Cryo-EM reveals that cardiac IGLV6-derived AL fibrils can be polymorphic

Parker Bassett*[1], Binh A. Nguyen*[1], Virender Singh*[1], Patrick T. Garrett[2], James J. Moresco[3], Shumaila Afrin[1], Maja Pękała[1], and Lorena Saelices[1].

[1]Center for Alzheimer's and Neurodegenerative Diseases, Department of Biophysics, Peter O'Donnell Jr. Brain Institute, University of Texas Southwestern Medical Center (UTSW), Dallas, TX, USA.

[2]Department of Molecular Medicine, The Scripps Research Institute (TSRI), La Jolla, California, USA.

[3]Center for the Genetics of Host Defense, University of Texas Southwestern Medical Center (UTSW), Dallas, Texas, USA.

*These authors contributed equally.

Correspondence to: Lorena Saelices Gómez, lorena.saelicesgomez@utsouthwestern.edu

ABSTRACT

AL amyloidosis is a systemic disease caused by the aggregation of free antibody light chains (LC) secreted by aberrant plasma cell clones into the bloodstream. These LC aggregates form amyloid fibrils that deposit in multiple organs, leading to organ failure and, ultimately, death. Investigating the structural basis of LC aggregation is a critical avenue for understanding the biopathology of AL amyloidosis. Structural insights into AL fibril formation may reveal mechanisms driving amyloid deposition and inspire novel therapeutic strategies. Previous studies using cryo-electron microscopy have uncovered diverse AL fibril structures extracted from multiple patients, highlighting variability in fibril architecture. Here, we present a novel cryo-EM structure of AL cardiac fibrils from a patient with AL amyloidosis. This structure reveals a unique fibril fold and the coexistence of multiple morphologies, including single- and double-protofilament forms. Notably, some of these double-protofilament fibrils exhibit an uncommon rotational symmetry, raising intriguing questions about the mechanisms governing fibril formation and evolution over time.

INTRODUCTION

Light chain amyloidosis (or AL amyloidosis) is a systemic amyloid disease that accounts for over 75% of systemic amyloid patients and is predicted to affect an increasing amount of people in the US.[1,2] The disease is marked by the generation of an aberrant plasma clone, which freely secretes light chain (LC) protein into the bloodstream, removing the native protein from the protective folding environment of the cell. Outside of the cell, these free LC proteins are free to misfold and aggregate into amyloid fibrils, eventually depositing in various organs and initiating the disease.[3,4] The disease most often affects the heart and kidneys, and can be particularly aggressive, with some patients with advanced cardiac involvement only surviving for a few weeks after onset.[2] In conjunction, it is widely believed that AL is underdiagnosed due to a similar phenotype to other diseases, which underscores the importance of studying the pathogenesis of the disease for improving the rates of diagnosis and aiding prognosis.[4]

LC protein is a roughly 25 kDa protein secreted from plasma cells that most often exists coupled with a heavy chain protein within the larger antibody protein structure, ultimately consisting of two light chain and two heavy chain subunits per mature immunoglobulin G protein.[5] Each LC subunit contains a variable light chain domain genetically fused to a constant light chain domain during a process known as V(D)J recombination. V(D)J recombination shuffles together the different combinations of variable domain and constant domain genes from the diverse set of germline-encoded genes.[3,6] These germline genes are organized into many different families based on similar sequence identities, and typically named as IGLV and a number (ex. IGLV6).[6,7] One gene from these families is selected randomly for genetic recombination, which when coupled with other downstream mutational processes, results in a vastly diverse repertoire of antibody structures and sequences.

Although the amyloidogenic LC sequence varies from patient to patient, various germline sequences have been shown to correlate with organ involvement and clinical phenotype. For example, within the lambda family, the IGLV1 gene family is predominantly associated with renal involvement, and IGLV3 is predominantly associated with cardiac involvement.[8] The most common lambda variable region implicated in AL however is IGLV6, which is predominantly associated with both cardiac and renal involvement.[3,8–10] These phenotypic trends are likely driven

by the unique features within a germline family, since each individual gene is more similar to those within the same family than those classified into others. Recent cryo-EM studies show that AL fibrils from IGLV3, although containing different LC sequences, are structurally related.[11,12] This observation suggests that LC sequences derived from a similar germline are structurally related despite somatic variability. Thus, we hypothesize that the germline sequence drives the formation of fibrils with a related structural conformation, which may in turn target a specific set of organs and contribute to the phenotypic variability.

Here we report new cryo-EM structures of AL fibrils, extracted from the explanted heart of a patient with AL amyloidosis, with germline IGLV6. These fibrils display multiple morphologies with single or double protofilaments, one of which present with an uncommon rotational symmetry, raising new questions about the fibril formation process in AL amyloidosis and its effects on clinical phenotype.

