area aperture.

irradiated suspension with DCM (Figure S6). This observation suggested significant water solubility for compound **1**, which was demonstrated by ¹H NMR using an internal standard. We discovered that suspensions prepared in this manner contain up to 47% of **1** in equilibrium in solution (Figure S7). While this observation does not affect the topotactic reaction and electron diffraction analysis shown in Figure 1, it has important implications on the observed lack of photosalience, as it may be possible that crystals deposited from aqueous suspension adhered to the surface of the TEM grid by solubilized and subsequently dried material that may act as glue.¹² To test this possibility, we prepared additional samples by direct loading of polycrystalline powders. TEM grids were gently dropped and submerged in the powder of **1** in a sample vial and then removed with tweezers. Excess powder was removed by tapping the tweezers on the edge of the vial.

Following the same diffraction data collection procedure described above, crystal structures were solved before and after irradiation of a single crystal (Figure 2). We speculate that the necessity to combine data from multiple crystal specimens in the case of the suspension samples may be due to the water-solubility of **1**, which may lead to a higher percentage of crystal defects, twinning, or potential deposit of amorphous phases.

For the next set of experiments, we collected images of microcrystals from the powder sample before and after irradiation. TEM grids were loaded in the same manner as described above and then screened in diffraction mode to find suitable crystals. For smaller crystals, a single diffraction pattern at 0° tilt angle was collected to verify crystallinity. This was not possible for larger crystals that were too thick to enable adequate penetration of the electron beam, but the images display clearly faceted specimens. Furthermore, knowing from the bulk solid irradiation that >99% of the powder sample is indeed crystalline, we assumed that any material deposited and imaged on the TEM grid was also crystalline. High-resolution real-space images were obtained for crystals ranging in size from ca. 400 nm to >100 μm . After the standard UV irradiation procedure, a second set of images was taken at the

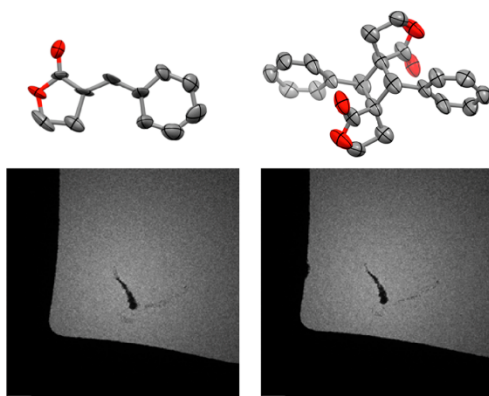


Figure 2. (Top) Single-crystal electron diffraction structures of **1** and **2**. (Bottom) Low-resolution images of the single microcrystal before and after irradiation obtained in diffraction mode.

same locations on the TEM grid. Representative pre- and post-irradiation images of each crystal or set of crystals are shown in Figure 3. The complete set of images can be found in Figures

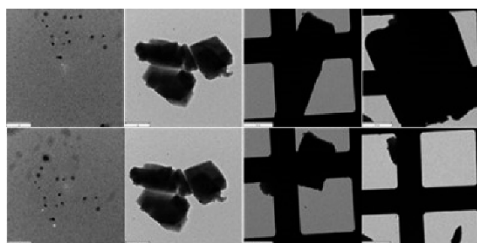


Figure 3. Representative images of nanocrystals and microcrystals before (top) and after (bottom) irradiation. The scale bar for the first two images from the left is 1 μm and that for the last two images is 30 μm .

S8 and S9. The results shown in Figure 3 confirm that there is a size-dependent threshold for the observation of the photosalient effect. All the nanocrystals and the small microcrystals were found to be in the exact same position and orientation after irradiation, whereas a majority of crystals $\sim 100 \mu\text{m}$ or longer along their largest dimension had either fractured or moved out of the frame entirely.

To verify that the motion of the larger crystals was not induced by the handling and movement of the grid holder, a control experiment was conducted in which images were taken before and after handling the grid holder in the same manner but without irradiation. These images can be found in the SI, and they show that crystals in the same size range remain in the same positions.

With this relatively limited set of crystal specimens, an exact description of the shape and size of crystals of **1** at which photosalience begins to occur cannot yet be established. Notably, it has been shown that crystals of other compounds which are submicron in at least one dimension, such as crystalline nanorods, still exhibit a bending motion.^{13–16} Consideration of the aspect ratios of the crystals at these

smaller length scales may also be a factor which controls photosalience. Therefore, it should be noted that there may be a dependence based on the third dimension normal to the plane of the TEM grid, which cannot be quantified from these images. Nonetheless, these results demonstrate a striking difference of behavior of crystals in different size regimes. It can be said that crystals of **1** with dimensions that range from nanocrystalline, or submicron in all dimensions, up to $\sim 50 \mu\text{m}$ in the largest dimension may not exhibit photosalience. As shown in Figure 4, SEM imaging of a powder sample of **1**

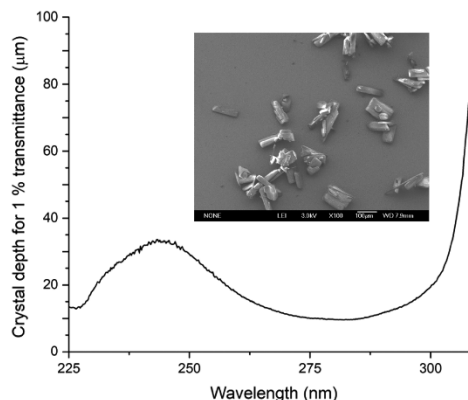


Figure 4. Light penetration depth corresponding to 1% transmittance as a function of wavelength for crystals of **1**. The inset is a SEM image from a polycrystalline powder viewed with 20° tilt to show the approximate crystal depth.

reveals the approximate depth of crystals that are similar in size to those which exhibited photosalience. It is interesting to note that this depth seems to be similar to that which would lead to almost complete absorption of incident photons. Taking the extinction coefficient of **1** and calculating its concentration in crystals, one can estimate that a specimen with a depth of ca. 10 μm would have a 1% transmittance at the λ_{max} of 282 nm, under the assumption that dependence on molecular orientation within the crystal (which leads to a bias in the average orientation of the transition dipole moment) has minimal impact. The UV–vis spectrum of **1** along with the calibration curve to determine the extinction coefficient can be found in the SI. The absorption spectrum and this calculation was also utilized to produce the plot in Figure 4, which shows a crystal depth corresponding to 1% transmittance as a function of wavelength.

