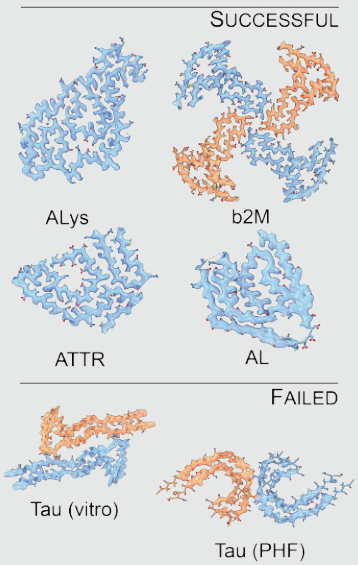
Helical Reconstruction of Amyloids in cryoSPARC

Jan-Hannes Schäfer[a*], Robert T. O'Neill [b], Joseph P. Donnelly [b], Evan T. Powers [b], Jeffery W. Kelly[b*], Gabriel C. Lander[a*],

[a]Department of Integrative Structural and Computational Biology, Scripps Research; La Jolla, CA, USA.
[b]Department of Chemistry, Chi-Huey Wong Laboratories for Biomedical Research, Scripps Research; La Jolla, CA, USA.

Abstract

Amyloid-mediated proteotoxicity underlies more than 50 human diseases. Cryo-electron microscopy (cryo-EM) analyses have yielded hundreds of in vitro and patient-derived amyloid structures, establishing direct links between filament morphologies and specific pathological conditions. Despite the growing popularity of the processing software cryoSPARC for single-particle analyses, RELION remains the dominant software platform for performing helical reconstruction of amyloid structures, highlighting an area for further development. Here, we present comprehensive processing guidelines for helical reconstruction of helical amyloids using cryoSPARC. Through systematic re-processing and validation of publicly deposited datasets, we demonstrate current capabilities and identify key limitations, emphasizing the need for amyloid-specific parameter optimization within cryoSPARC workflows. Our findings showcase a potential for developing unsupervised processing workflows to meet the demanding throughput requirements of time-resolved in vitro studies and large-scale compound screening initiatives, thereby accelerating therapeutic drug development. Ultimately, our goal is to shift the focus of amyloid cryo-EM from computationally intensive processing challenges toward addressing fundamental biological questions that enhance our capacity for treatment discovery.



Keywords

helical reconstruction, amyloids, cryoSPARC

Introduction

Many macromolecules assemble into filaments with helical symmetry to enable essential cellular functions including motility, morphological organization, and macromolecular trafficking[1]. While these universal assemblies typically increase protein stability and allow responsiveness to physiological changes through directed assembly and disassembly[2], aberrant protein homeostasis can result in pathological fibril formation following irreversible conformational changes. The resulting amyloid assemblies are characterized by stable cross-β-strand arrangements and have a propensity to accumulate into deposits called plaques. Such plaques are implicated in over 50 human diseases, including Alzheimer's disease (AD), Parkinson's disease (PD), and transthyretin amyloidosis (ATTR)[3,4]. Since amyloid fibrils are generally recalcitrant to crystallization, and are too large for NMR analysis, cryo-electron microscopy (cryo-EM) has emerged as the primary method for resolving the molecular details of amyloidogenic disease pathologies. While today's single-particle workflows can achieve true atomic resolution[5], helical reconstruction differs substantially from the traditional workflows used for single particle analyses of macromolecules, requiring specialized processing procedures for accurate structure determination. Helical symmetry optimization tools were previously implemented in RELION to specifically target amyloids, which enabled the first reconstruction of patient-derived tau fibrils[6,7]. These initial studies led to a flurry of ex-vivo amyloid structures, identifying a link between amyloid folds and underlying disease pathologies and accelerating our understanding of these devastating diseases[8]. Amyloids present unique structural challenges, as they usually present as helical assemblies with constrained screw operators, shallow helical rise values (Δz), and narrow twist ranges ($\Delta\Phi$) (Fig.1A). This distinctive symmetry, combined with substantial structural polymorphism observed in ex vivo samples, makes automated filament picking and accurate initial model generation particularly challenging[9]. Additional complexity arises from substantial structural polymorphism (Fig.1B): Differences in peptide-packing result in polymorphs at the protofila-

*Correspondence to:
glander@scripps.edu, jkelly@scripps.edu, jschaefer@scripps.edu

ment level, while ultrastructural differences arise from varying numbers of interacting protofilaments within the amyloid fibril. Compositional polymorphs result from ultrastructures of different amyloid proteins and non-amyloid components[10]. Posttranslational modifications such as glycosylation, disulfide bridges in light-chain amyloids[11], or α-synuclein phosphorylation in PD[12] add another level of heterogeneity (Fig. 1C). These structural complexities have necessitated amyloid-specific cryo-EM processing tools, which have been successfully developed and implemented within RELION[7,13,14]. While the Electron Microscopy Data Bank (EMDB) currently reports over 12,700 RELION-generated reconstructions compared to approximately 9,600 from cryoSPARC[15], the majority of helical reconstructions — including nearly all amyloid structures — have been processed using RELION (https://www.ebi.ac.uk/emdb/emsearch/charts/, Fig. 1D). Of the 352 amyloid reconstructions documented in the Amyloid Atlas[16], only bacterial PSMα1 polymorphs (EMD-43835)[17], FapC (EMD-49649)[18], and one ex vivo light-chain amyloid reconstruction (EMDB-70557)[19] have been processed using cryoSPARC. One possible reason for this discrepancy stems from a user's greater ability to tightly control for parameter optimization within RELION, whereas cryoSPARC prioritizes generalized processing guidelines. Effective software development relies on community involvement and feedback. To expand cryoSPARC into a truly universal processing pipeline capable of handling both single-particle analysis and helical reconstruction, comprehensive guidelines for reconstructing amyloids need to be established and formalized. This work aims to systematically evaluate the current capabilities of cryoSPARC (v4.7) for resolving amyloid structures while identifying key limitations in existing processing workflows. We use five RELION-processed EMPIAR datasets as case studies and ground truth references to cross-validate reconstructions generated within cryoSPARC. Our comprehensive analysis guides users through the complete processing pipeline: from initial assessment of collected micrographs and rational selection of helical symmetry parameters, through classification of heterogeneous samples and reconstruction of homogeneous polymorphs, to final validation using established community-driven metrics. By providing these systematic guidelines and benchmarks, we hope to accelerate the processing of challenging helical assemblies and unlock new structure-driven insights into amyloid disease pathology and therapeutic intervention strategies.

Results

Processing guidelines for amyloids using cryoSPARC

The following approaches for reconstructing amyloids in cryoSPARC were identified through systematic parameter optimization and provide a foundation for target-specific adjustments, though universal applicability may vary. A comprehensive workflow is depicted in Fig. 2.

Preprocessing and initial considerations:

For image pre-processing, beam-induced motion was corrected using Patch Motion Correction, followed by Patch CTF Estimation. Amyloids typically accumulate in thicker ice regions near grid hole edges, necessitating less stringent selection criteria than conventional single-particle analysis. We retained micrographs with CTF fits above 6 Å and discarded those with astigmatism values exceeding 600 Å. Picking templates were generated using template-free Filament Tracing on 20 representative micrographs. The tracer identifies helical filaments as short, straight, partially overlapping segments using an approach adapted from SPRING[20]. These extracted segments serve as input for iterative helical real-space reconstruction (IHRSR)[21]. Both RELION and cryoSPARC implement the IHRSR algorithm, segmenting filaments into overlapping regions that can be averaged while preserving the ability to separate heterogeneous populations. This segmentation approach is feasible because single projections of helical assemblies contain complete information for reconstructing the asymmetric unit[21].

Filament Tracing to 2D Classification:

Filament tracing efficiency improved when diameters slightly below actual filament dimensions were selected, combined with increased standard deviation values for Gaussian blur (0.4-0.5) to enhance filament contour prominence. To separate visually distinct polymorphs, the initial extraction box should exceed the approximate crossover distance, which can be estimated from high-defocus cryo-EM micrographs or negative-stain transmission electron microscopy pre-screening. Two rounds of 2D classification effectively generate well-aligned class averages that enable separation of different filament types. Optimal results were achieved using a large number of classes (K=100) with an increased initial uncertainty factor (e.g., 4) and 40 expectation-maximization iterations, enabling hard classification during the final iteration. This approach effectively reduces per-class heterogeneity, as evidenced by high effective sample size (ESS) values approaching unity and Fourier ring correlation (FRC) values near the Nyquist frequency.

