Journal of Structural Biology: X 14 (2026) 100169

Contents lists available at ScienceDirect

Journal of Structural Biology: X

journal homepage: www.sciencedirect.com/journal/journal-of-structural-biology-x

Tailocins are recurrent contaminants in ex vivo amyloid extraction

Jan-Hannes Schaefer [a,b,*], Robert T. O'Neill [b], Zi Gao [b], Danielle A. Grotjahn [a], Evan T. Powers [b], Gabriel C. Lander [a], Jeffery W. Kelly [b]

[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

ARTICLE INFO

Keywords:
Tailocin
Helical reconstruction
Contamination
cryoSPARC

ABSTRACT

Amyloid fibrils extracted from patient tissue are enriched by physicochemical fractionation methods and enzymatic tissue digestion with collagenase, resulting in a mixture of sedimented complexes. Reprocessing of an ex vivo lysozyme amyloid dataset revealed non-amyloid tubular assemblies, which we reconstructed to 3.5 Å. Sequence-agnostic model building with CryoAtom and fold search with FoldSeek identified a phage tail tube fold. Testing the crude *Clostridium histolyticum* collagenase used for tissue digestion, negative-stain TEM and LC-MS/MS showed proteins of an F-type tailocin, allowing us to assign the helical reconstruction to a major tail tube protein (TTP) of *C. histolyticum*. We also observed tailocin contamination in cardiac light-chain amyloid extracts, following the same collagenase treatment. Overall, we identify tailocins as recurrent reagent-borne contaminants in collagenase-treated ex vivo amyloid preparations and propose strategies to detect and avoid such artefacts in single-particle cryo-EM by optimizing the experimental design and data annotation toward a database of common contaminants.

1. Introduction

Extraction of disease-relevant amyloid fibrils from patient tissue has become the standard for cryo-electron microscopy studies, linking fibril morphology to a specific pathology (Lashuel, 2026; Scheres et al., 2023). While amyloid fibrils from patients with amyloidosis of the central nervous system can generally only be obtained from postmortem specimens, systemic amyloids can be extracted from living patients following organ transplants, operative biopsy, or lately punch-biopsies (Nguyen et al., 2026; Yao et al., 2026). Transthyretin amyloidosis (ATTR) and Light-Chain amyloidosis (AL) manifest across several organ systems, allowing patient-derived structures to now span hereditary and wild-type disease, multiple organs from single individuals, and different polymorphs (Iakovleva et al., 2021; Nguyen et al., 2024; Schmidt et al., 2019; Steinebrei et al., 2023). Established amyloid extraction protocols utilize tissue digestion with collagenase, a step introduced by Annamalai and colleagues (Annamalai et al., 2016) as a modification of the classical Pras protocol (Schubert et al., 1968) and subsequently adopted throughout the field (Fändrich and Schmidt, 2021; Nguyen et al., 2024; Schmidt et al., 2019; Steinebrei et al., 2023; Steinebrei et al., 2022). Extraction protocols traditionally use collagenase from *Clostridium histolyticum* culture filtrates rather than purified enzyme preparations, on the assumption that non-collagenolytic components are either inert or removed during downstream sample preparation.

Reprocessing cryo-EM datasets can reveal unexpected findings, as demonstrated by the identification of vault particles in ex vivo tau samples (Lövestam and Scheres, 2025). Similarly, this approach proved informative for deposited datasets of lysozyme amyloids associated with systemic lysozyme amyloidosis (ALys). Here, we present the helical reconstruction of the major tail tube protein (TTP) of a tailocin from *Clostridium histolyticum,* which was co-purified with amyloid fibrils extracted from patient tissue affected by lysozyme amyloidosis. Using the same collagenase treatment and extraction approach, we identified tailocins in cardiac amyloid extracts from a patient with light-chain amyloidosis, showing that the same reagent-derived contaminant can recur across distinct ex vivo amyloid preparations that use *Clostridium histolyticum* collagenase.

2. Results and discussion

2.1. Structure of a Clostridium histolyticum tailocin

During re-processing of an ex vivo lysozyme variant D87G (Karimi-Farsijani et al., 2024) dataset (EMPIAR-11785), associated with ALys,

* Corresponding author at: Department of Integrative Structural and Computational Biology, Scripps Research; La Jolla, CA, USA.
E-mail address: jschaefer@scripps.edu (J.-H. Schaefer).

https://doi.org/10.1016/j.yjsbx.2026.100169
Received 22 June 2026; Received in revised form 29 October 2026; Accepted 2 November 2026
Available online 7 November 2026
2590-1524/© 2026 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

we identified non-amyloid fibrillar assemblies by 2D classification (Fig. 1A). Approximately one-third of micrographs contained these distinct fibrillar structures. Helical refinement using a rise of 37.23 Å and twist of 14.17° (**Fig. S1**) yielded a 3.5 Å resolution reconstruction of 10 nm diameter tubes with a C6-symmetric protofilament arrangement and an inner diameter of 3.5 nm (Fig. 1B).

To identify the assembly, we used automated model building with CryoAtom (Su et al., 2026), with the human proteome (UP000005640) as a reference database. No convincing matches were identified, arguing against our initial interpretation that the fibrils represented stacks of common amyloid-associated proteins, such as serum amyloid P component (SAP), or intermediates of lysozyme amyloid formation. We then used the fitted poly-alanine backbone trace from automated model building as input for Foldseek (Van Kempen et al., 2024) to identify proteins with similar folds, which indicated the presence of phage tail tube proteins (TTPs) with *Siphoviridae*-like morphology. Because the global resolution was limited to approximately 3.5 Å, the structure-to-sequence assignment remains probabilistic and relies in part on alignment of aromatic residues (**Fig. S2E**), but the high map-model fit, assessed with the Q-score, supports our assignment (**Fig. S2D**).