RESULTS

Characterization of cardiac tissue and fibrils extracted from a patient with AL

We sourced cardiac tissue from the explanted heart of a patient at UT Southwestern Medical Center diagnosed with AL amyloidosis and cardiac involvement. We then treated the tissue to a water-based extraction previously used in our lab.[13,14] Transmission electron microscopy (TEM) showed fibrillar deposits after extraction, confirming the presence of fibrils after extraction from tissue (Fig. 1a and Suppl. Fig. 1). As further validation of the amyloid identity, we used western blot, where we observed the presence of a λ LC species between 10 and 15 kDa, indicating the presence a truncated light chain monomer (Fig. 1b). Additionally, intact mass analysis indicated the predominant protein species in our purified fibrils had a mass of 12,842.74 Da, roughly matching the size of the λ-light chain band observed in the western blot data (Fig. 1c).

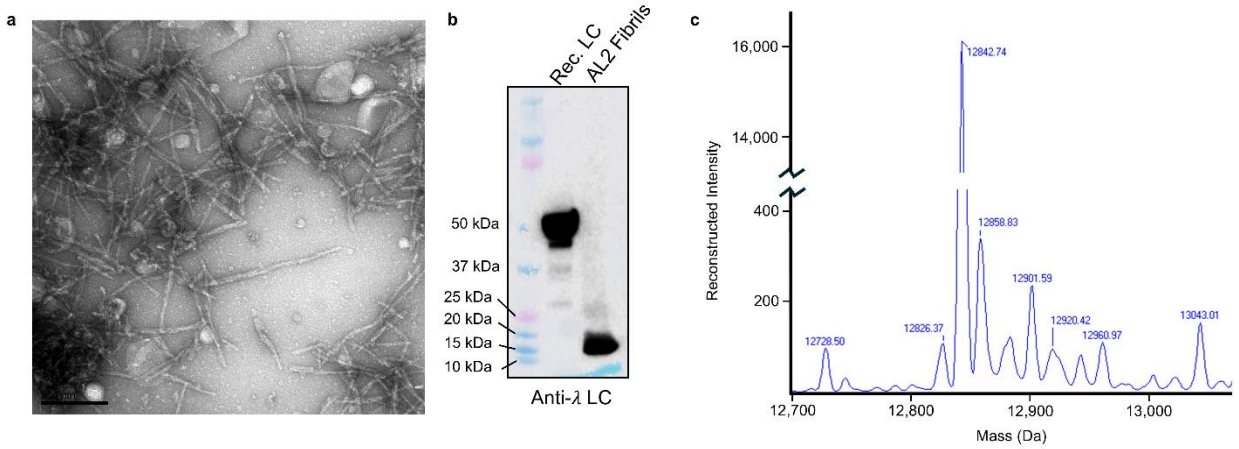

Figure 1. Fibrils extracted from cardiac tissue of a patient with AL consist of λ LC fragments. a) TEM micrograph of fibrils extracted from cardiac tissue of a patient with AL amyloidosis. 20,000x magnification, scale bar 200 nm. **b)** Western blot with anti-λ LC antibody targeting a LC protein control and fibrils extracted from the patient sample AL2. **c)** Intact mass analysis shows the dominant protein species in the extracted fibril solution to have a mass of 12,842 Da, closely matching the band in panel 1b.