Initial Reconstruction Generation and Polymorphism Management:

Unlike RELION, cryoSPARC does not support initial model generation with helical priors using projection overlaps across entire helical crossovers[7]. Instead, cryoSPARC generates

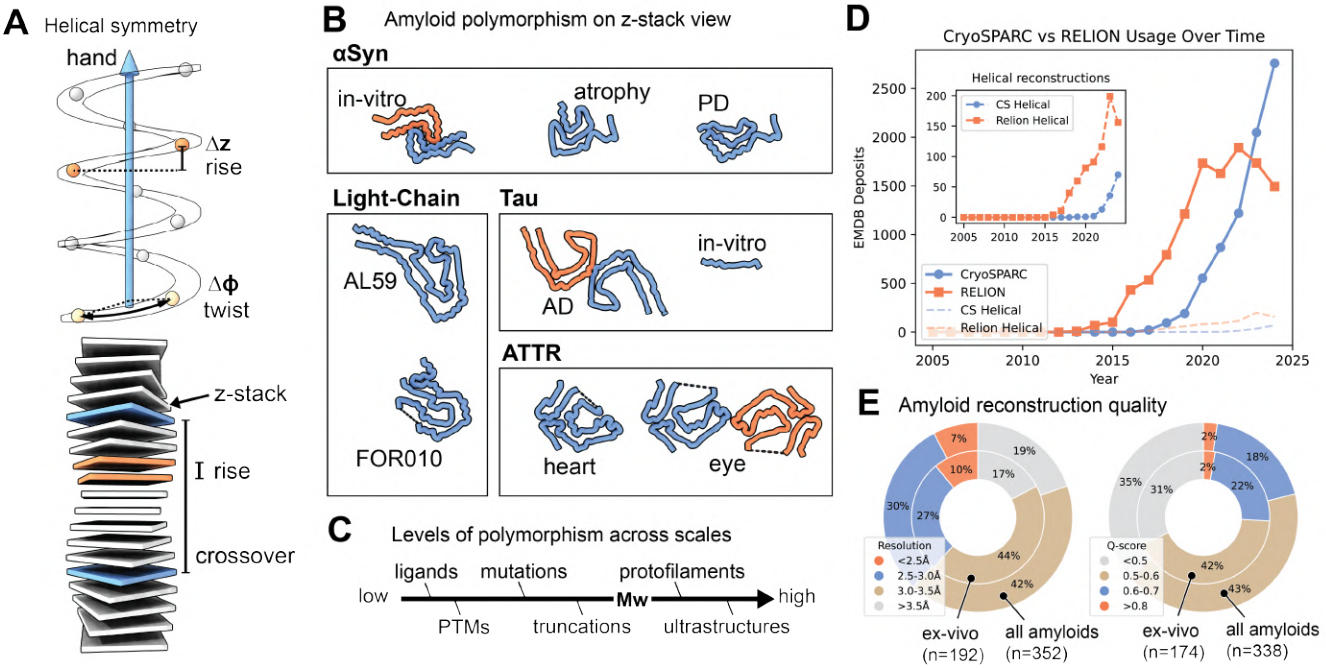


Figure 1. Architecture, diversity, and reconstructions of amyloids. (**A**) Schematic representation of helical symmetry parameters: helical rise (Δz, orange) and helical twist (ΔΦ, tan) with central helical z-axis (blue). Arrow indicates helix handedness. Amyloid z-stack model showing rise (orange) and crossover distance (blue). (**B**) Cross-βpeptide arrangements of amyloid polymorphs grouped by amyloid protein. Single-chain polymorphs shown in blue; double-chain polymorphs in orange/blue. AL59 and FOR010 represent patient identifiers from light-chain amyloidosis cases. (**C**) Contributors to amyloid polymorphism organized by molecular weight (Mw), including post-translational modifications (PTMs). (D) Cumulative EMDB depositions categorized by processing software: RELION (orange) and cryoSPARC (blue). Data for 2025 omitted for annual comparison consistency. Helical reconstructions highlighted in inset. (E) Validation metrics for deposited amyloid reconstructions using Amyloid Atlas data. Q-scores and resolutions at FSC 0.143 for ex vivo and all annotated amyloid entries displayed as nested pie charts.

cylindrical starting models through its Helical Refinement job. Selecting homogeneous projections that span an entire crossover distance proves most efficient for initial model generation (Fig. S1). Using smaller extraction boxes may produce mixed starting models from conflicting projection sets. Similar to RELION's expectation-maximization optimizer, the initial volume must fall within the convergence search range to generate accurate final reconstructions[7]. Separating inherent heterogeneity (polymorphism) requires manual grouping of matching 2D projections in cryoSPARC. Advanced tools such as FilamentTools for bi-hierarchical classification of filament 2D projections in RELION5[14] or Clustering of Helical Polymers (CHEP)[22] are not yet available in cryoSPARC. Consequently, users must manually select class averages representing suspected polymorphs, with inherent risk of false-positive selections. The initial Helical Refinement should proceed without imposing axial or screw symmetry constraints, without limiting shifts along the helical axis, and with maximum out-of-plane tilt searches set to 60°.

Symmetry Parameter Determination:

While large extraction boxes with high binning enable initial amyloid identification and rough separation of protomer-level polymorphs, helical segments should be re-extracted using smaller box sizes (approximately 300 Å or three times the amyloid diameter) with reduced binning to resolve β-stack separation. This approach typically matches the Nyquist frequency to values below the expected separation distance or helical rise of amyloids (approximately 4.8 Å). To select projections displaying visible stack separation, segments were re-extracted, re-centered, and subjected to 2D classification using maximum alignment resolution of 4 Å, initial class uncertainty factor of 4, and hard classification enabled during the final iteration Fig. S1. The Average Power Spectra job type applied to well-resolved projections enables determination of helical rise from prominent layer lines. By measuring the distance from the meridian to the layer line of interest (d/px), the rise (Δz/Å) can be calculated as: $\Delta z = (\text{box size})/(d) \times 2 \times (\text{pixel size})$. Approximating helical twist ($\Delta\varphi$) requires the crossover distance, which must be estimated from low-pass filtered micrographs or complementary methods such as negative-stain transmission electron microscopy or atomic force microscopy. The twist can then be calculated using: $\Delta\varphi = (\Delta z \times 180°)/(\text{crossover distance})$. While reciprocal space measurements provide one approach for deriving screw parameters, real-space helical indexing tools like HI3D offer alternatives by using projections of asymmetric helical reconstructions without requiring manual approximations, accessible through its web-based server[23]. For improved HI3D accuracy, finer sampling parameters should be employed: angular