To test collagenase as the source of contamination, we performed a mock fibril extraction of the commercial *C. histolyticum* collagenase used for the lysozyme amyloid (EMPIAR-11785) and light-chain amyloid extraction. We assumed a reasonably consistent composition of the bacterial secretion used for the manufacturing of collagenase across batches. Our enrichment was designed to retain higher-molecular-weight assemblies with sedimentation properties similar to amyloid fibrils. Subsequent negative-stain transmission electron microscopy (nsTEM) of the enriched material revealed abundant phage tail-like assemblies that resembled the non-amyloid fibrils observed in the lysozyme amyloid dataset. Further processing of the nsTEM data yielded 2D class averages similar to those observed in our cryo-EM data (Fig. 2A). Power-spectrum analysis of the prominent layer line revealed a helical rise of about 34 Å, close to the 37 Å rise from our previous cryo-EM reconstruction (Fig. 2E, left panel). Together, these results indicate collagenase as the reagent-based source of the ALys contamination.

To identify the present contaminant, we constructed a candidate sequence database from the *C. histolyticum* proteome (**Table S2**) and inspected the corresponding predicted structures using AlphaFold3 (**Fig. S3**). Subsequently, we performed automatic model building with CryoAtom and the three available *C. histolyticum* proteomes as input. Resulting models were used for another round of FoldSeek. Manual inspection of the FoldSeek targets with the highest confidence scores (**Fig. S3**) supported assignment of the structure to a *C. histolyticum* major tail tube protein MttP, hereafter called TTP, which is part of an F-type non-contractile tailocin element (UniProt: A0A4U9RSA4, Fig. 2B).

The same enriched sample and a control were subjected to tryptic digestion and mass spectrometry. Bottom-up LC-MS/MS identified 60 proteins from *C. histolyticum*. Common contaminants were excluded. We detected six proteins encoded in the F-type tailocin locus NCTC503_02168–02181 (Fig. 2C-D), and all were enriched after mock purification of amyloids. The major tail protein TTP (A0A4U9RSA4), which we characterized structurally, was enriched 18-fold ($\log_2$ fold change 4.18, adjusted $p = 0.028$), with 69% sequence coverage by MS/MS from 20 peptides (**Fig. S4**). The subunits assigned by inferred homology with monocin (Lee et al., 2016; Sigal et al., 2024; Gu et al., 2025) to the side fiber / receptor-binding protein (02171), the baseplate hub (02172), the distal tail protein (02173) and the tape measure protein (02174) were enriched 13- to 17-fold, reaching significance for 02171, 02173 and 02174 (adjusted $p = 0.034$, 0.0073 and 0.0091); the comparable enrichment of 02172 (13-fold) did not (adjusted $p = 0.22$). Locus 02177 showed the largest point estimate of the cluster (20-fold) but likewise did not reach significance (adjusted $p = 0.21$); the quantification of 02177 and of 02174 depended in part on imputed values (Fig. 2C). The six subunits of a single contiguous locus move together over a narrow 13- to 20-fold range while spanning an approximately 30-fold range in absolute abundance, consistent with their co-enrichment as components of one particle. Rather than the assembly chaperone suggested by its position in the monocin operon, 02177 encodes a second tail-tube copy (TTP*) that shares 88.1% sequence identity with locus 02179 (TTP) **Fig. S4B**. Local map-model fit analysis of non-conserved positions using the Q-score supports locus 02179, rather than 02177, as the main TTP in the cryo-EM reconstruction (**Fig. S4B**).

Tailocins are phage tail-like bacteriocins and act like molecular needles, evolved to puncture membranes of competitor cells under nutrient-scarce conditions. Bacteria encode one or rarely two gene clusters for tailocins. Lytic release of tailocins occurs in a small fraction of cells, gaining overall population fitness. These capsid-free phage tail-like assemblies consist of a tail tube and baseplate with fibers selective for cell surface markers like the O-antigens of lipopolysaccharides (Granato et al., 2019; Patz et al., 2019; Scholl, 2017). Since the observed structure lacks a sheath, it can be categorized as a non-contractile F-type tailocin, structurally related to tails of phages of the *Siphoviridae* family (Backman et al., 2024). Limited by the number of tailocins, we were only able to reconstruct the TTP as part of the structural gene module (Fig. 2B,D). While visible at the micrograph level and in 2D class averages, the tail cap (Fig. 1A, panel 1) and tail tip (Fig. 1A, panel 3) were not reconstructed. Each protomer within the hexameric TTP stack contains a conserved β-sandwich domain that lines the central channel and an extended β-hairpin involved in inter-stack interactions (Fig. 1C). Overall, we identified a commercial enzyme widely used in amyloid research as a source of contamination with bacterial tailocin particles.

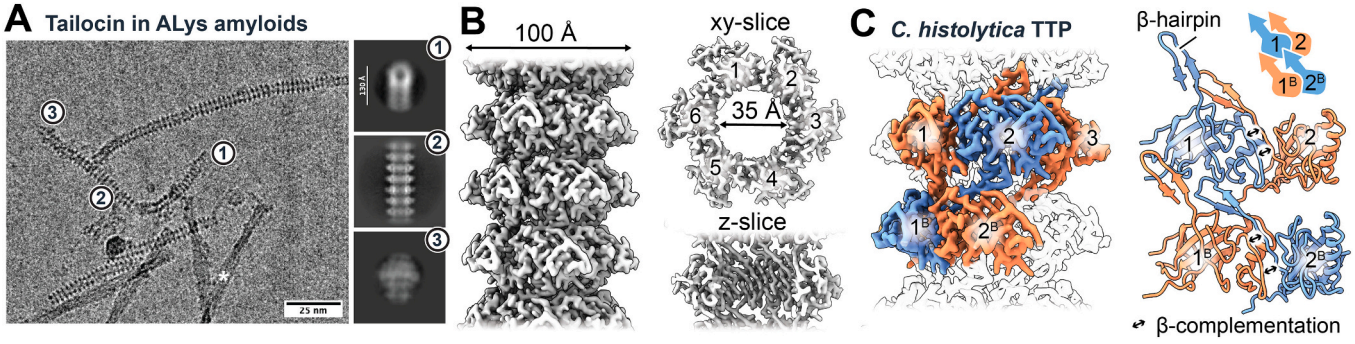


Fig. 1. Reconstruction of a major tail tube protein from a *Clostridium histolyticum* tailocin. (**A**) Micrograph from EMPIAR-11785 showing 2D class averages of the tail cap region (1), tail tube (2), and tail tip (3) of a tailocin. Asterisk indicates ALys amyloid. Scale bars: micrograph 25 nm, 2D class 130 Å. (**B**) 3.5 Å helical reconstruction of 10 nm C6-symmetric tubes with a 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. Inter-protofilament contacts through β-complementation are indicated by black double arrows.