This size-dependent threshold demonstrates that photosalience is an emergent phenomenon that relies not only on a particular arrangement of a photochemically active compound within a crystal but also on the mechanical and photochemical properties of the crystal itself. The smaller crystals may be more elastic, and therefore better able to accommodate any strain that is formed between reactant and product crystal grain boundaries. The diffraction associated with lattice deformation or ripples on the nanoscale has recently been associated with rigid peptide nanocrystals.¹⁷ Furthermore, the higher surface-area-to-volume ratio of the smaller crystals would promote more homogeneous irradiation throughout the crystal, which would mitigate the segregation of different crystal domains.¹⁸

In conclusion, we have demonstrated the use of microED to determine crystal structures from a single microcrystal before and after an SCSC reaction. In addition, a TEM imaging study established that photosalience is an emergent multiscale phenomenon with a size threshold. Given the renewed interest in mechanically responsive crystals for use as nano- and micromechanical actuators,^{19–23} this size-dependent behavior may place a lower limit on the size of mechanical devices that can be constructed with such crystals as actuating components. This is particularly true if this size dependence is a general effect observed in all mechanically responsive crystals.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.cgd.1c01378>.

Sample preparation and characterization including powder X-ray diffraction, TEM images, crystallographic data (PDF)

X-ray data for butyrolactone monomer **1** (CIF)

X-ray data for butyrolactone dimer **2** (CIF)

■ AUTHOR INFORMATION

Corresponding Authors

Miguel A. Garcia-Garibay – Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90024-1569, United States; orcid.org/0000-0002-6268-1943; Email: mgg@chem.ucla.edu

Jose Rodriguez – Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90024-1569, United States; orcid.org/0000-0002-0248-4964; Email: jrodriguez@mbi.ucla.edu

Authors

Vince M. Hipwell – Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90024-1569, United States

Chih-Te Zee – Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90024-1569, United States

Complete contact information is available at <https://pubs.acs.org/10.1021/acs.cgd.1c01378>

Funding

This work was supported by NSF grant CHE-1855342.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Support from the UCLA Graduate Division for V.M.H. and C.Z. with dissertation year fellowships is gratefully acknowledged.

■ REFERENCES

- (1) Naumov, P.; Chizhik, S.; Panda, M. K.; Nath, N. K.; Boldyreva, E. Mechanically Responsive Molecular Crystals. *Chem. Rev.* **2015**, *115*, 12440–12490.
- (2) Rath, B. B.; Vittal, J. J. Single-Crystal-to-Single-Crystal [2 + 2] Photocycloaddition Reaction in a Photosalient One-Dimensional Coordination Polymer of Pb(II). *J. Am. Chem. Soc.* **2020**, *142*, 20117–20123.

- (3) Zhu, L.; Al-Kaysi, R. O.; Bardeen, C. J. Photoinduced Ratchet-Like Rotational Motion of Branched Molecular Crystals. *Angew. Chemie - Int. Ed.* **2016**, *55*, 7073–7076.

- (4) Hasebe, S.; Hagiwara, Y.; Komiya, J.; Ryu, M.; Fujisawa, H.; Morikawa, J.; Katayama, T.; Yamanaka, D.; Furube, A.; Sato, H.; Asahi, T.; Koshima, H. Photothermally Driven High-Speed Crystal Actuation and Its Simulation. *J. Am. Chem. Soc.* **2021**, *143*, 8866–8877.

- (5) Desta, I. T.; Chizhik, S. A.; Sidelnikov, A. A.; Karothu, D. P.; Boldyreva, E. V.; Naumov, P. Mechanically Responsive Crystals: Analysis of Macroscopic Strain Reveals “Hidden” Processes. *J. Phys. Chem. A* **2020**, *124*, 300–310.

- (6) Kearsley, S. K.; Desiraju, G. R. Determination of an Organic Crystal Structure with the Aid of Topochemical and Related Considerations: Correlation of the Molecular and Crystal Structures of α -Benzylidene- γ -Butyrolactone and 2-Benzylidencyclopentanone With Their Solid State Photoreactivity. *Proc. R. Soc. London A* **1985**, *397*, 157–181.

- (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*, 1587–1592.

- (8) Rodriguez, J. A.; Eisenberg, D. S.; Gonen, T. Taking the Measure of MicroED. *Curr. Opin. Struct. Biol.* **2017**, *46*, 79–86.

- (9) Chin, K. K.; Natarajan, A.; Gard, M. N.; Campos, L. M.; Shepherd, H.; Johansson, E.; Garcia-Garibay, M. A. Organic Molecular Nanocrystals: Triplet–Triplet Absorption, Phosphorescence, and Circular Dichroism of Chiral Crystals of Benzophenone. *Chem. Comm* **2007**, 4266–4268.

- (10) Nath, N. K.; Runčevski, T.; Lai, C. Y.; Chiesa, M.; Dinnebier, R. E.; Naumov, P. Surface and Bulk Effects in Photochemical Reactions and Photomechanical Effects in Dynamic Molecular Crystals. *J. Am. Chem. Soc.* **2015**, *137*, 13866–13875.

- (11) Kasai, H.; Nalwa, H. S.; Oikawa, H.; Okada, S.; Matsuda, H.; Minami, N.; Kakuta, A.; Ono, K.; Mukoh, A.; Nakanishi, H. A Novel Preparation Method of Organic Microcrystals. *Jpn. J. Appl. Phys.* **1992**, *31*, L1132–L1134.