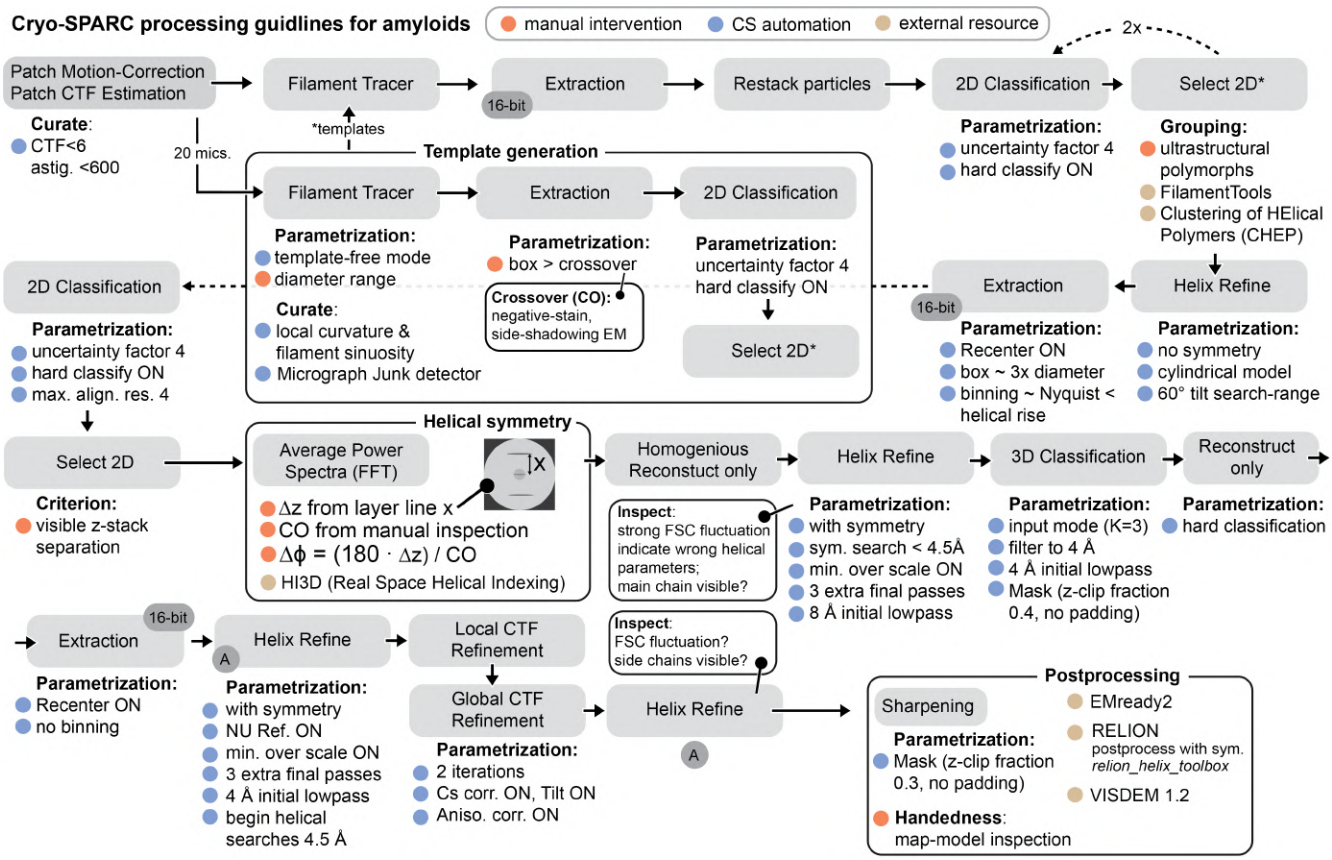


Figure 2. Processing guidelines for amyloids using cryoSPARC. Individual cryoSPARC job types are displayed in grey boxes, with recommended non-default parameters marked by blue bullets. Steps requiring manual intervention or supplementary experiments are denoted by orange bullet points, and external resources are denoted by tan bullet points.

steps of 0.5° and axial steps of 0.2 Å[24]. Before proceeding to helical refinement, particle quality should be assessed using the "Homogeneous Reconstruction Only" job with enforced helical rise applied to the selected particle subset.

Final amyloid reconstruction through refinements and classifications:

To assess the accuracy of the approximated helical parameters, helical refinements should be performed with helical symmetry enabled, 8 Å initial low-pass filtering, per-particle scale minimization enabled, and three additional final passes. Maximum out-of-plane tilt should be set to 60°. Symmetry searches should commence only at 4.5 Å resolution to maintain convergence toward separated cross-β stacks. Strong fluctuations in FSC curves with high discrepancies between tight and corrected FSC curves for auto-masking may indicate incorrect helical symmetry (Fig. S1,S2, see Tau-PHF example). For samples where there is substantial compositional or conformational heterogeneity, 3D classification without alignment may facilitate particle separation into distinct structural classes, to enable symmetry search convergence on subsets to achieve meaningful reconstructions. Reported resolution estimates at this processing stage are unlikely to reflect the actual resolution of the structures, and reconstructions should be critically examined qualitatively for main-chain visibility and stack separation. Classification using input mode with three classes (K=3), filtered to 4 Å with a low-pass filter at 4 Å and focused masking (z-clip fraction of 0.4, no extra padding) may successfully separate low-quality particles or distinct polymorphs. A "Reconstruction Only" job with hard classification enabled and enforced rise values should be performed to inspect reconstructions and select consensus structures for further processing. The selected particle stack should then be re-extracted and re-centered without binning, followed by another round of helical refinement with symmetry searches enabled (starting at 4.5 Å), 4 Å low-pass filtering, non-uniform refinement enabled for three additional final passes, and per-particle scale minimization enabled. Local and global CTF refinement should be attempted, as these jobs may further enhance reconstruction quality. However, after each CTF refinement step, subsequent helical refinements should be performed to evaluate improvements. The resulting reconstructions require careful manual inspection, comparing observed density quality against prior reconstructions, rather than relying on the reported FSC-based resolution estimates, to ensure meaningful structural interpretation. Global CTF refinement with three iterations should include both tilt and spherical aberration correction. In our testing, reference-based motion correction was not performed since not all selected datasets included movie data, though it may provide further

resolution improvements where movie frames are available. Finally, reconstructions can be post-processed using B-factor sharpening within cryoSPARC, employing masks with z-clip fractions of 0.3 and no additional padding. Alternatively, EM-ready2[25] or VISDEM1.2 (https://github.com/gschroe/visdem) can be used to improve reconstruction interpretability and facilitate model building. Most amyloids are left-handed (exhibiting negative twist values), providing a logical starting point for twist value optimization. Depending on final reconstruction quality, backbone traces can directly inform map handedness determination (typically below 3 Å resolution). For reconstructions with lower interpretability, supplementary methods such as negative-stain transmission electron microscopy or atomic force microscopy may be necessary to determine absolute handedness. Database comparisons using resources such as the Amyloid Atlas[16]may also facilitate model building and validation.

Reconstructing amyloids to high resolution using cryoSPARC

To establish robust processing guidelines for reconstructing amyloids in cryoSPARC, six EMPIAR datasets were selected for systematic evaluation. These include five RELION-processed depositions and one recent cryoSPARC-processed deposition: ex vivo paired-helical filament (PHF-type) tau, recombinant intermediate-amyloid tau (MIA-type), cardiac transthyretin amyloids (ATTR), recombinant β2-microglobulin variant V27M, ex vivo lysozyme D87G, and ex vivo light-chain Lambda6 amyloids (Table 1). The datasets were selected to represent the diversity of amyloid diseases, origins (in vitro vs. in vivo), helical symmetries, data quality, availability, and dataset sizes. Following the processing guidelines described above, light-chain, transthyretin, lysozyme, and β2-microglobulin amyloids were successfully reconstructed with high confidence, showing well-resolved side chains (Fig. 3A, Fig. S1S2). Recombinant tau fibrils were resolved sufficiently to visualize the main chain and achieve stack separation, though side-chain placement remains ambiguous. However, the ex vivo tau dataset failed to yield a correct reconstruction (Fig. 3A), highlighting the common problem of convergence to incorrect local refinement minima[7]. While β-stack separation for ex vivo PHF tau was clearly observable in 2D class averages (Fig. S1), refinement did not converge on correct helical parameters, resulting in an incorrect reconstruction exhibiting severe overfitting (Fig. S1). In a recent publication, we characterized cardiac ATTR filaments from three different patients (wild-type and familial V122I variant transthyretin), linking amyloid deposits to microangiopathy and capillary occlusion, which is currently under revision. These amyloids share the common spearhead fold characteristic of transthyretin amyloids, which we reconstructed to high resolution using the presented workflow (Fig. 3B, Table 1). Thus, our processing guidelines enabled meaningful reconstruction of eight out of nine selected datasets in cryoSPARC. Notably, many amyloids exhibit structural polymorphism[26]. Based on visible protofilaments in 2D class averages, no oligomer-level polymorphism was observed for β2-microglobulin, lysozyme, ATTR, or tau (Fig. S1). Consistent with the original work, the processed light-chain amyloids assemble into single- and double-protomeric arrangements[19], but only single-chain amyloids yielded successful reconstructions. Re-processing identified a pseudo-C2 symmetric protomer arrangement in light-chain amyloids (Figure S1) but did not produce a high-quality reconstruction. This polymorphic arrangement was previously reported for a different Lambda6 germline light chain, which also failed to yield reconstructions suitable for atomic model building. The authors suggested a unique 2-fold rotational symmetry perpendicular to the fibril axis[27].