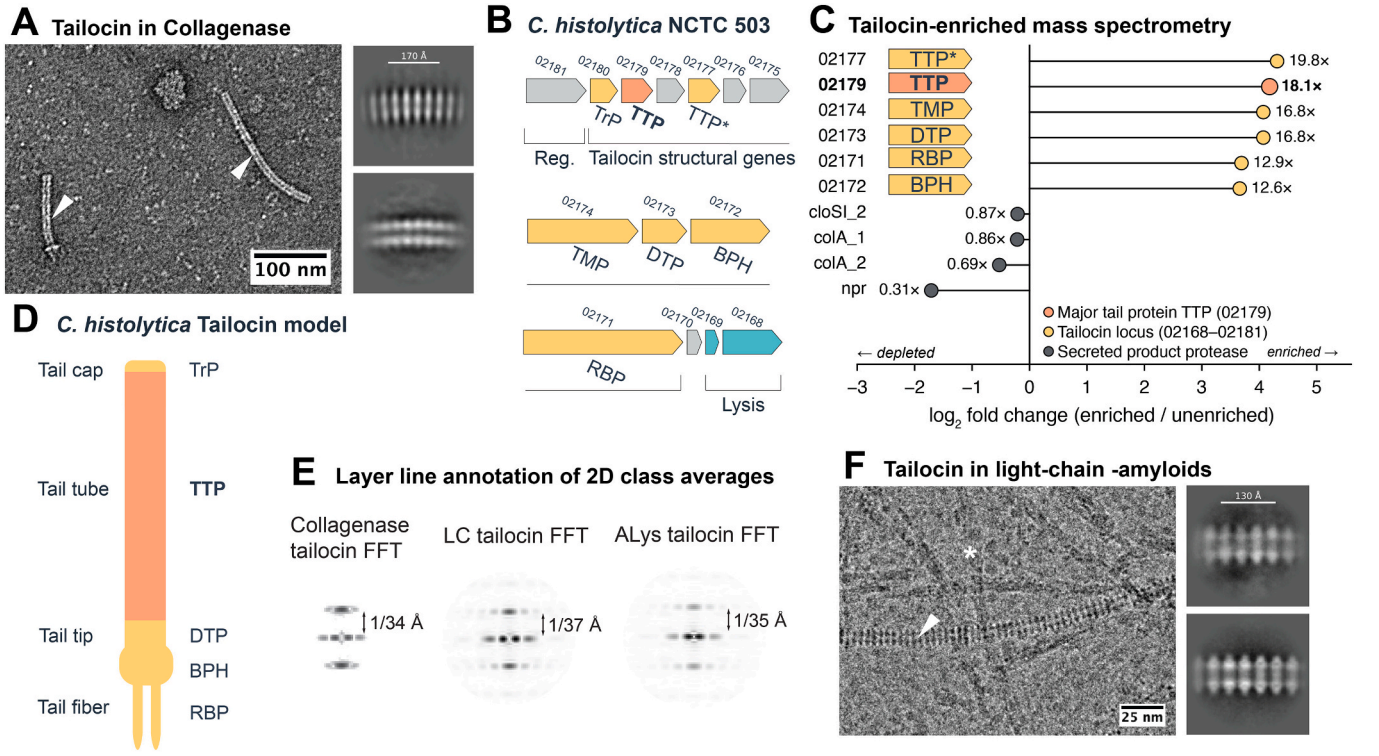


Fig. 2. Tailocins are a recurrent contaminant in collagenase-treated ex vivo amyloid extracts. **(A)** Negative-stain TEM micrograph and 2D class averages of tailocin (white arrow) from crude collagenase (Sigma-Aldrich C2674). Scale bars: 100 nm on micrograph, 170 Å on 2D class. **(B)** Schematic representation of a *C. histolyticum* NCTC 503 tailocin element. Grey elements are of unassigned function; yellow indicates structural genes, and genes involved in lysis are in mint. The tail tube protein (TTP) is highlighted in orange. A gene duplicate of the TTP is annotated as TTP*. Tail terminator TrP, tape measure protein TMP, baseplate hub BPH, and tail tip/side fibers RBP are annotated. **(C)** $\log_2$ fold change (limma, two replicates per condition) for all tailocin locus proteins (yellow), the major tail tube protein (orange) and secreted product-protease proteins (gray). Real fold-change given as values. Structural tailocin genes annotated according to Fig. 2B. **(D)** Schematic representation of a tailocin particle with the TTP highlighted in orange. **(E)** Comparison of contrast-inverted power spectra of the nsTEM collagenase tailocin (left), cryo-EM of light-chain tailocin (middle panel) and lysozyme amyloids ALys-derived tailocin (right) 2D class averages using the *relion_display* function in RELION-5 (Burt et al., 2024). The annotated layer-line approximates the helical rise of the helical assembly. **(F)** Micrograph of light-chain amyloids (asterisk) and tailocin (arrow). Scale bar 25 nm. 2D class averages of phage tail fibrils, scale bar 130 Å.

2.2. Clostridium histolyticum tailocin is a recurrent contaminant in amyloid extracts

We next asked whether tailocins are a recurrent contaminant in ex vivo amyloids. To address this, we analyzed cardiac light-chain amyloids (AL), which cause systemic amyloidosis and arise from the overproduction of immunoglobulin light-chain fragments. Clinically, AL amyloidosis can affect the heart, digestive tract, kidneys, nervous system, and soft tissue (Al Hamed et al., 2021).

Following an established collagenase-based extraction protocol, we enriched light-chain amyloids from the apex of a heart explanted from an AL amyloidosis patient for cryo-EM. Using reference-based particle picking, we also identified tailocin in the AL dataset (Fig. 2F). While nearly one-third of ALys micrographs contained observable tailocins, only approximately 5% of our AL micrographs contained the tailocins (Fig. 2F). These percentages were estimated by linking particles assigned to tailocin 2D class averages back to their original micrographs. Because of the limited number of particles in the AL dataset, we were unable to reconstruct the AL-associated major tail-tube protein. Using the 2D class averages of the proposed tailocin to generate an average power spectrum in Fourier space, we estimated a helical rise of approximately 34 Å for collagenase tailocins (Fig. 2E, left panel). The same approach yielded a rise of about 35 Å for ALys-derived tailocin and approximately 37 Å for light-chain-derived tailocin (Fig. 2E, right and middle panel, respectively), further supporting our assessment of tailocin contamination from the use of crude *C. histolyticum* secretion in ex vivo amyloid extracts.