- (12) Bllloch, N. D.; Mitchell, H. T.; Tymm, C. C.; Van Citters, D. W.; Mirica, K. A. Crystal Engineering of Molecular Solids as Temporary Adhesives. *Chem. Mater.* **2020**, *32*, 9882–9896.

- (13) Al-Kaysi, R. O.; Bardeen, C. J. General Method for the Synthesis of Crystalline Organic Nanorods Using Porous Alumina Templates. *Chem. Commun.* **2006**, *11*, 1224–1226.

- (14) Al-Kaysi, R. O.; Bardeen, C. J. Reversible Photoinduced Shape Changes of Crystalline Organic Nanorods. *Adv. Mater.* **2007**, *19*, 1276–1280.

- (15) Good, J. T.; Burdett, J. J.; Bardeen, C. J. Using Two-Photon Excitation to Control Bending Motions in Molecular-Crystal Nanorods. *Small* **2009**, *5*, 2902–2909.

- (16) Kim, T.; Al-Muhanna, M. K.; Al-Suwaidan, S. D.; Al-Kaysi, R. O.; Bardeen, C. J. Photoinduced Curling of Organic Molecular Crystal Nanowires. *Angew. Chemie - Int. Ed.* **2013**, *52*, 6889–6893.

- (17) Gallagher-Jones, M.; Ophus, C.; Bustillo, K. C.; Boyer, D. R.; Panova, O.; Glynn, C.; Zee, C.-T.; Ciston, J.; Mancina, K. C.; Minor, A. M.; Rodriguez, J. A. Nanoscale mosaicity revealed in peptide microcrystals by scanning electron nanodiffraction. *Commun. Biol.* **2019**, *2*, 26.

- (18) Al-Kaysi, R. O.; Müller, A. M.; Bardeen, C. J. Photochemically Driven Shape Changes of Crystalline Organic Nanorods. *J. Am. Chem. Soc.* **2006**, *128*, 15938–15939.

- (19) Manrique-Juárez, M. D.; Rat, S.; Salmon, L.; Molnár, G.; Quintero, C. M.; Nicu, L.; Shepherd, H. J.; Bousseksou, A. Switchable Molecule-Based Materials for Micro- and Nanoscale Actuating Applications: Achievements and Prospects. *Coord. Chem. Rev.* **2016**, *308*, 395–408.

- (20) Kim, T.; Zhu, L.; Al-Kaysi, R. O.; Bardeen, C. J. Organic Photomechanical Materials. *ChemPhysChem* **2014**, *15*, 400–414.

- (21) Nath, N. K.; Panda, M. K.; Sahoo, S. C.; Naumov, P. Thermally Induced and Photoinduced Mechanical Effects in Molecular Single Crystals - A Revival. *CrystEngComm* **2014**, *16*, 1850–1858.
- (22) Commins, P.; Desta, I. T.; Karothu, D. P.; Panda, M. K.; Naumov, P. Crystals on the Move: Mechanical Effects in Dynamic Solids. *Chem. Commun.* **2016**, *52*, 13941–13954.
- (23) Sato, O. Dynamic Molecular Crystals with Switchable Physical Properties. *Nat. Chem.* **2016**, *8*, 644–656.



Supporting Information
**Scale-Dependent Photosaliency and Topotactic Reaction of
Microcrystalline Benzylidenebutyrolactone Determined by
Electron Microscopy and Electron Diffraction**

Vince M. Hipwell,^a Chih-Te Zee,^a Jose Rodriguez*,^a and Miguel A. Garcia-Garibay**^a

^aDepartment of Chemistry and Biochemistry, University of California, Los Angeles, California 90024-1569.

*Email: jrodriguez@mbi.ucla.edu, mgg@chem.ucla.edu

Contents:	Page Number
Materials and Methods	S2 – S3
Molecular Characterization Data	S3 – S4
¹ H NMR and ¹³ C{ ¹ H} NMR spectra	S4 – S6
Powder X-Ray Diffraction Data	S7
Analysis of Aqueous Suspensions	S8
TEM Images	S9 – S10
UV-Vis Data	S11
Crystallographic Data	S12 – S14
References	S15

General Methods

All reagents were used as received without further purification. The reaction to synthesize **1** was carried out in oven-dried glassware under argon atmosphere. ^1H and $^{13}\text{C}\{^1\text{H}\}$ NMR spectra were acquired on a Bruker Avance spectrometer at 500 MHz (^1H) and 126 MHz ($^{13}\text{C}\{^1\text{H}\}$). All chemical shifts are reported in ppm on the δ -scale relative to the residual solvent signal as reference (CDCl_3 δ 7.26 and δ 77.16 for proton and carbon, respectively). Powder X-ray diffraction patterns were measured using a Panalytical X'Pert Pro X-ray Powder Diffractometer. UV-Vis absorption spectra were recorded on an Ocean Optics Flame-T spectrometer with the OceanView software package. UV irradiation of TEM grids was performed in a photochemical reactor with the unfiltered output of a non-calibrated Hanovia 450 W medium pressure mercury vapor lamp in a quartz immersion well. TEM grids were placed at a distance of 30 cm from the lamp and exposed for 5 minutes. Control experiments with dry powder samples on glass slides using the same irradiation conditions were shown to lead to ca. 99% conversion. TEM images and electron diffraction data was collected on a Tecnai TF30 microscope.

Collection of MicroED data by continuous rotation

The crystals of **1** were introduced to a grid (Ted Pella, Carbon Type-A 300 mesh, Cu) by gentle physical contact between the grid and the crystalline powder or by dropcasting of a crystal suspension onto the grid and wicking excess liquid. The grid was placed in a single tilt sample holder. Diffraction patterns and crystal images were collected at room temperature using a FEI Tecnai F30 operated at 300 keV in diffraction mode. Diffraction patterns were recorded while continuously rotating at 0.3 degrees per second using a bottom mount TemCam-F416 CMOS camera (TVIPS). Individual frames were acquired with 3 second exposures per image using selected area apertures corresponding to approximately 5 μm at the sample plane. After collecting full datasets for **1**, the single tilt sample holder was transferred to the Hanovia photochemical reactor and irradiated for 10 minutes. The single tilt sample holder was then returned into the microscope; data was collected for **2** in the same manner as **1**.