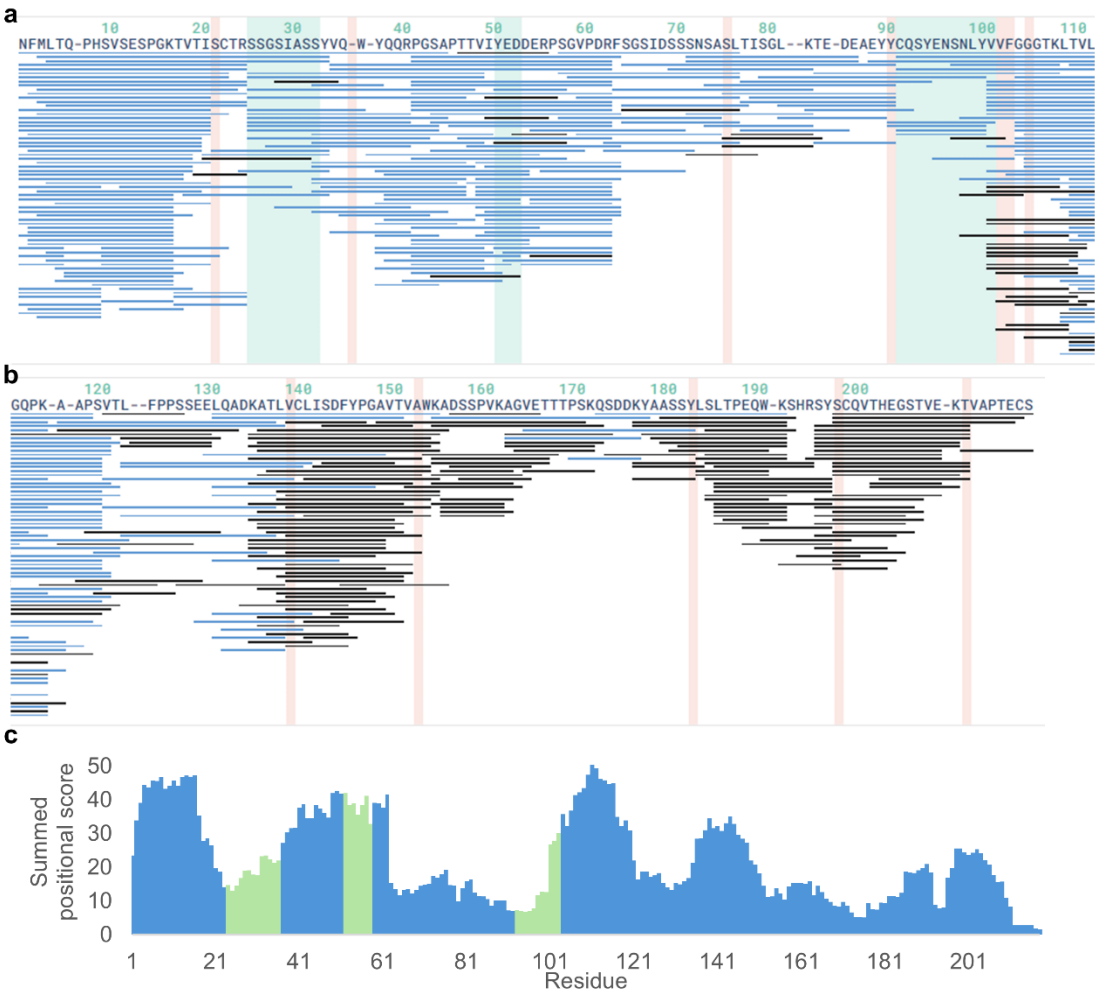

Figure 2. Sequence prediction using *de novo* mass spectrometry. Short peptide reads from enzymatic digestion are mapped to the predicted sequence of the AL fibrils. The sequence is shown at the top for the variable and joining regions **(a)** and the constant region **(b)**. Blue bars indicate peptides that uniquely match the predicted sequence, while black bars indicate peptides that match more than one immunoglobulin sequence. **c)** Depth of coverage for the sequence reads. For all panels complementarity-determining regions are shaded in green.

Sequence determination of extracted amyloid fibrils

Bone marrow aspirate was unavailable for this patient, so neither the clonal protein sequence involved in the amyloid, nor the germline genetic identity was known prior to structural determination. In order to determine the amino acid sequence of the fibrils, we used a mass spectrometry (MS)-based approach with the software Casanovo and Stitch.[15–18] Digestion of the purified ex-vivo fibrils with trypsin, chymotrypsin, and elastase gave a variety of MS spectra that were used in a *de novo* sequence determination screening for the sequence comprising the entire length of the LC protein (Fig. 2). As a result, we were able to determine the most likely germline involved in the *ex vivo* fibrils, as well as the individual mutations likely present in the final aggregating protein. Based on our results, we determined the fibril sequence was derived from the IGLV6-57 germline sequence, which is supported by the cryo-EM data as will be discussed later. The constant region is predicted to be IGLC2. The joining region was determined to be IGLJ2, however this is an ambiguous assignment since the sequences of IGLJ2 and IGLJ3 are extremely similar in identity. Database searches with our MS data allowed us to identify additional

possible proteins in our extracted fibril solution, including collagen VI (COLVI), consistent with previous observations of interactions between AL fibrils and COLVI (Suppl. Table 1).[12] This was further supported by our Stitch results, which included a high decoy score, and a significant portion of the peptide reads that could be attributed to COLVI.

Cryo-EM data processing and modeling of the single protofilament structure

Early Cryo-EM screening confirmed the presence of diverse configurations of fibrils, suggesting multiple morphologies (Fig. 3a and Suppl. Fig. 2). Particle picking using Topaz resulted in over 1.4 million particles. 2D classification using RELION revealed both single and double protofilament structures (Suppl. Fig. 3a). We selected 2D classes and stitched them together manually to determine the approximate fibril crossover length, twist, and rise for single and double morphologies (Fig. 3b and Suppl. Table 2). We were able to generate an initial model in RELION for a double protofilament structure for use in further steps, but the single protofilament structure was too curved for RELION to generate a suitable initial model (Suppl. Fig. 3b). We then performed 3D classification, which revealed three unique structural classes, two of which comprised a double protofilament structure (Fig. 3c). A scheme for the overall helical reconstruction process is seen in Suppl. Fig. 4.