2.3. Strategies to avoid artefacts in cryo-EM

Modern single-particle analysis (SPA) cryo-EM data processing enables reconstruction of minimally processed samples, in many cases omitting affinity-based purification steps entirely. This capability can reveal transient or unstable complexes, but minimally invasive extraction methods also increase the risk that reagent- or preparation-derived artefacts will be overinterpreted as biological findings. Even when sample heterogeneity is deliberately explored, unexpected results should be carefully assessed within the context of sample preparation before they are interpreted as novel biological assemblies. The following set of strategies may help the cryo-EM community identify contaminants, avoiding mis- or overrepresentation of new structures:

Reagent-only controls: Mock-protocols or extraction blanks are widely used in sequence-based fields (Salter et al., 2014), as they rely on low-input samples that carry a high risk of reagent-derived signals being misinterpreted as novel biology (Duan et al., 2024; Naccache et al., 2013). Current SPA cryo-EM pipelines lack an equivalent but would benefit from extraction blanks, implemented in nsTEM prescreening. Additionally, testing lot-level reagent quality is advisable, as production quality, especially for biologics, can vary.

Genetically encoded contaminants: Host-derived proteins can mimic affinity tags, competing with the intended target. Documented examples are *E. coli* ArnA and AcrB as enriched contaminants in immobilized metal affinity chromatography (Belcher et al., 2026; Caliseki et al., 2025). Alternative purification routes, such as ion-exchange chromatography or tandem affinity purification, have proven effective amongst

several other strategies.

Shared properties enrich contaminants: Enrichment strategies involving detergent insolubility (Jiang et al., 2022), protease resistance (Biasini et al., 2009), or a high sedimentation coefficient (Kedersha and Rome, 1986; Liu et al., 2025; Lövestam and Scheres, 2025) will co-purify abundant macromolecules with shared properties of the intended target molecule. Consequently, a protocol selective for physical properties should be described as enriching a class of particles, not a single target.

Sequence-agnostic model building: Sequence-agnostic backbone tracing followed by map-based sequence identification allows an unbiased map-model fit from a defined candidate library, covering the proteome of the expression host and intended target sequence (Jamali et al., 2024; Su et al., 2026; Van Kempen et al., 2024). This approach is especially valuable for cases in which the composition of the sample is unknown in advance or for unexpected contaminants or co-binders of the intended target (Ho et al., 2020; Kyrilis et al., 2021; Skalidis et al., 2022). The ranked candidate list should be reported in full, including candidates of high backbone but low per-residue map-to-model fit. In cases of limited resolution, probabilistic sequence assignment and fold searches with FoldSeek (Van Kempen et al., 2024) can be explored. For the reported tail-tube fold, which is shared across siphoviral tails and their bacteriocin derivatives (Zinke et al., 2022), a single fold assignment can implicate an entire class of contaminants.

Annotate deposited maps as contaminants, toward a contamination database for cryo-EM: Contaminant reconstructions are often discarded, causing laboratories to unknowingly rediscover them independently. Proteomics addressed this problem with a shared contaminant database, and crystallography with ContaMiner and ContaBase (Hungler et al., 2016). However, cryo-EM has no equivalent resource. Depositing identified contaminant maps to the EMDB with an explicit reagent-derived annotation, as has been done for a phage tail traced to a contaminated buffer (Jenkins et al., 2024) (PDB 9MFE), would convert the experience of individual research groups into a searchable community asset, allowing depositors to check new maps against known artefacts as a routine step.

3. Materials and methods

3.1. Amyloid light-chain extraction

This study, conducted with anonymized ex vivo human cardiac tissue samples obtained by a generous gift from Protego Bio from a donor who provided informed consent, was approved by the Institutional Review Board of Scripps Research (protocol No.: IRB-24-8382, continuing IRB-16-6864 and IRB-12-5994). Light-chain amyloids were extracted from human heart tissue using a water-based protocol (Annamalai et al., 2016). Briefly, 125–250 mg of cardiac tissue from the apex was cut, washed three times with 0.5 ml Tris-Calcium buffer, and incubated overnight with *C. histolyticum* collagenase (Sigma, Aldrich C2674, 0.5 mg/ml) at 37 °C. The sample was pelleted at 3100 x g for 30 min at 4 °C, and the supernatant was discarded. The pellet was resuspended in 0.5 ml Tris-EDTA, centrifuged, and the supernatant was discarded. After 10 total extraction cycles, the pellet was resuspended in 0.5 ml ice-cold water and stored at 4 °C.

3.2. Negative staining TEM of Clostridium histolyticum collagenase

Clostridium histolyticum collagenase (Sigma Aldrich, C2674) was resuspended in TBS buffer (50 mM Tris-HCl, 5 mM $CaCl_2$) at a final concentration of 4 mg/ml. To enrich higher-molecular-weight components, 0.8 ml sample was sedimented for 30 min at 200,000 x g; the pellet was washed with TBS buffer, resuspended in 80 µl TBS, and assessed by negative-stain electron microscopy following established protocols (Chowdhury et al., 2015). In brief, 100× and undiluted samples were applied to glow-discharged 400 square Mesh Copper/Rhodium Maxaform grids (Electron Microscopy Sciences) and stained with 2% uranyl formate. Micrographs were collected automatically with Leginon (Suloway et al., 2005) on a Thermo Fisher Talos F200C transmission electron microscope operating at 200 keV and equipped with a Gatan K2 direct electron detector at a nominal magnification of 73,000, corresponding to a pixel size of 1.981 Å. Micrographs were imported in cryoSPARC (v.4.7) with constant CTFs and inverted contrast. Templates were generated from manually picked particles. 2D classification without Force max over poses/shifts yielded well-aligned 2D averages.