MicroED data processing

The collected TVIPS movies were converted to individual images in SMV format, which are compatible with X-ray data processing software. The diffraction images were indexed and integrated with XDS. The datasets were initially indexed up to 1.0 Å and then integrated up to 0.95 Å (Kabsch, 2010). The reflection outputs from XDS were sorted and merged in XSCALE. Single datasets from a single crystal, composed of 100 images for both **1** and **2**, were merged in XSCALE and produce two final datasets with acceptable completeness up to 0.95 Å. While the completeness was acceptable, regions containing the zonal absences that would confirm the presence of glide planes were incomplete. In response to this ambiguity, the reflection lists were kept in space group P 1 2₁1 when submitted to SHELXT, which yielded *Ab initio* solutions (Sheldrick, 2015). In both cases, SHELXT assigned P 1 2₁/n 1 as the most likely space group for the best outputted solutions. The atomic positions were refined against measured data using SHELXL (Sheldrick, 2015).

Molecular Characterization Data

(E)-3-benzylidenedihydrofuran-2(3H)-one (1): Synthesis **1** was conducted as previously reported with minor modifications.¹ A solution of benzaldehyde (1.92 mL, 18.9 mmol) and γ -butyrolactone (2 eq) in 20 mL toluene at 0 °C was treated with a suspension of 1.5 eq sodium methoxide in 20 mL toluene, then allowed to warm to room temperature and stir for 2 hrs. The reaction was diluted with 30 mL ethyl acetate, then 20 mL 9:1 H₂O:H₂SO₄ (concentrated) was added slowly to the reaction mixture. This mixture was washed with brine then dried over MgSO₄. The solvent was removed to give the crude product as a pale yellow solid, which was washed with ether to give **1** as a white solid (676 mg, 21% yield). NMR spectra match literature precedent.¹ ¹H NMR (500 MHz, CDCl₃): δ 7.59 (t, *J* = 2.9 Hz, 1H), 7.51 (m, 2H), 7.44 (m, 3H), 4.48 (t, *J* = 7.2 Hz), 3.26 (td, *J* = 7.4, 3.0 Hz, 2H). ¹³C{¹H} NMR (126 MHz, CDCl₃): δ 172.6, 136.9, 134.8, 130.1, 130.0, 129.1, 123.7, 65.6, 27.6.

6,12-diphenyl-2,9-dioxadispiro[4.1.47.15]dodecane-1,8-dione (2): **2** was synthesized by irradiation of a dry powder of **1**. 10 mg of **1** was ground into a thin layer between two microscope slides and irradiated for 1 hr. with the Hanovia photochemical reactor to give **2** as a white powder in quantitative yield. NMR spectra match literature precedent.² ¹H NMR (500

MHz, CDCl₃): δ 7.37 (m, 8H), 7.31 (m, 2H), 4.39 (s, 2H), 4.19 (t, *J* = 8.8 Hz, 2H), 4.03 (m, 2H), 2.82 (dd, *J* = 13.1, 5.1 Hz, 2H), 2.26 (m, 2H) ppm. ¹³C{¹H} NMR (126 MHz, CDCl₃): δ 178.0, 135.2, 129.5, 129.0, 128.4, 65.4, 54.1, 52.1, 35.0 ppm.

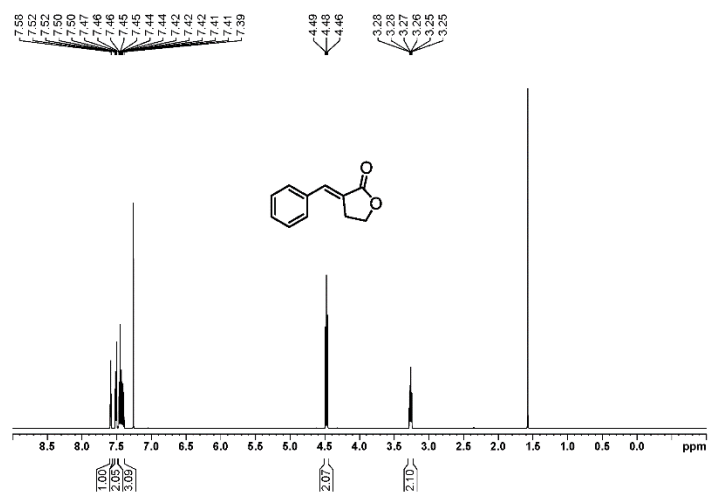


Figure S1. ¹H NMR spectrum (500 MHz, CDCl₃) of **1**.

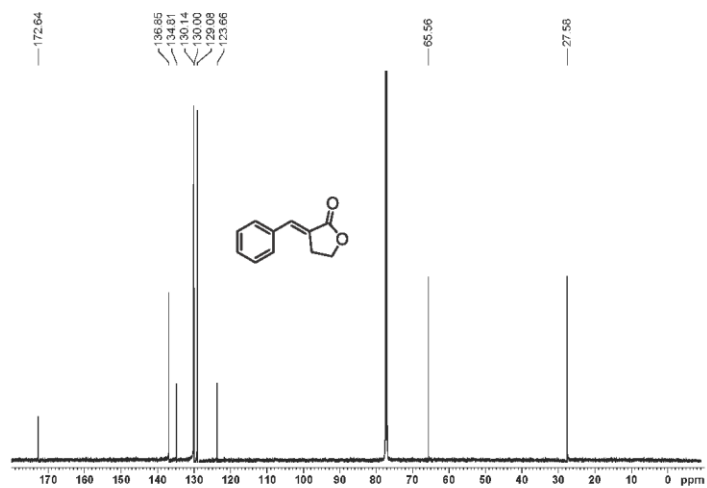


Figure S2. $^{13}\text{C}\{^1\text{H}\}$ NMR (126 MHz, CDCl_3) of 1.