Validating helical reconstructions using cryoSPARC

Helical reconstructions are inherently susceptible to artifacts and subsequent misinterpretation due to the imposition of high-order helical and axial symmetry constraints that inflate the reported resolutions. A recent preprint reported errors in approximately 14% of helical reconstructions deposited in the EMDB, identified through real-space helical indexing and manual data curation[24]. Given the prevalence of such biases, FSC-based resolution estimation alone cannot reliably assess helical reconstruction quality (Fig. 3C, Fig. S1) and should not serve as the sole validation criterion. Therefore, helical symmetry correctness must be supported by comprehensive map-model assessment, including clear cross-β stack separation, continuous backbone traces, side chain resolvability, and proper stereochemistry[7]. To quantify the interpretability of our amyloid reconstructions, Q-scores[28] were calculated using ChimeraX and compared to original deposited data using corresponding identifiers from Table 1. Q-scores were calculated for single cross-β stacks and plotted in Fig. 3C (detailed values in Table S1). While most reconstructed amyloids support the fitted atomic models with Q-scores above 0.6, the original RELION reconstructions yielded higher interpretability values. The re-processed light-chain AL-L6 from cryoSPARC showed modest improvement compared to the original deposition. Comparison of applied helical symmetry parameters reveals minor discrepancies in helical rise values, potentially contributing to reduced map interpretability. Additionally, the pseudo-2_1 screw symmetry of tau reconstructions was processed using C2 symmetry, which may impact overall interpretability.

Identification of a potential ALys Amyloidosis intermediate

During re-processing of ex vivo lysozyme variant D87G, associated with hereditary systemic lysozyme amyloidosis (ALys)[29], non-amyloidogenic fibrils were identified through 2D classification (Fig. 4A). Approximately one-third of micrographs contained these distinct fibrillar structures. Helical refinement using parameters of 37.23 Å rise and 14.17° twist

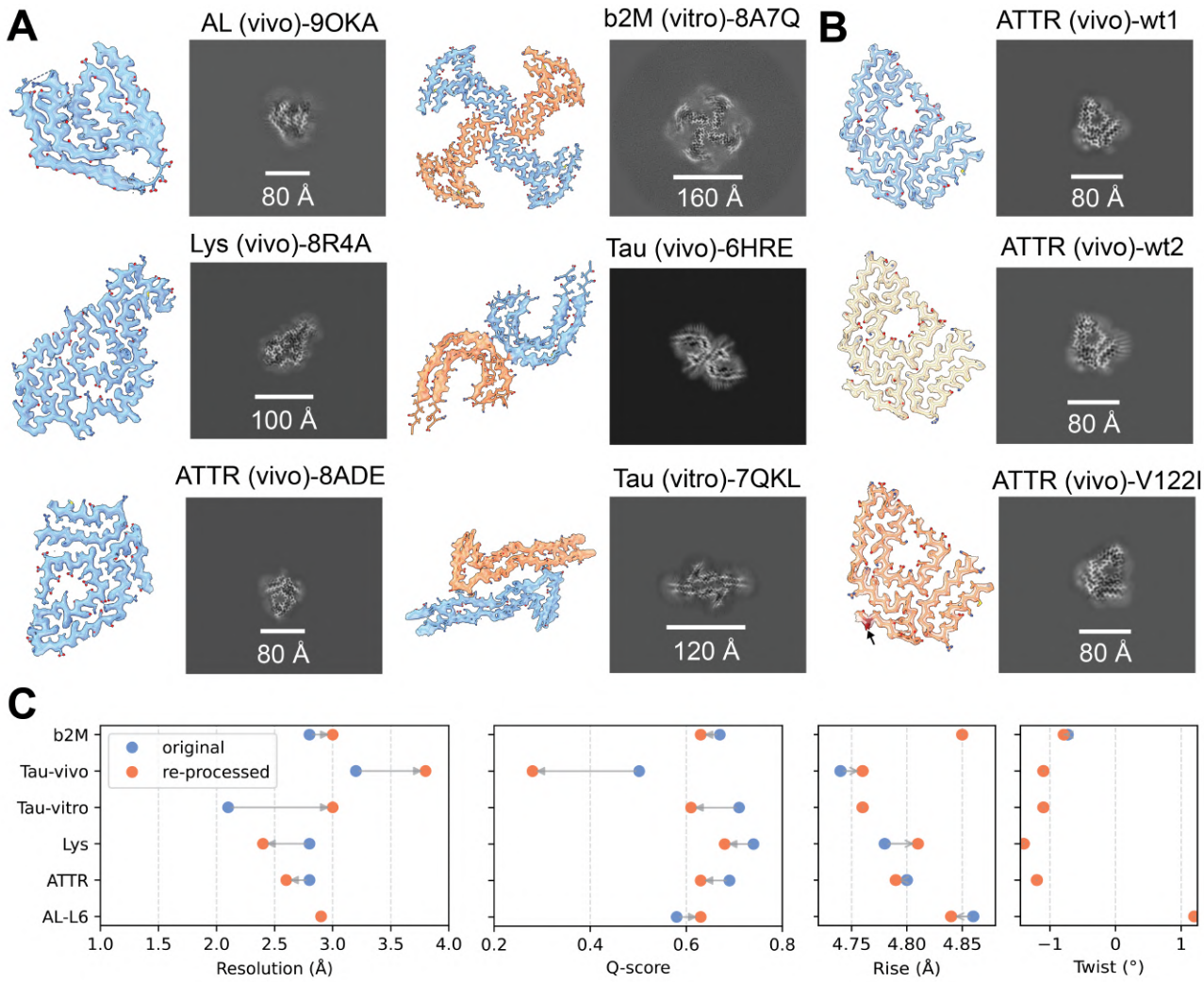


Figure 3. Performance of amyloid processing using cryoSPARC. (**A**) Map-model fits and xy-slices with scale bars for refined reconstructions of amyloid datasets from EMPIAR. Light-chain amyloids (AL, PDB-9OKA), β2-microglobulin variant V27M (β2M, PDB-8A7Q), lysozyme variant D87G (Lys, PDB-8R4A), ex vivo PHF-type tau (PDB-6HRE), transthyretin amyloids (ATTR, PDB-8ADE), and recombinant tau MIA-type amyloids (PDB-7QKL). Single-protofilament polymorphs shown in blue; multi-protofilament polymorphs in blue and orange. (**B**) Cardiac ATTR reconstructions from our previous work showing two wild-type transthyretin samples and familial variant V122I (indicated by arrow). (**C**) Validation of re-processed EMPIAR datasets (orange) compared to original depositions (blue, from Table 1) with resolutions estimated according to an FSC 0.143 cutoff, Q-scores calculated with ChimeraX, and helical symmetry parameters.

(Fig. S3) yielded a 4 Å resolution reconstruction revealing 10 nm diameter tubes with C6-symmetric protofilament arrangement and a 3.5 nm inner diameter (Fig. 4B). Post-processing with EMready2 enhanced interpretability, enabling automated model building with ModelAngelo[30]. Neither human lysozyme C nor annotated human reference proteomes as input produced convincing side-chain placements, necessitating construction of a poly-alanine model for visualization of this unidentified complex (Fig. 4C). The resulting model suggests intra-protofilament packing driven by an elongated β-hairpin motif and inter-protofilament packing through extensive β-complementation lining the tubular pore. While residue-level interpretation remains limited by current resolution, this structure warrants further investigation as it may represent a novel disease-relevant intermediate in ALys amyloidosis pathogenesis or common amyloid-binders like Serum-amyloid protein (SAP).