3.3. Comparative power spectrum analysis

2D class averages of ALys, AL, as well as negative-stain data of *Clostridium histolyticum* collagenase non-amyloid fibrils were opened with the *relion_display* function in RELION-5 (Burt et al., 2024) and converted to their power spectrum. The first layer line distance was manually measured from the equator and converted to the helical rise value according to established protocols (Sanchez et al., 2025), using: rise [Å] = (box size [px] / FFT layer line distance) * 2 * pixel size [Å/px].

3.4. Mass-spectrometric characterization of Clostridium histolyticum collagenase

MS Sample preparation. Collagenase and mock enriched collagenase were prepared for bottom-up proteomic analysis by denaturation in TBS containing 2% (*w/v*) SDS. Cysteine residues were reduced with 5 mM tris (2-carboxyethyl)phosphine (TCEP) for 30 min at room temperature and subsequently alkylated with 10 mM iodoacetamide (IAA) for 30 min at room temperature in the dark. Proteins were digested overnight at 37 °C with trypsin at an enzyme-to-protein ratio of 1:100 (*w*/w). Digestion was terminated by addition of trifluoroacetic acid (TFA) to a final concentration of 2%, followed by incubation for 5 min at 37 °C with shaking at 800 rpm. Peptides were subsequently desalted using C18-based cleanup, dried by vacuum centrifugation, and reconstituted in 0.1% formic acid prior to LC–MS/MS analysis.

LC–MS/MS analysis. Peptides were analyzed by nanoLC–MS/MS using a Vanquish Neo UHPLC system coupled to an Orbitrap Exploris 480 mass spectrometer (Thermo Fisher Scientific) operated in data-dependent acquisition (DDA) mode. Peptides were loaded using a trap-and-elute configuration with a 300-µm × 5-mm trap column and separated on a 75-µm × 50-cm analytical column at a flow rate of 350 nL/min. Mobile phase A consisted of 0.1% FA (Formic acid) and mobile phase B consisted of 80% acetonitrile. The gradient began at 2% B, increased to 25% B over 75.1 min and to 38% B by 112.1 min, followed by a column wash at 99% B. The total method duration was 120 min. Peptides were ionized by nanospray ionization in positive-ion mode at 1.9 kV. MS1 spectra were acquired in the Orbitrap at a resolution of 120,000 over an m/z range of 320–1600 with a normalized AGC target of 300%. Precursors with charge states of 2–6 were selected for fragmentation using a 3 s cycle-time method, with dynamic exclusion of previously selected precursors for 30 s. MS/MS spectra were acquired following higher-energy collisional dissociation (HCD) using a normalized collision energy of 28%, a precursor isolation window of 1.6 m/z, an Orbitrap resolution of 15,000, a normalized AGC target of 100%, and a maximum injection time of 50 ms.

Database search and identification of *C. histolyticum*-derived peptides. Raw LC–MS/MS data were processed using FragPipe v24.0 with MSFragger v4.4.1. Spectra were searched against a target-decoy database containing the relevant *C. histolyticum* proteome (UniProt proteome UP000308489), together with common contaminant sequences. Trypsin specificity was specified with up to two missed cleavages, and peptides between 7 and 50 amino acids were considered. Precursor and fragment mass tolerances were set to ±20 ppm and 20 ppm, respectively. Carbamidomethylation of cysteine (+57.02146 Da) was specified as a fixed modification, while methionine oxidation (+15.9949 Da) and protein N-terminal acetylation (+42.0106 Da) were included as variable modifications. Database-search results were rescored using Percolator v3.7.1,

followed by protein inference using ProteinProphet. Protein identifications were filtered using a target-decoy strategy with a 1% protein-level false discovery rate. Label-free peptide/protein quantification was performed using IonQuant v1.11.20 with match-between-runs enabled.

3.5. Cryo-electron microscopy of cardiac light-chain amyloids

UltraAuFoil 200 mesh R 2/2 grids were glow-discharged under vacuum for 30 s at 15 mA in a Pelco 311 easiGlow 91000 glow discharge cleaning system (Ted Pella). 3.5 µl sample was applied to the front of the grid and 0.5 µl to the back, incubated for 1 min, and backblotted with Whatman 1 filter paper for 4–6 s after the liquid spot on the filter paper stopped spreading. Grids were manually plunge-frozen in a 4 °C cold room with >95% humidity. AutoGrids were assembled and immediately screened.

Cryo-EM data were collected on a Talos Arctica TEM (Thermo Fisher) operating at 200 keV. Movies were recorded using a Falcon 4i direct electron detector (Thermo Fisher) at a nominal magnification of 150,000, corresponding to a pixel size of 0.94 Å. Movies were saved in the electron-event representation (EER) format and recorded with a total electron exposure of 50 e^{-} per $Å^2$. All datasets were collected automatically using EPU (v.3.9, Thermo Fisher) with a defocus range of −0.8 to −2.0 µm. EPU's fast exposure navigation was used to collect data with an 8 µm image shift. All cryo-EM datasets were processed using cryoSPARC (v.4.6).

For preprocessing, movies were dose-fractionated into 40 frames and motion-corrected using patch-based motion correction, followed by patch-based CTF estimation in cryoSPARC Live. Micrographs with CTF fits worse than 6 Å and astigmatism greater than 600 Å were discarded. The remaining accepted micrographs were used for further processing. Per-dataset micrograph statistics are compiled in **Table S1**.

3.6. Processing of tubular filaments in EMPIAR-11785

From the 2013 processed micrographs in EMPIAR-11785, approximately 30% contained non-amyloid segments and were selected for further processing. To increase the number of particles, filaments were traced using a diameter range of 40–80 Å without templates and with 0.5× Gaussian blur. Segments were extracted in a 208-pixel box, Fourier-cropped to 100 pixels, and subjected to two rounds of 2D classification with an initial class uncertainty factor of 4 and hard classification during the final iteration. A total of 30,000 selected segments were re-extracted with recentering in a 256-pixel box without binning. Helical refinement was performed without symmetry using a cylindrical model and a 60° tilt search-range. The resulting reconstruction was used as input for real-space helical indexing in HI3D (Behrmann et al., 2012) with default parameters, 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* (Liebschner et al., 2019) was used and identified the highest correlation with C6 point-group symmetry. After another round of 2D classification, 21,000 segments were used in helical refinement with C6 symmetry, a rise of 38 Å and a 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 3.5 Å reconstruction (FSC 0.143). The map was sharpened automatically within cryoSPARC using a B-factor of −56 $Å^2$. The full processing workflow and validation are shown in **Fig. S1, Fig. S2** and **Table S1**.