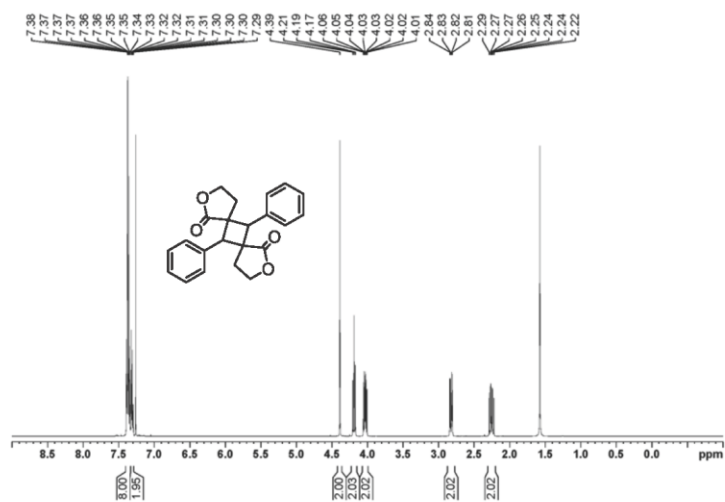


Figure S3. ^1H NMR spectrum (500 MHz, CDCl_3) of 2.

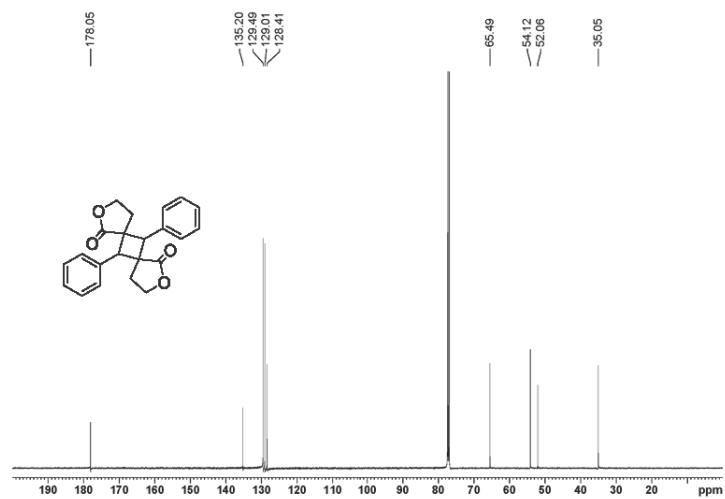


Figure S4. $^{13}\text{C}\{^1\text{H}\}$ NMR (126 MHz, CDCl_3) of **2**.

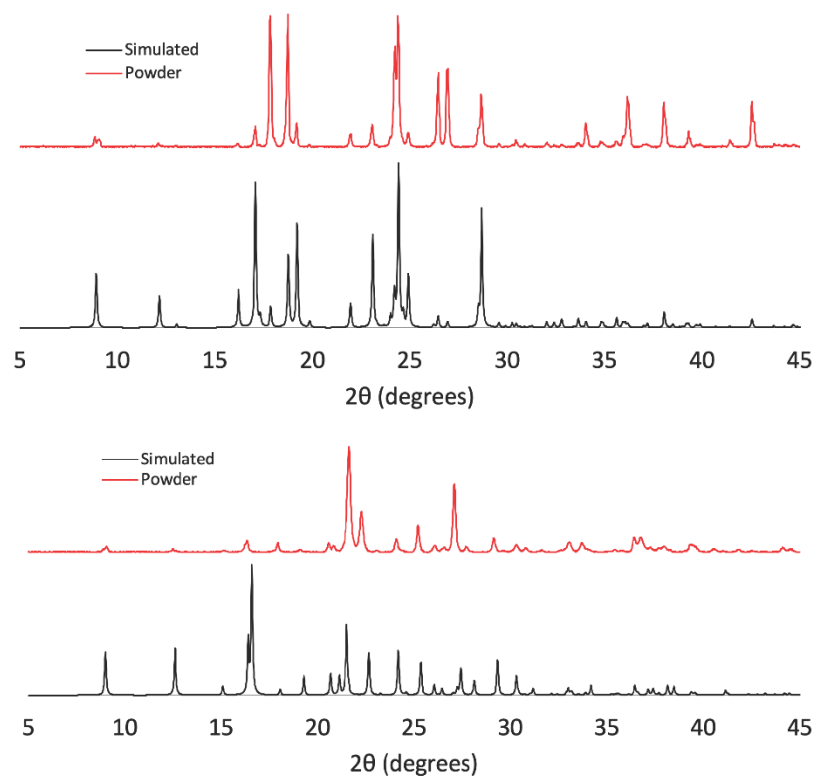


Figure S5. PXRD diffractograms of **1** (top) and **2** (bottom) from powder samples and calculation from single-crystal structure data. Differences in peak intensity between experimental and simulated spectra are likely to be the result of preferred orientation of the plate-like crystals.