Material & Methods

Data selection and curation

For database searches, the EMDB REST API (https://www.ebi.ac.uk/emdb/api/) was utilized to automatically extract the Q-score and resolution values for matching EMDB entries of amyloids reported by the Amyloid Atlas[16]. For testing and validating our processing guidelines in cryoSPARC, we re-processed the amyloid reconstructions from six EMPIAR datasets, listed in Table 1. Reported Tau amyloids were originally deposited with pseudo-21-screw

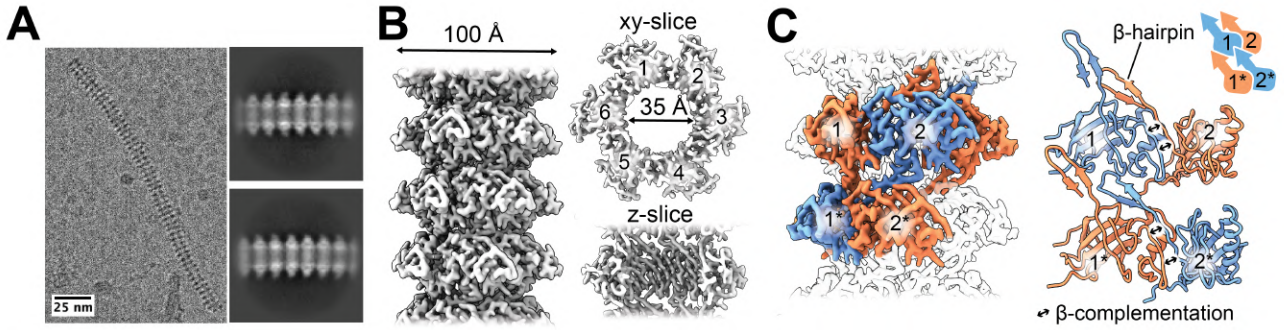


Figure 4. Reconstruction of an unidentified tubular assembly in ex vivo lysozyme amyloids. (**A**) Micrograph from EMPIAR-11785 showing 2D class averages of a non-amyloidogenic tubular fibril (scale bar: 25 nm). (**B**) 4 Å helical reconstruction of 10 nm C6-symmetric tubes with 3.5 nm inner diameter, displayed along the helical axis and as xy- and z-slices. Numerals indicate individual protofilaments. (**C**) Cryo-EM reconstruction color-coded by subunit packing using a poly-alanine model automatically fitted into the density using ModelAngelo. Inter-protofilament contacts through β-complementation are indicated by black double arrows.

symmetry, which was converted to their appropriate C2 symmetry.

Amyloid processing

All datasets were processed in cryoSPARC v.4.6-7 on NVIDIA Tesla L4 GPUs and Intel Xeon CPU E5-2698 v4 with up to 4 GPUs for multi-GPU CS job-types. Processing of EMPIAR datasets was performed following the established guidelines shown in Fig. 2 and described in the Results section. A comprehensive validation of the individual datasets is shown in Fig. S1,S2 and Table S1. Results were visualized using ChimeraX[31] and ggplot in Python.

Processing of tubular filaments in EMPIAR-11785

EMPIAR-11785 was processed following the established guidelines shown in Fig. 2. From the 2,013 processed micrographs, about 30% contained non-amyloid segments and were selected for further processing. To increase the number of particles, filaments were traced using a diameter of 40-80 Å without templates and 0.5x gaussian blur. Segments were extracted in a 208 pixel box, Fourier-cropped to 100 pixel and subjected to two rounds of 2D classification with an initial class uncertainty factor of 4 and hard classification during the last iteration. 30K selected segments were re-extracted with re-centering in a 256 pixel box and no binning. Helix refine was performed without symmetry using a cylindrical model and a 60° tilt search-range. The resulting reconstruction was used as input in Real-space helical indexing in HI3D with default parameters[23], suggesting a twist of 13.74° and a rise of 37.93 Å. These values were used in helix refine with non-uniform refinement (NUR) enabled. To test for higher-order symmetry, Phenix's symmetry search tool phenix.map_symmetry[32] was used, resulting in the highest correlation with C6 point-group symmetry. After another round of 2D classification, 21K segments were used in helical refinement with C6, a rise of 38 Å and twist of 13.7°. Helical indexing with the new C6 point group and higher-resolution map suggested a rise of 37.19 Å and twist of 14.21°. The final helical refinement with a 20° tilt search-range resulted in a 4 Å reconstruction (FSC 0.143). The map was sharpened automatically within cryoSPARC using a B-factor of –56 Å2. The full processing workflow and validation is shown in Fig. S3 and Table S1. Resolvability of the map was increased with postprocessing in EMready2[25] and utilized for automatic model building with Model Angelo[30]. Human lysozyme C (Uniprot: P61626) and the annotated human reference proteome (UP000005640) were used as input in FASTA or the HMM search procedure, respectively.

Discussion

In our previous work on ex vivo ATTR amyloids, we demonstrated cryoSPARC's capability to reconstruct amyloids from complex tissue extracts to high resolution (Fig. 3B). This success motivated the development of systematic processing guidelines for amyloid reconstruction using cryoSPARC as a viable alternative to RELION. Following our published ATTR work, we re-processed deposited amyloid datasets using our established workflow, successfully reconstructing four of six selected amyloid datasets to high resolution, one dataset to reasonable resolution, and only one dataset failing to achieve suitable resolution for model building. Notably, our new processing workflow, available as import-ready .json files (https://github.com/schaefer-jh/CS-amyloids/), achieved 2.4 Å global resolution for ALys amyloids (Fig. 3A, Fig. S1) in approximately 12 hours using three GPUs for multi-GPU processing of 2,000 movies. By integrating our recently published deep neural network CryoSift (https://cryosift.org)[33] to automatically select suitable 2D class averages through its integrated mode with cryoSPARC-tools, unsupervised high-throughput processing of amyloids with known helical symmetry becomes feasible in cryoSPARC, potentially accelerating future drug development efforts. While many of the presented non-default processing parameters facilitate reconstruction of homogeneous amyloids and protofilament-level polymorphs in cryoSPARC, reliable classification of polymorphs at the peptide level remains challenging. To enhance cryoSPARC's

Table 1. EMPIAR datasets of helical filaments. Helical crossover (CO), resolution (d). * denotes our own work, processed in cryoSPARC (v. 4.6–7). Helical crossover was calculated as $\mathrm{CO} = \frac{180}{\mathrm{twist}} \times \mathrm{rise}$.

EMPIAR	EMDB	d (Å)	PDB	Protein (source)	mic. #	Å/px	Dose (e/Å^2)	Rise (Å)	Twist (°)	sym	CO (Å)
10940	14046	2.1	7QKL	Tau (vitro)	331	0.824	40	4.76	-1.1	C2	778
10230	0259	3.2	6HRE	Tau (vivo)	507	1.15	60	4.74	-1.1	C1	776
11785	18883	2.8	8R4A	Lysozyme (vivo)	2,013	1.04	49	4.78	-1.4	C1	614
11383	15224	2.8	8A7Q	b2m-V27M	611	0.94	43	4.85	-0.73	C2	1,196
12815	70557	2.9	9OKA	AL (vivo)	7,076	0.90	40	4.86	+1.2	C1	729
12217	15361	2.8	8ADE	ATTR (vivo)	1,943	1.04	45	4.80	-1.2	C1	720
12909*	71953	3.2	9PX6	ATTR (vivo) wt1	1,910	0.94	50	4.85	-1.26	C1	693
12911*	71960	3.4	9PX7	ATTR (vivo) wt2	2,426	0.94	50	4.85	-1.28	C1	682
12912*	71962	3.0	9PX9	ATTR V122I (vivo)	2,326	0.94	50	4.85	-1.24	C1	704

capabilities as a complementary platform for amyloid reconstruction, several targeted improvements would expand its utility alongside existing tools like RELION. Key improvements include implementing non-cylindrical starting-model generation using projection-level information from overlapping 2D class averages, similar to relion_helix_inimodel2d in RELION[7] or denovo3d in Helicon[34]. Coupled with automatic clustering of 2D class averages, analogous to RELION's FilamentTools[14], this would create more homogeneous particle groups, helping overcome the common problem of convergence to local refinement minima by reducing the search range for initial model generation while minimizing user error in class selection[7,9]. While large-scale polymorphs such as single and double protofilaments of light-chain amyloids (Fig. S1,S2) were readily classified at the projection level, previous studies on light-chain Lambda6[27] demonstrated that double protofilaments contain additional peptide-level morphologies that require further 3D classification. These polymorphs and structural intermediates likely play crucial roles in amyloid disease progression[35] and have been successfully dissected using time-course cryo-EM experiments[14]. Separation and high-resolution reconstruction of ex vivo amyloid mixtures is essential for therapeutic development; however, cryoSPARC currently lacks 3D classification with helical rise and twist optimization, making separation of unique but similar filament types difficult or impossible. Implementing classification routines with helical symmetry optimization, both with alignment in Heterogeneous Refinement and without alignment in 3D Classification jobs, would enable broader access to amyloid structure determination using cryoSPARC. Re-processing cryo-EM datasets can reveal unexpected findings, such as vault particles discovered in ex vivo tau samples[36]. This proved true for the deposited lysozyme amyloids associated with systemic ALys amyloidosis as well. Through 2D classification, we identified previously unreported non-amyloid fibrils, which we reconstructed to 4 Å resolution, revealing a hexameric tubular structure with β-sheets stabilizing inter- and intra-protomer interfaces (Fig. 4C). Due to limited complementary data and resolution constraints, incomplete side-chain information prevented identification of the protein forming these tubular structures. We speculate this helical assembly may represent a disease-relevant intermediate in ALys amyloidosis pathogenesis, though further investigation is required for confirmation. Our study presents current capabilities and limitations of cryoSPARC for amyloid reconstruction using optimized non-default parameters. The established guidelines should enable reconstruction of diverse amyloids, though further amyloid-specific software optimization is needed to advance cryoSPARC's potential as a universal platform for helical reconstructions. These community-driven guidelines aim to democratize amyloid structure determination, enabling broader participation in amyloidosis research and accelerating therapeutic intervention development.