3.7. Construction of a candidate sequence database

Sequence candidates were assembled from the *Clostridium histolyticum* reference proteome (UniProt UP000308489) carrying phage-, tail-, capsid-, portal-, terminase-, holin- or bacteriocin-related annotations. The corresponding models were retrieved from the AlphaFold Protein Structure Database, with a list of annotated non-redundant candidates provided in supplementary **Table S2**.

3.7.1. Sequence-agnostic identification of the fibrillar species

A main-chain model was built de novo into the sharpened map using CryoAtom v.2.1.0 (Su et al., 2026) in no-sequence mode, and the resulting per-residue amino-acid probability profiles were used for hidden Markov model searches (hmm_search) against three databases in turn: (I) the 22-protein candidate set; (II) the three available Clostridium histolyticum proteomes (UP000308489, UP001732247 and UP001732502); and the human reference proteome (UP000005640). In parallel, AlphaFold Protein Structure Database models for all 22 candidates were rigid-body docked into a single asymmetric unit of the map in ChimeraX v.1.10.1 and ranked by the map-model fit Q-score plugin. Structures shown in Fig. S3.

3.8. Model building and refinement

Map resolvability was improved through postprocessing in EMready2 (Cao et al., 2026). The AlphaFold model of A0A4U9RSA4 was manually built into the density in Coot v.0.9.8.95 (Emsley and Cowtan, 2004) and ISOLDE (Croll, 2018). The helical assembly was generated by applying the refined helical parameters with the *sym* command in ChimeraX, and the resulting dodecamer model was refined against the sharpened map in *phenix.real_space_refine* (Liebschner et al., 2019) with secondary-structure, Ramachandran and non-crystallographic symmetry restraints. Further validation was performed using the Q-score (Pintilie et al., 2020) for residue-wise quality estimation. Results were visualized using ChimeraX (Pettersen et al., 2021).

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors used Claude to assist in writing code for data extraction and visualization. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

CRediT authorship contribution statement

Jan-Hannes Schaefer: Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Robert T. O'Neill:** Writing – review & editing, Investigation. **Zi Gao:** Writing – review & editing, Methodology, Formal analysis, Data curation. **Danielle A. Grotjahn:** Writing – review & editing, Supervision. **Evan T. Powers:** Writing – review & editing, Supervision, Funding acquisition. **Gabriel C. Lander:** Writing – review & editing, Supervision, Funding acquisition. **Jeffery W. Kelly:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Funding for this research was provided by a NINDS grant NS095892 to GCL, NIDDK grant DK046335 and NHLBI grant HL157566 to JWK, and the German Research Council project number 556478029 to JHS. RTO is grateful for the financial support from the George E. Hewitt Foundation for Medical Research. The authors thank Jean-Christophe Ducom and Charles Bowman at Scripps for computational support. We thank the two anonymous reviewers, who substantially improved the

manuscript by prompting us to investigate the collagenase as a source of the observed tailocin contamination.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.yjsbx.2026.100169.

Data availability

Data will be made available on request.

References

Al Hamed, R., Bazarbachi, A.H., Bazarbachi, A., Malard, F., Harousseau, J.-L., Mohty, M., 2021. Comprehensive review of AL amyloidosis: some practical recommendations. Blood Cancer J. 11, 97. https://doi.org/10.1038/s41408-021-00486-4.

Annamalai, K., Gührs, K., Koehler, R., Schmidt, M., Michel, H., Loos, C., Gaffney, P.M., Sigurdson, C.J., Hegenbart, U., Schönland, S., Fändrich, M., 2016. Polymorphism of amyloid fibrils in vivo. Angew. Chem. Int. Ed. 55, 4822–4825. https://doi.org/10.1002/anie.201511524.

Backman, T., Burbano, H.A., Karasov, T.L., 2024. Tradeoffs and constraints on the evolution of tailocins. Trends Microbiol. 32, 1084–1095. https://doi.org/10.1016/j.tim.2024.04.001.

Behrmann, E., Tao, G., Stokes, D.L., Egelman, E.H., Raunser, S., Penczek, P.A., 2012. Real-space processing of helical filaments in SPARX. J. Struct. Biol. 177, 302–313. https://doi.org/10.1016/j.jsb.2011.12.020.

Belcher, E.R., Hardwick, S.W., Maia De Oliveira, T., Hyvönen, M., 2026. Affinity purification contaminants identified by cryo-EM and mass spectrometry. Biosci. Rep. 46, BSR20260232. https://doi.org/10.1042/BSR20260232.

Biasini, E., Tapella, L., Mantovani, S., Stravalaci, M., Gobbi, M., Harris, D.A., Chiesa, R., 2009. Immunopurification of pathological prion protein aggregates. PloS One 4, e7816. https://doi.org/10.1371/journal.pone.0007816.

Burt, A., Toader, B., Warshamanage, R., Von Kügelgen, A., Pyle, E., Zivanov, J., Kimanius, D., Bharat, T.A.M., Scheres, S.H.W., 2024. An image processing pipeline for electron cryo-tomography in RELION -5. FEBS Open Bio 14, 1788–1804. https://doi.org/10.1002/2211-5463.13873.

Caliseki, M., Borucu, U., Yadav, S.K.N., Schaffitzel, C., Kabasakal, B.V., 2025. Off-target structural insights: ArnA and AcrB in bacterial membrane-protein cryo-EM analysis. Acta Crystallogr. D Struct. Biol. 81, 545–557. https://doi.org/10.1107/S2059798325007089.

Cao, H., Zhu, Y., Li, T., et al., 2026. EMReady2: improvement of cryo-EM and cryo-ET maps by local quality-aware deep learning with Mamba. Nat Commun 17, 6493. https://doi.org/10.1038/s41467-026-71794-1.