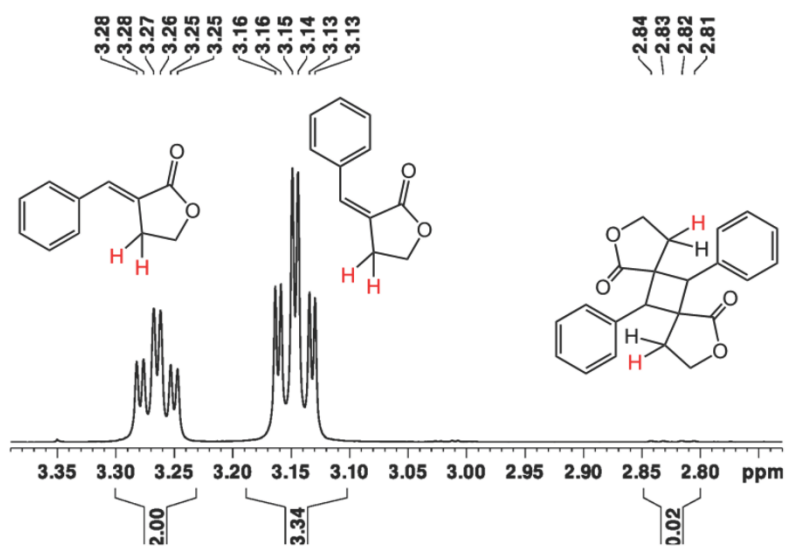


Figure S6. Partial ^1H NMR spectrum following UV irradiation of aqueous suspension of **1**. Hydrogens labeled red correspond to the diagnostic peaks used to establish relative quantities of the three compounds shown.

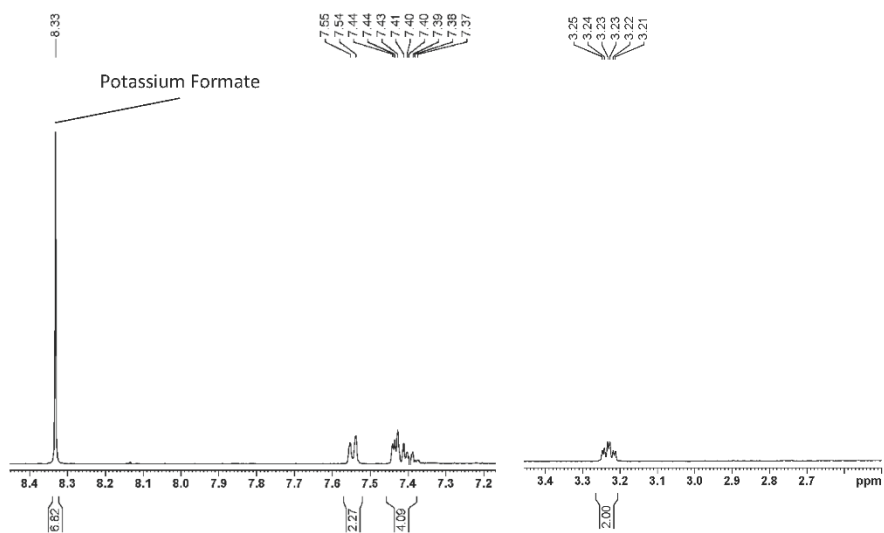


Figure S7. Partial ^1H NMR spectrum of suspension of **1** in D_2O with Potassium Formate (3.23 eq) internal standard.

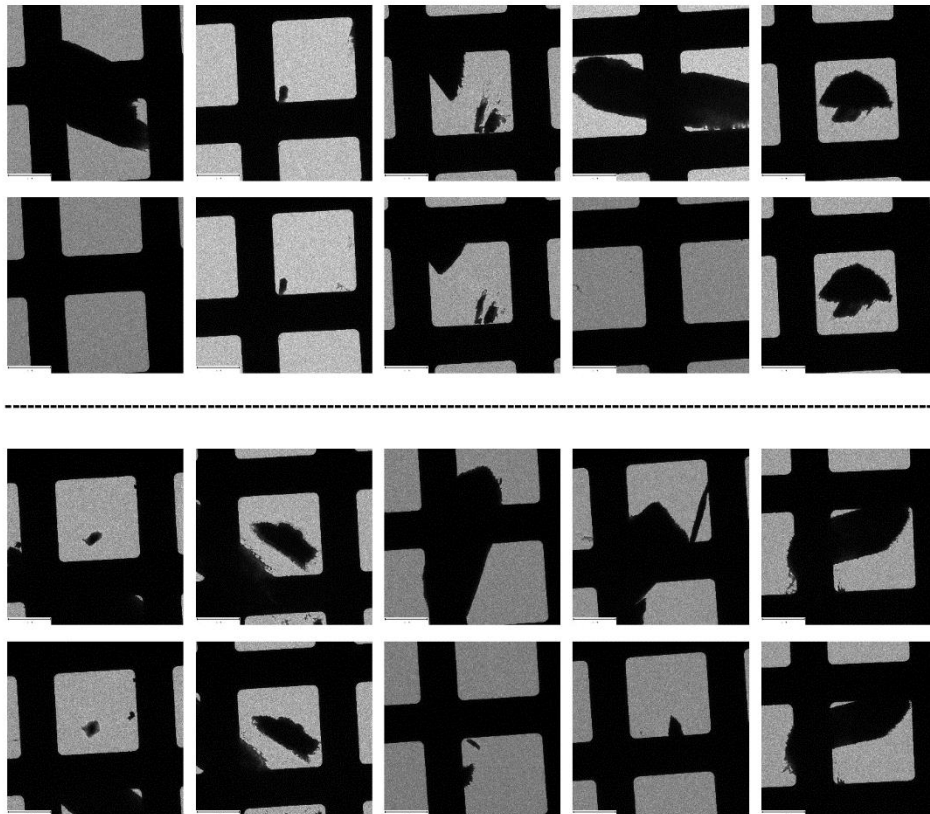


Figure S8. TEM images of microcrystals of **1** on the TEM grid before UV irradiation (top) and the same position on the TEM grid following the UV irradiation (bottom). The scale bar in each image is 30 μm .