Acknowledgements

Funding for this research was provided by a NINDS grant NS095892 to GCL and the German Research Council project number 556478029 to JHS. The authors thank Jean-Christophe Ducom and Charles Bowman at Scripps for computational support. We thank Olga Gursky and Chad W. Hicks for sharing a pre-release of their EMPIAR-12815 data.

Data & software availability

All re-processed datasets from the Electron Microscopy Public Image Archive (EMPIAR) were deposited in the Electron Microscopy Data Bank (EMDB). The datasets from Table S1 are available using the following EMDB accession numbers with their respective EMPIAR-ID: ATTR wt1 (EMD-71953, EMPIAR-12909), ATTR wt2 (EMD-71960, EMPIAR-12911), ATTR V122I (EMD-71962, EMPIAR-12912), Unknown (EMD-72138, EMPIAR-11785), AL-L6 (EMD-72157, EMPIAR-19992), ATTR (EMD-72158, EMPIAR-12217), b2M V27M (EMD-72159, EMPIAR-11383), Lys D87G (EMD-72161, EMPIAR-11785), Tau MIA (EMD-72164, EMPIAR-10940), Tau PHF (EMD-72174, EMPIAR-10230). An import-ready .json workflow for cryoSPARC (generated in v.4.7.0) is available on https://github.com/schaefer-jh/CS-amyloids/ and can be used per the developers' instructions (https://guide.cryosparc.com/application-guide-v4.0+/workflows).

Competing interests

The authors declare no competing interests.

References

[1] McManus, J. J., Charbonneau, P., Zaccarelli, E., and Asherie, N. *Current Opinion in Colloid & Interface Science* **22**, 73–79 4 (2016).

[2] Ghosal, D. and Löwe, J. *The EMBO journal* **34**, 2312–20 9 (2015).

[3] Iadanza, M. G., Jackson, M. P., Hewitt, E. W., Ranson, N. A., and Radford, S. E. *Nature reviews. Molecular cell biology* **19**, 755–773 12 (2018).

[4] Knowles, T. P. J., Vendruscolo, M., and Dobson, C. M. *Nature reviews. Molecular cell biology* **15**, 384–96 6 (2014).

[5] 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., and Scheres, S. H. W. *Nature* **587**, 152–156 11 (2020).

[6] Fitzpatrick, A. W. P., Falcon, B., He, S., Murzin, A. G., Murshudov, G., Garringer, H. J., Crowther, R. A., Ghetti, B., Goedert, M., and Scheres, S. H. W. *Nature* **547**, 185–190 7 (2017).

[7] He, S. and Scheres, S. H. W. *Journal of structural biology* **198**, 163–176 6 (2017).

[8] Taylor, A. I. P. and Staniforth, R. A. *Frontiers in neuroscience* **16**, 878869 (2022).

[9] Lövestam, S. and Scheres, S. H. W. *Faraday discussions* **240**, 243–260 11 (2022).

[10] Zielinski, M., Röder, C., and Schröder, G. F. *The Journal of biological chemistry* **297**, 100938 8 (2021).

[11] Radamaker, L., Karimi-Farsijani, S., Andreotti, G., Baur, J., Neumann, M., Schreiner, S., Berghaus, N., Motika, R., Haupt, C., Walther, P., Schmidt, V., Huhn, S., Hegenbart, U., Schönland, S. O., Wiese, S., Read, C., Schmidt, M., and Fändrich, M. *Nature communications* **12**, 6434 11 (2021).

[12] Zhao, K., Lim, Y.-J., Liu, Z., Long, H., Sun, Y., Hu, J.-J., Zhao, C., Tao, Y., Zhang, X., Li, D., Li, Y.-M., and Liu, C. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 20305–20315 8 (2020).

[13] Kimanius, D., Dong, L., Sharov, G., Nakane, T., and Scheres, S. H. W. *The Biochemical journal* **478**, 4169–4185 12 (2021).

[14] Lövestam, S., Li, D., Wagstaff, J. L., Kotecha, A., Kimanius, D., McLaughlin, S. H., Murzin, A. G., Freund, S. M. V., Goedert, M., and Scheres, S. H. W. *Nature* **625**, 119–125 1 (2024).

[15] Punjani, A., Rubinstein, J. L., Fleet, D. J., and Brubaker, M. A. *Nature methods* **14**, 290–296 3 (2017).

[16] Sawaya, M. R., Hughes, M. P., Rodriguez, J. A., Riek, R., and Eisenberg, D. S. *Cell* **184**, 4857–4873 9 (2021).

[17] Hansen, K. H., Byeon, C. H., Liu, Q., Drace, T., Boesen, T., Conway, J. F., Andreasen, M., and Ümit Akbey. *Proceedings of the National Academy of Sciences of the United States of America* **121**, e2406775121 8 (2024).

[18] Hansen, K. H., Golcuk, M., Byeon, C. H., Tunc, A., Plechinger, E. B., Dueholm, M. K. D., Conway, J. F., Andreasen, M., Gur, M., and Ümit Akbey. *Science Advances* **11** 9 (2025).

[19] Hicks, C. W., Prokaeva, T., Spencer, B., Jayaraman, S., Huda, N., Wong, S., Chen, H., Sanchorawala, V., Lavatelli, F., and Gursky, O. *bioRxiv : the preprint server for biology* 7 (2025).

[20] Huber, S. T., Kuhm, T., and Sachse, C. *Journal of structural biology* **202**, 1–12 4 (2018).

[21] Egelman, E. H. *Journal of structural biology* **157**, 83–94 1 (2007).

[22] Pothula, K. R., Geraets, J. A., Ferber, I. I., and Schröder, G. F. *Progress in biophysics and molecular biology* **160**, 16–25 3 (2021).

[23] Sun, C., Gonzalez, B., and Jiang, W. *Scientific reports* **12**, 8162 5 (2022).

[24] Li, D., Pérez, M. M., Zhang, X., Li, J., and Jiang, W. *bioRxiv : the preprint server for biology* 5 (2025).

[25] Cao, H., Li, T., He, J., and Huang, S. *Biophysical Journal* **124**, 625a 2 (2025).

[26] Annamalai, K., Gührs, K., Koehler, R., Schmidt, M., Michel, H., Loos, C., Gaffney, P. M., Sigurdson, C. J., Hegenbart, U., Schönland, S., and Fändrich, M. *Angewandte Chemie International Edition* **55**, 4822–4825 4 (2016).

[27] Bassett, P., Nguyen, B. A., Singh, V., Garrett, P. T., Moresco, J. J., Kaur, G., Afrin, S., Pekała, M., and Saelices, L. 12 (2024).

[28] Pintilie, G., Zhang, K., Su, Z., Li, S., Schmid, M. F., and Chiu, W. *Nature methods* **17**, 328–334 3 (2020).

[29] Dumoulin, M., Kumita, J. R., and Dobson, C. M. *Accounts of chemical research* **39**, 603–10 9 (2006).

[30] Jamali, K., Käll, L., Zhang, R., Brown, A., Kimanius, D., and Scheres, S. H. W. *Nature* **628**, 450–457 4 (2024).