Chowdhury, S., Ketcham, S.A., Schroer, T.A., Lander, G.C., 2015. Structural organization of the dynein–dynactin complex bound to microtubules. Nat. Struct. Mol. Biol. 22, 345–347. https://doi.org/10.1038/nsmb.2996.

Croll, T.I., 2018. *ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps*. Acta Crystallogr. D Struct. Biol. 74, 519–530. https://doi.org/10.1107/S2059798318002425.

Duan, J., Keeler, E., McFarland, A., Scott, P., Collman, R.G., Bushman, F.D., 2024. The virome of the kitome: small circular virus-like genomes in laboratory reagents. Microbiol. Resour. Announc. 13, e01261–23. https://doi.org/10.1128/mra.01261-23.

Emsley, P., Cowtan, K., 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132. https://doi.org/10.1107/S0907444904019158.

Fändrich, M., Schmidt, M., 2021. Methods to study the structure of misfolded protein states in systemic amyloidosis. Biochem. Soc. Trans. 49, 977–985. https://doi.org/10.1042/BST20201022.

Granato, E.T., Meiller-Legrand, T.A., Foster, K.R., 2019. The evolution and ecology of bacterial warfare. Curr. Biol. 29, R521–R537. https://doi.org/10.1016/j.cub.2019.04.024.

Gu, Z., Ge, X., Wang, J., 2025. Structure of an F-type phage tail-like bacteriocin from listeria monocytogenes. Nat. Commun. 16, 1695. https://doi.org/10.1038/s41467-025-57075-3.

Ho, C.-M., Li, X., Lai, M., Terwilliger, T.C., Beck, J.R., Wohlschlegel, J., Goldberg, D.E., Fitzpatrick, A.W.P., Zhou, Z.H., 2020. Bottom-up structural proteomics: cryoEM of protein complexes enriched from the cellular milieu. Nat. Methods 17, 79–85. https://doi.org/10.1038/s41592-019-0637-y.

Hungler, A., Momin, A., Diederichs, K., Arold, S., T., 2016. *ContaMiner and ContaBase: a webserver and database for early identification of unwantedly crystallized protein contaminants*. J. Appl. Crystallogr. 49, 2252–2258.https://doi.org/10.1107/S1600576716014965.

Iakovleva, I., Hall, M., Oelker, M., Sandblad, L., Anan, I., Sauer-Eriksson, A.E., 2021. Structural basis for transthyretin amyloid formation in vitreous body of the eye. Nat. Commun. 12, 7141. https://doi.org/10.1038/s41467-021-27481-4.

Jamali, K., Käll, L., Zhang, R., Brown, A., Kimanius, D., Scheres, S.H.W., 2024. Automated model building and protein identification in cryo-EM maps. Nature 628, 450–457. https://doi.org/10.1038/s41586-024-07215-4.

Jenkins, M.C., Dutta, T., Bobe, D., Kopylov, M., 2024. Identification and cryoEM structure determination of *Escherichia* phage YDC107 tail found in a bacteria-contaminated buffer. https://doi.org/10.1101/2024.12.10.627647.

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., 2022. Amyloid fibrils in FTLD-TDP are composed of TMEM106B and not TDP-43. Nature 605, 304–309. https://doi.org/10.1038/s41586-022-04670-9.

Karimi-Farsijani, S., Sharma, K., Ugrina, M., Kuhn, L., Pfeiffer, P.B., Haupt, C., Wiese, S., Hegenbart, U., Schönland, S.O., Schwierz, N., Schmidt, M., Fändrich, M., 2024. Cryo-EM structure of a lysozyme-derived amyloid fibril from hereditary amyloidosis. Nat. Commun. 15, 9648. https://doi.org/10.1038/s41467-024-54091-7.

Kedersha, N.L., Rome, L.H., 1986. Isolation and characterization of a novel ribonucleoprotein particle: large structures contain a single species of small RNA. J. Cell Biol. 103, 699–709. https://doi.org/10.1083/jcb.103.3.699.

Kyrilis, F.L., Semchonok, D.A., Skalidis, I., Tüting, C., Hamdi, F., O'Reilly, F.J., Rappsilber, J., Kastritis, P.L., 2021. Integrative structure of a 10-megadalton eukaryotic pyruvate dehydrogenase complex from native cell extracts. Cell Rep. 34, 108727. https://doi.org/10.1016/j.celrep.2021.108727.

Lashuel, H.A., 2026. Cracking the code of native amyloid fibrils: advances and next steps to enable pathology-informed therapeutic and diagnostic. Nat. Struct. Mol. Biol. 33, 567–576. https://doi.org/10.1038/s41594-026-01792-w.

Lee, G., Chakraborty, U., Gebhart, D., Govoni, G.R., Zhou, Z.H., Scholl, D., 2016. F-type Bacteriocins of listeria monocytogenes: a new class of phage tail-like structures reveals broad parallel coevolution between tailed bacteriophages and high-molecular-weight Bacteriocins. J. Bacteriol. 198, 2784–2793. https://doi.org/10.1128/JB.00489-16.

Liebschner, D., Afonine, P.V., Baker, M.L., Bunkóczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.W., Jain, Swati, McCoy, A.J., Moriarty, N.W., Oeffner, R.D., Poon, B.K., Prisant, M.G., Read, R.J., Richardson, J.S., Richardson, D.C., Sammito, M.D., Sobolev, O.V., Stockwell, D.H., Terwilliger, T.C., Urzhumtsev, A.G., Videau, L.L., Williams, C.J., Adams, P.D., 2019. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr. D Struct. Biol. 75, 861–877. https://doi.org/10.1107/S2059798319011471.

Liu, X., Nizamudeen, Z.A., Hill, C.J., Parmenter, C., Arkill, K.P., Lambert, D.W., Hunt, S., 2025. Assessment of vault particles in cancer cell line-derived extracellular vesicle preparations. J. Extracell. Vesicles 14, e70142. https://doi.org/10.1002/jev2.70142.

Lövestam, S., Scheres, S.H.W., 2025. Cryo-EM structure of the vault from human brain reveals symmetry mismatch at its caps. Structure 33, 1643–1648.e1. https://doi.org/10.1016/j.str.2025.07.014.