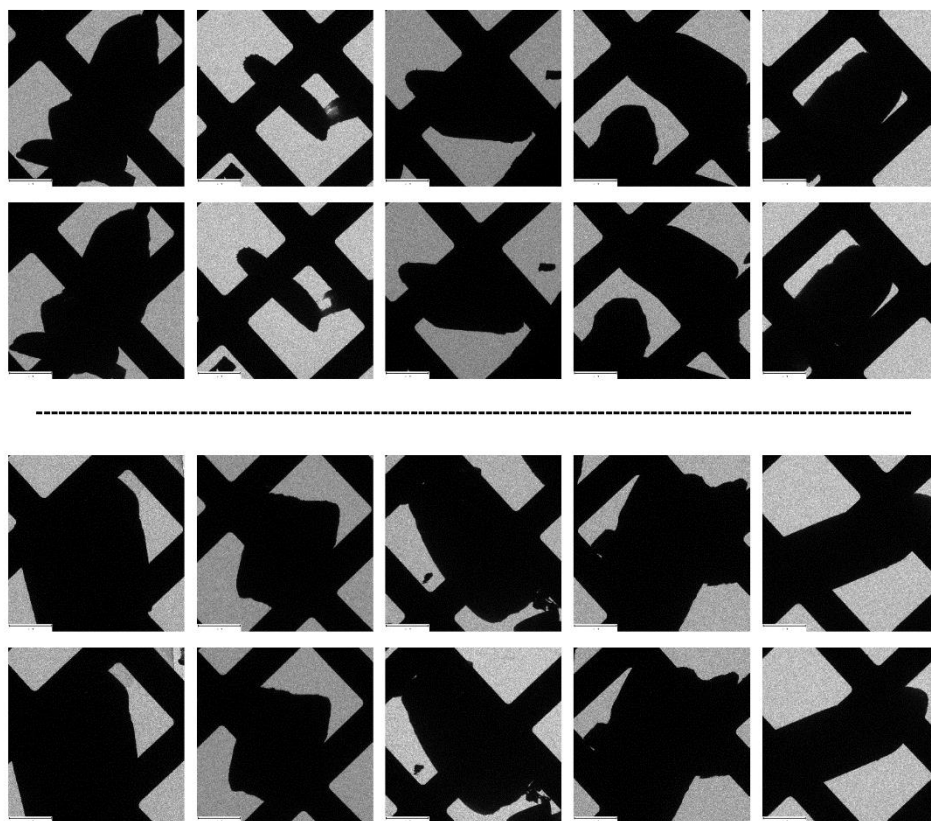


Figure S9. TEM images of microcrystals of **1** on the TEM grid (top) and the same position on the TEM grid following handling of the sample holder in the same manner as was done for UV irradiation experiments (bottom). The scale bar in each image is 30 μm .

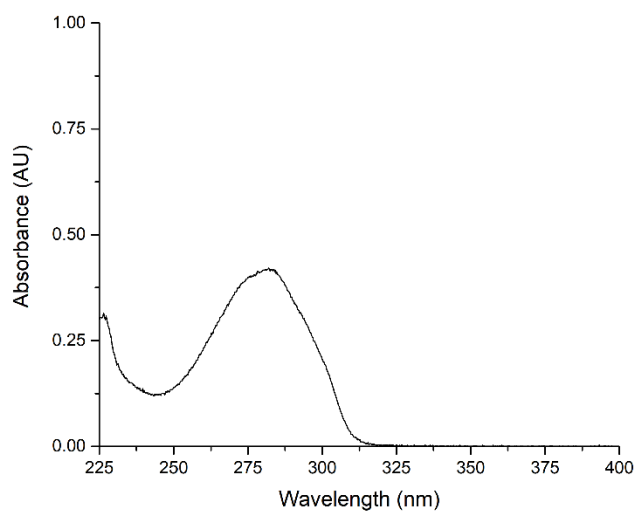


Figure S10. UV-Vis spectrum of **1** in acetonitrile solution (0.0025 mg/mL).

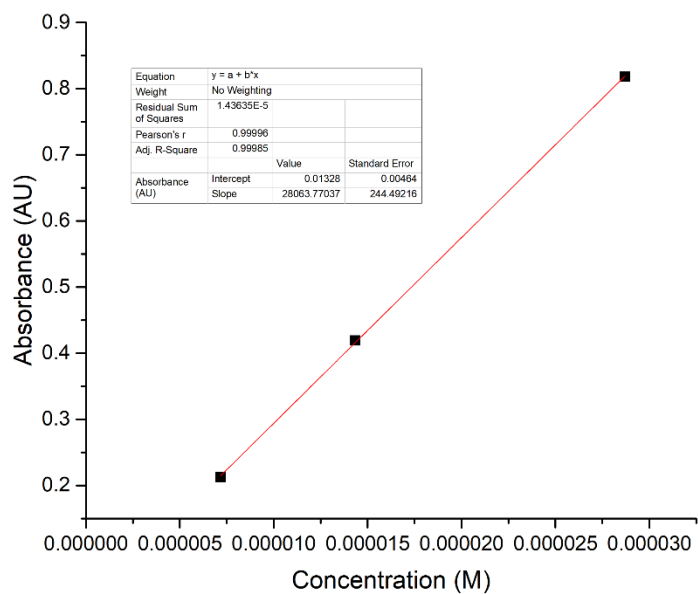


Figure S11. Calibration curve for determination of molar extinction coefficient of **1**.

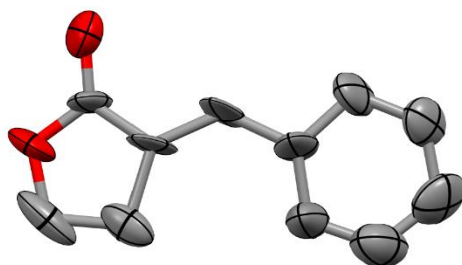


Figure S12. Single-crystal electron diffraction structure of **1** determined by microED.

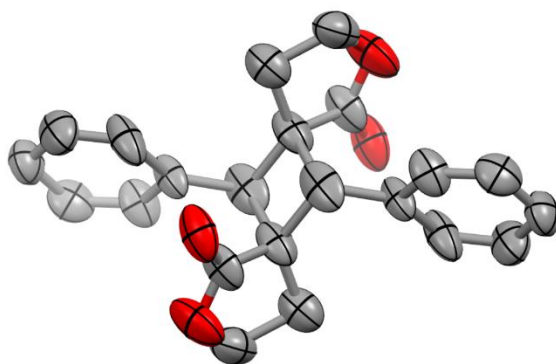


Figure S13. Single-crystal electron diffraction structure of **2** determined by microED.