[31] Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., and Ferrin, T. E. *Protein Sci.* **30**, 70–82 1 (2021).

[32] Liebschner, D., Afonine, P. V., L, G. B. M., Bunkóczi, Chen, V. B., Croll, T. I., Hintze, B., Hung, L. W., Swati, A. J. J., McCoy, Moriarty, N. W., D, B. K. O. R., Poon, Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J., and Adams, P. D. *Acta Crystallogr. D Struct. Biol.* **75**, 861–877 10 (2019).

[33] Schäfer, J.-H., Calza, A., Hom, K., Damodar, P., Peng, R., Bogdanović, N., Lander, G. C., Stagg, S. M., and Cianfrocco, M. A. 8 (2025).

[34] Li, D., Zhang, X., and Jiang, W. 4 (2025).

[35] Tycko, R. *Neuron* **86**, 632–45 5 (2015).

[36] Lövestam, S. and Scheres, S. H. *Structure* 8 (2025).

Supporting Information

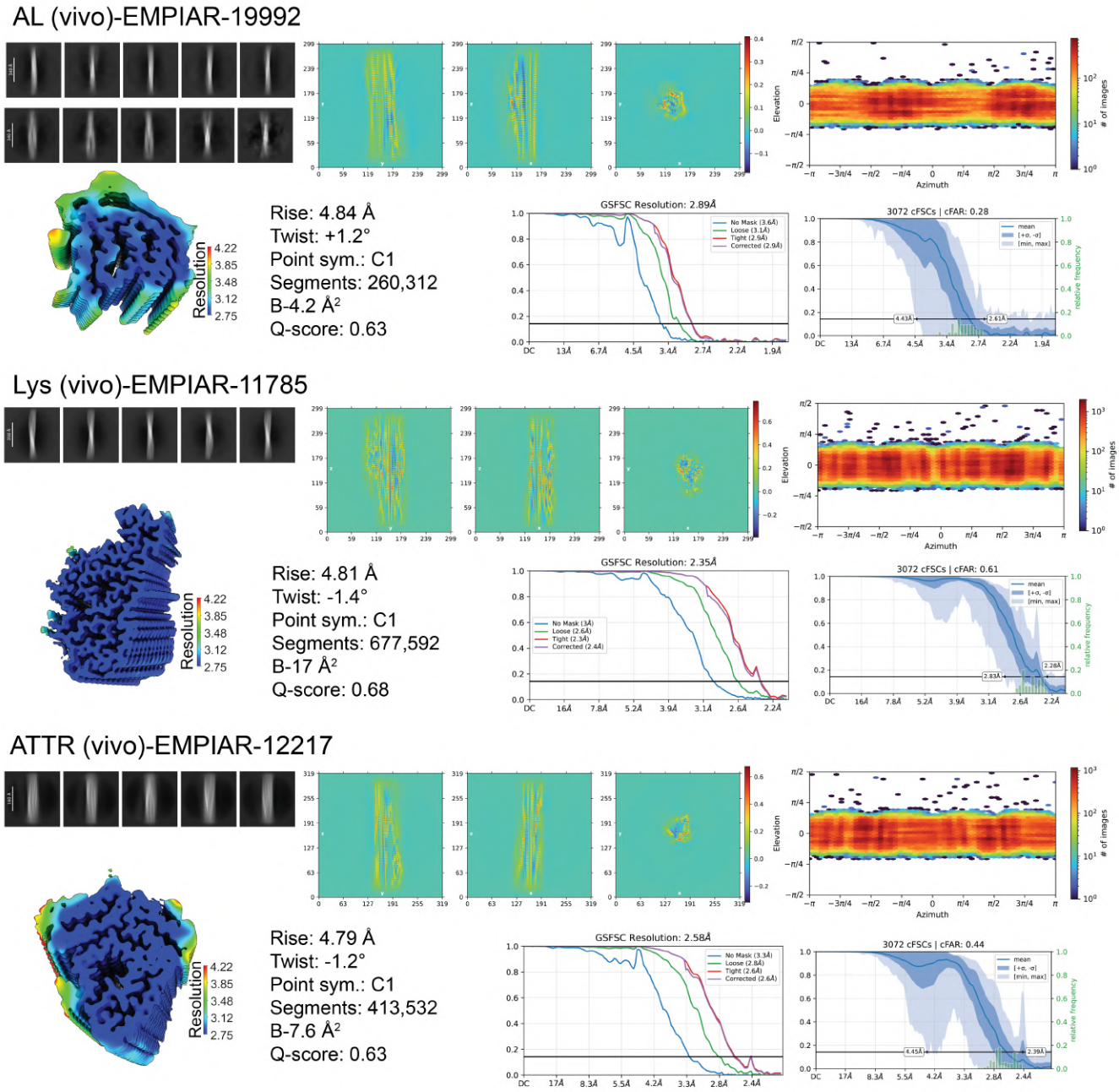


Figure S1. Validation of amyloid reconstructions using cryoSPARC. Comprehensive validation for each EMPIAR dataset includes 2D class averages, real-space slices along major orthogonal planes (zy, zx, xy), viewing directional distribution, local resolution maps (color-coded consistently across all datasets), resolution estimated using gold-standard FSC (cutoff at 0.143), and cFAR reports.

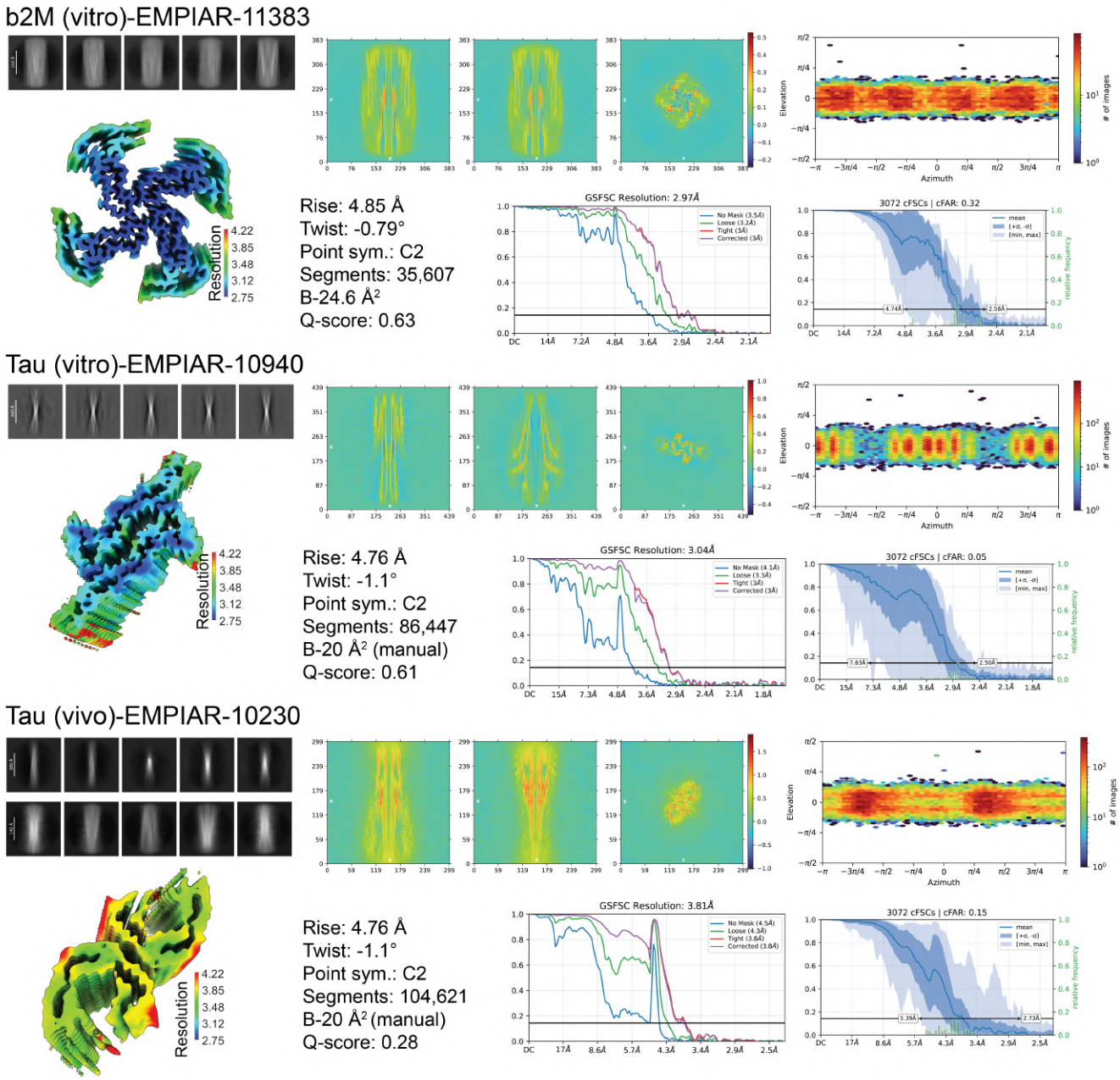


Figure S2. (continued).