Naccache, S.N., Greninger, A.L., Lee, D., Coffey, L.L., Phan, T., Rein-Weston, A., Aronsohn, A., Hackett, J., Delwart, E.L., Chiu, C.Y., 2013. The perils of pathogen discovery: origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction spin columns. J. Virol. 87, 11966–11977. https://doi.org/10.1128/JVI.02323-13.

Nguyen, B.A., Fernandez-Ramirez, M.D.C., Bassett, P., Singh, V., Singh, P., Pękała, M., Villalon, L., Ahmed, Y., Lemoff, A., Evers, B., Lopez, C., Kluve-Beckerman, B., Saelices, L., 2026. Cryo-EM reveals structural variability of apolipoprotein A-I amyloid fibrils across organs, mutations, and clinical presentations. Nat. Commun. 17, 5324. https://doi.org/10.1038/s41467-026-72150-z.

Nguyen, B.A., Singh, V., Afrin, S., Yakubovska, A., Wang, L., Ahmed, Y., Pedretti, R., Fernandez-Ramirez, M.D.C., Singh, P., Pękała, M., Cabrera Hernandez, L.O., Kumar, S., Lemoff, A., Gonzalez-Prieto, R., Sawaya, M.R., Eisenberg, D.S., Benson, M.D., Saelices, L., 2024. Structural polymorphism of amyloid fibrils in ATTR amyloidosis revealed by cryo-electron microscopy. Nat. Commun. 15, 581. https://doi.org/10.1038/s41467-024-44820-3.

Patz, S., Becker, Y., Richert-Pöggeler, K.R., Berger, B., Ruppel, S., Huson, D.H., Becker, M., 2019. Phage tail-like particles are versatile bacterial nanomachines – a mini-review. J. Adv. Res. 19, 75–84. https://doi.org/10.1016/j.jare.2019.04.003.

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

Pintilie, G., Zhang, K., Su, Z., Li, S., Schmid, M.F., Chiu, W., 2020. Measurement of atom resolvability in cryo-EM maps with Q-scores. Nat. Methods 17, 328–334. https://doi.org/10.1038/s41592-020-0731-1.

Salter, S.J., Cox, M.J., Turek, E.M., Calus, S.T., Cookson, W.O., Moffatt, M.F., Turner, P., Parkhill, J., Loman, N.J., Walker, A.W., 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol. 12, 87. https://doi.org/10.1186/s12915-014-0087-z.

Sanchez, J., Pierson, J., Borcik, C., Rienstra, C., Wright, E., 2025. High-resolution Cryo-EM structure determination of a-Synuclein—A prototypical amyloid fibril. BIO-Protoc. 15. https://doi.org/10.21769/BioProtoc.5171.

Scheres, S.H.W., Ryskeldi-Falcon, B., Goedert, M., 2023. Molecular pathology of neurodegenerative diseases by cryo-EM of amyloids. Nature 621, 701–710. https://doi.org/10.1038/s41586-023-06437-2.

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

Scholl, D., 2017. Phage tail–like bacteriocins. Annu. Rev. Virol. 4, 453–467. https://doi.org/10.1146/annurev-virology-101416-041632.

Schubert, M.P.M., Zucker-Franklin, D., Rimon, A., Franklin, E.C., 1968. The characterization of soluble amyloid prepared in water. J. Clin. Invest. 47, 924–933. https://doi.org/10.1172/JCI105784.

Sigal, N., Lichtenstein-Wolfheim, R., Schlussel, S., Azulay, G., Borovok, I., Holdengraber, V., Elad, N., Wolf, S.G., Zalk, R., Zarivach, R., Frank, G.A.,

Herskovits, A.A., 2024. Specialized listeria monocytogenes produce tailocins to provide a population-level competitive growth advantage. Nat. Microbiol. 9, 2727–2737. https://doi.org/10.1038/s41564-024-01793-9.

Skalidis, I., Kyrilis, F.L., Tüting, C., Hamdi, F., Chojnowski, G., Kastritis, P.L., 2022. Cryo-EM and artificial intelligence visualize endogenous protein community members. Structure 30, 575–589.e6. https://doi.org/10.1016/j.str.2022.01.001.

Steinebrei, M., Baur, J., Pradhan, A., Kupfer, N., Wiese, S., Hegenbart, U., Schönland, S. O., Schmidt, M., Fändrich, M., 2023. Common transthyretin-derived amyloid fibril structures in patients with hereditary ATTR amyloidosis. Nat. Commun. 14, 7623. https://doi.org/10.1038/s41467-023-43301-3.

Steinebrei, M., Gottwald, J., Baur, J., Röcken, C., Hegenbart, U., Schönland, S., Schmidt, M., 2022. Cryo-EM structure of an ATTRwt amyloid fibril from systemic non-hereditary transthyretin amyloidosis. Nat. Commun. 13, 6398. https://doi.org/10.1038/s41467-022-33591-4.

Su, B., Huang, K., Peng, Z., Amunts, A., Yang, J., 2026. CryoAtom improves model building for cryo-EM. Nat. Struct. Mol. Biol. 33, 351–361. https://doi.org/10.1038/s41594-025-01713-3.

Suloway, C., Pulokas, J., Fellmann, D., Cheng, A., Guerra, F., Quispe, J., Stagg, S., Potter, C.S., Carragher, B., 2005. Automated molecular microscopy: the new Leginon system. J. Struct. Biol. 151, 41–60. https://doi.org/10.1016/j.jsb.2005.03.010.

Van Kempen, M., Kim, S.S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C.L.M., Söding, J., Steinegger, M., 2024. Fast and accurate protein structure search with Foldseek. Nat. Biotechnol. 42, 243–246. https://doi.org/10.1038/s41587-023-01773-0.

Yao, Y., Zhao, Q., Yao, S., Xu, Y., Liu, K., Cao, T., Sun, B., Zhou, J., Liu, C., Li, D., 2026. Biopsy-resolved cryo-EM structures of amyloid fibrils provide molecular insights into AL amyloidosis. Proc. Natl. Acad. Sci. U. S. A. 123, e2515454123. https://doi.org/10.1073/pnas.2515454123.

Zinke, M., Schröder, G.F., Lange, A., 2022. Major tail proteins of bacteriophages of the order Caudovirales. J. Biol. Chem. 298, 101472. https://doi.org/10.1016/j.jbc.2021.101472.