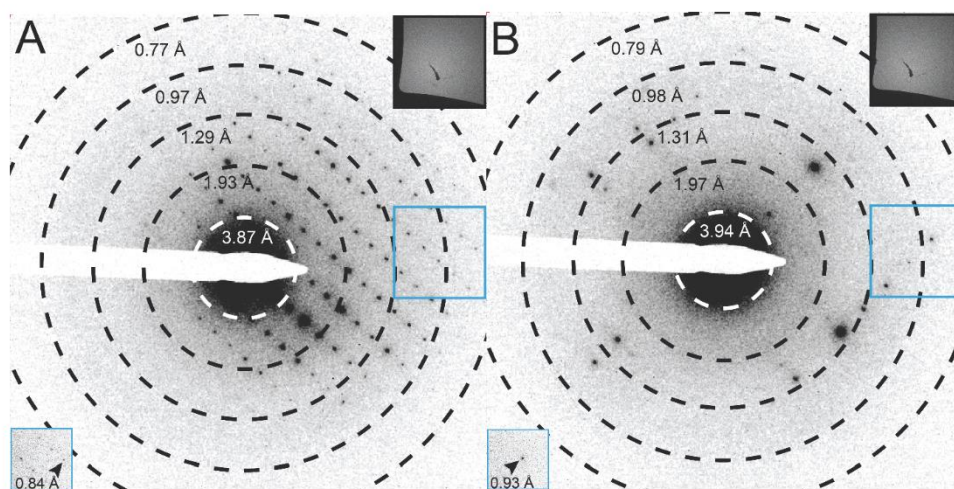


Figure S14. Single crystal diffraction patterns of compound **1** (A) and **2** (B) measured during electron diffraction data collection. The patterns correspond to a 0.9° wedge of reciprocal space. The blue insets in the bottom right show magnified images of the boxed region in each diffraction pattern. The black insets contain images of the same crystal before and after irradiation in the Hanovia photochemical reactor from which data for **1** and **2** were collected. Resolution circles are indicated by the dotted rings.

Table S1. Selected parameters of the crystal structures of **1** and **2**.^{*} As described in the MicroED data processing section (S2-S3), structures solutions were explored in the space groups $P 1 2_1/n 1$ and $P 1 2_1 1$, with the latter being consistent with previous X-ray diffraction analysis,⁶ and also considered the most likely space group for optimal solutions.

Data Collection and Processing				
Compound	1		2	
Molecular Formula	C ₁₁ H ₁₀ O ₂		C ₂₂ H ₂₀ O ₄	
No. of crystals merged	1		1	
Space Group	$P 1 2_1/n 1$	$P 1 2_1 1$	$P 1 2_1/n 1$	$P 1 2_1 1$
Cell Dimensions				
a, b, c (Å)	10.83 5.86 13.42	10.83 5.86 13.42	11.44 5.98 12.43	11.44 5.98 12.43
α, β, γ (°)	90 108.324 90	90 108.324 90	90 108.060 90	90 108.060 90
Resolution Limit (Å)	0.96 (1.03-0.96)	0.95 (1.02-0.95)	0.95 (1.02-0.95)	0.95 (1.01-0.95)
Wavelength (Å)				
Rmerge (%)	5.7 (29.9)	5.7 (30.2)	5.5 (23.0)	5.2 (23.7)
I/σ_I	6.14 (2.09)	6.09 (2.07)	7.28 (2.52)	7.41 (2.52)
Completeness	83.6 (85.7)	75.8 (81.0)	83.8 (85.9)	78.2 (81.5)
CC _{1/2}	99.9 (84.6)	99.9 (84.0)	99.5 (89.7)	99.7 (90.0)
No. total reflections	1511 (251)	1537 (255)	1483 (262)	1543 (233)
No. unique reflections	814 (150)	829 (153)	824 (159)	859 (141)
No. possible reflections	974 (175)	1093 (189)	983 (185)	1099 (173)
Multiplicity	1.85 (1.67)	1.85 (1.67)	1.80 (1.65)	1.79 (1.65)

Refinement				
Resolution (Å)	0.95		0.95	
No. of Reflections	807		804	
R1 (%)	22.26		16.38	
wR2 (%)	56.47		44.44	
GooF (Restrained GooF)	2.211 (2.211)		1.752 (1.752)	

^{*}Values in parentheses are for the highest-resolution shell unless otherwise indicated.

References

- (1) Kabsch, W. XDS. *Acta Crystallogr. D* **2010**, *66*, 125–132.
- (2) Sheldrick, G. M. SHELXT – Integrated space-group and crystal-structure determination. *Acta Crystallogr. A* **2015**, *71*, 3–8.
- (3) Sheldrick, G. M. Crystal structure refinement with SHELXL. *Acta Crystallogr. C* **2015**, *71*, 3–8.
- (4) Uhrner, F.; Lederle, F.; Namyslo, J. C.; Gjikaj, M.; Schmidt, A.; Hubner, E. G. Reaction of N-heterocyclic carbaldehydes with furanones – An investigation of reactivity and regioselectivity. *Tetrahedron* **2017**, 4472–4480.
- (5) Minami, T.; Matsumoto, M.; Suganuma, H.; Agawa, T. Reactions of an α -Phosphono- α -butyrolactone Carbanion with Acid Anhydrides, Epoxides, Carbonates, and a Lactone. *J. Org. Chem.* **1978**, *43*, 2149–2153.
- (6) Nath, N. K.; Runčevski, T.; Lai, C. Y.; Chiesa, M.; Dinnebier, R. E.; Naumov, P. Surface and Bulk Effects in Photochemical Reactions and Photomechanical Effects in Dynamic Molecular Crystals. *J. Am. Chem. Soc.* **2015**, *137*, 13866–13875.