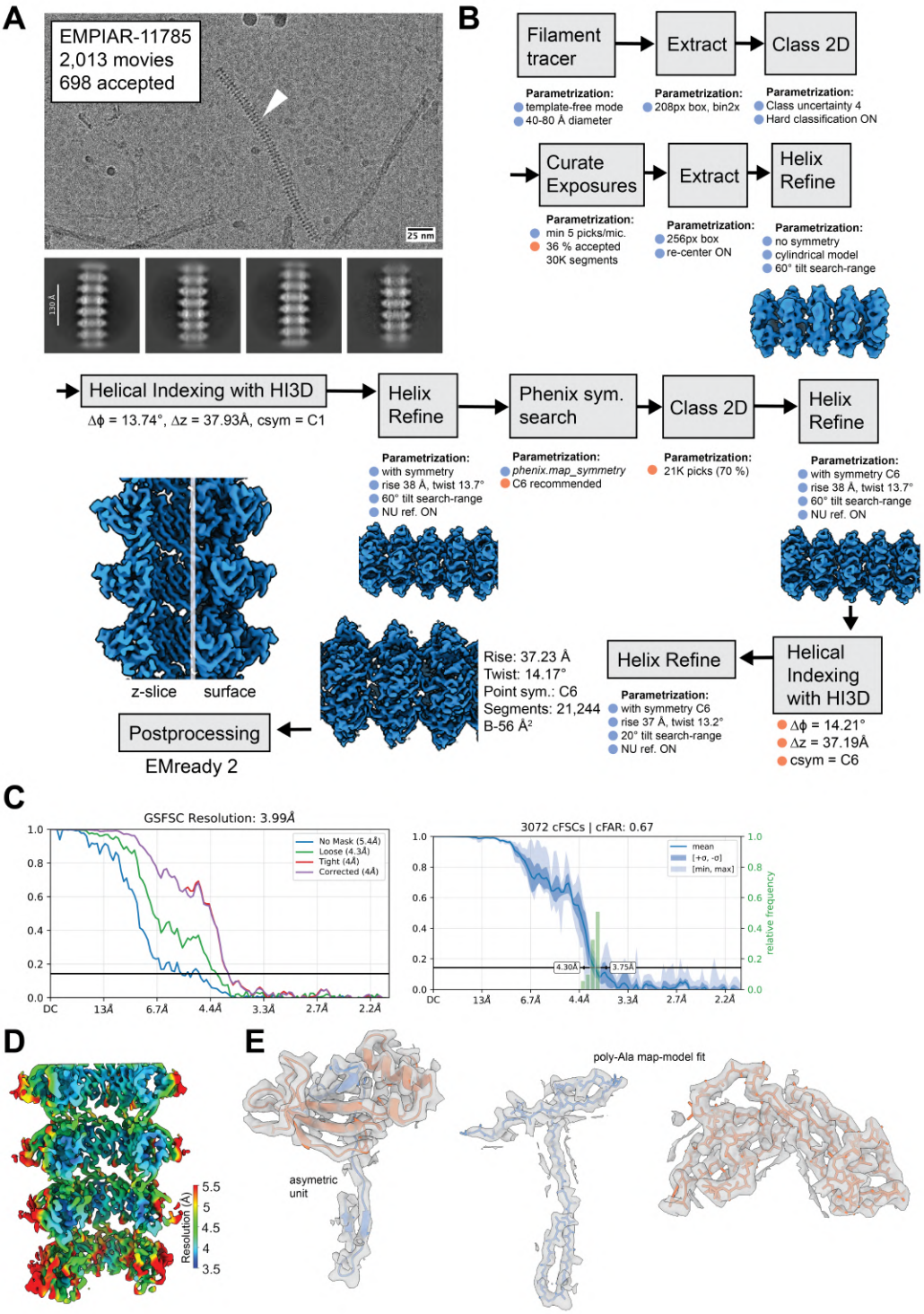


Figure S3. Validation of non-amyloid fibrils. (A) Micrograph and 2D class averages from EMPIAR-11785 positive for the non-amyloid filament (white arrow). 25 nm scale bar on micrograph and 130 Å scale bar on 2D class averages. (B) Processing workflow for helical reconstruction. (C) Gold-standard FSC resolution estimation at 0.143

Table S1. Data collection and processing parameters for ATTR and Tau filaments. Columns highlighted in green.

	ATTR wt1	**ATTR wt2**	**ATTR V122I**	**Unknown**	**AL-L6**	**ATTR**	**b2M V27M**	**Lys D87G**	**Tau MIA**	**Tau PHF**
EMDB ID	71953	71960	71962	72138	72157	72158	72159	72161	72164	72174
EMPIAR ID	12909	12911	12912	11785	12815	12217	11383	11785	10940	10230
Microscope	Arctica	Arctica	Arctica	Krios	Arctica	Krios	Krios	Krios	Krios	Krios
Camera	Falcon 4i	Falcon 4i	Falcon 4i	K2 Quantum	Falcon 4i	K2 Quantum	Falcon 4i	K2 Quantum	Falcon 4i	K2 Summit
Magnification (nom.)	150,000	190,000	190,000	130,000	130,000	130,000	130,000	130,000	96,000	105,000
Voltage (keV)	200	200	200	300	200	300	300	300	300	300
Total dose ($e^-/Å^2$)	50	50	50	49	40	45	43	49	40	60
Exposure rate (e^-/px/s)	11.07	11.07	10.81	5.9	10.82	3.1	7.6	5.9	–	13.8
Frames per movie	1,395	1,395	1,386	40	–	40	1,442	40	–	50
EER Fractions	40	40	40	NA	40	NA	40	NA	40	NA
Pixel size (Å/px)	0.94	0.94	0.94	1.04	0.9	1.04	0.94	1.04	0.824	1.15
Defocus range (µm)	-2.0 to -0.8	-2.0 to -0.8	-2.0 to -0.8	-1.0 to -2.0	-0.8 to -2.5	-1.2 to -2.5	-1.3 to -2.5	-1.0 to -2.0	-1.0 to -3.5	-1.7 to -2.8
Recorded movies	3,289	6,232	6,063	2,013	7,076	1,943	611	2,013	331	507
Processing software	CS (v.4.6)	CS (v.4.6)	CS (v.4.6)	CS (v.4.7)	CS (v.4.7)	CS (v.4.7)	CS (v.4.7)	CS (v.4.7)	CS (v.4.7)	CS (v.4.7)
Final particle images	171,067	231,893	335,693	21,244	260,312	413,532	35,607	677,592	86,447	104,621
Symmetry imposed	C1	C1	C1	C6	C1	C1	C2	C1	C2	C2
Helical twist (°)	-1.26	-1.28	-1.24	+14.17	+1.2	-1.2	-0.79	-1.4	-1.1	-1.1
Helical rise (Å)	4.85	4.85	4.85	37.23	4.84	4.79	4.85	4.81	4.76	4.76
Res. (FSC 0.143, Å)	3.2	3.4	3.0	4.0	2.9	2.6	3.0	2.4	3.0	3.8
Local res. range (Å)	2.7–4.9	3.0–5.2	2.6–4.8	3.8–4.3	2.6–4.4	2.4–4.5	2.6–4.7	2.3–2.8	2.5–7.6	2.7–5.4
cFAR	0.31	0.27	0.28	0.67	0.28	0.44	0.32	0.61	0.05	0.15
Sharpening B-factor ($Å^2$)	-46	-68	-36	-56	-4	-8	-25	-17	-20	-20