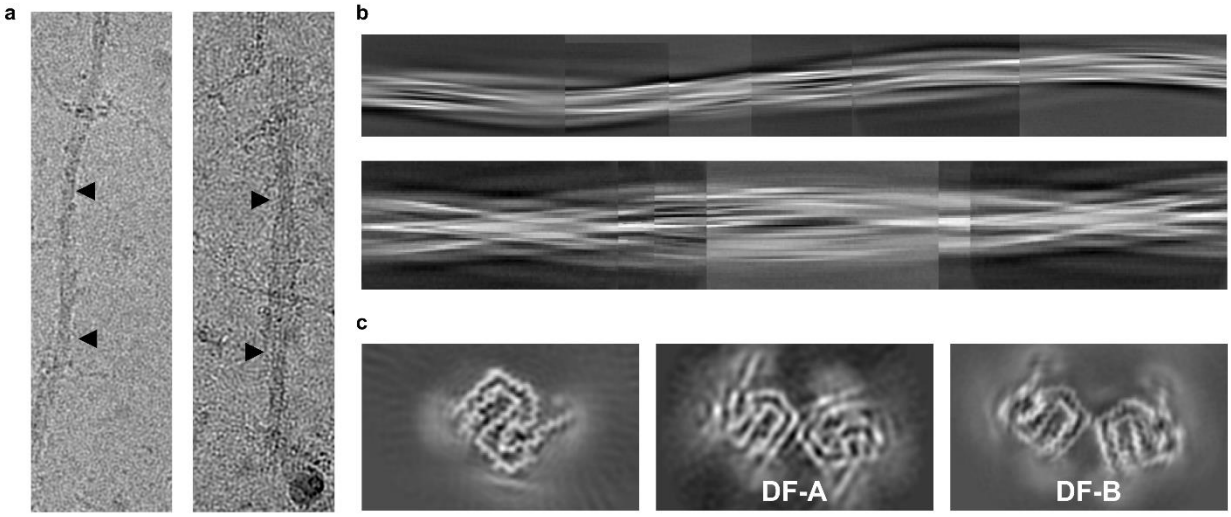

Figure 3. Cryo-EM of cardiac AL fibrils reveals multiple morphologies. a) Cryo-EM micrographs show both single (left) and double (right) protofilament structures. Crossover lengths are marked by triangles. **b)** Full-length crossovers were stitched from representative 2D classes. 2D classes showed both single (top) and double (bottom) protofilament structures. **c)** 3D classification revealed three dominant morphologies: a single protofilament structure (left), and two double protofilament structures related by differing symmetries (middle, right).

The single protofilament structure was the most abundant and we solved the structure at a resolution of 3.0Å based on the gold-standard FSC curve (Suppl. Fig. 5a). At this resolution, the cross-beta fibril structure was clearly resolved, and the overall conformation of the fibril core was determined (Fig. 4a). Local resolution of the map is best at the center of the core at 2.8Å, and some areas of the core were less resolved and more likely to be disordered (Suppl. Fig. 5b). The fibril topology was observed with a twist angle of -1.03˚ and a helical rise of 5.03Å (Fig. 4a). Additionally, we observed three disconnected densities unlikely to result from sidechain orientation (Fig. 4b, asterisks). These densities are regularly spaced alongside the layers of the fibril and in close proximity to polar residues (Suppl. Fig. 6).

We then used the MS-predicted sequence to model the atomic structure (Fig. 4b). Initial modeling using the IGLV6-57 sequence further supported the sequence prediction by matching side chain densities in the structural data with the predicted side chain positions. In total, the fibril core consists of two regions of the LC peptide chain spanning residues Asn 1 to Arg 40 and Asp 61 to approximately Leu 112. Structural data at the C-terminal region of the fibril core is less resolved and so residues 108-112 are modelled with sidechains but the exact structural orientation of the sidechains is unclear.

Analysis of the fibril core of the single protofilament structure

The structure of the fibril core in the single protofilament structure is centered around the intact disulfide bridge, spanning Cys 22 to Cys 91 (or Cys 23 to Cys 88 according to Kabat numbering). The single protofilament structure is stabilized by multiple interactions. Each monomer of the fibril adopts a structure that is not planar, instead interlocking with the monomers above and below along the fibril axis. While the N-terminal 1-40 fragment is placed in a middle position (layer n, orange), the C-terminal 61-112 fragment is located at the lower and upper layers (layers n-1, blue; and n+1, green) (Fig. 4b,c). This creates opportunities for side chains to form interactions with immediately adjacent monomers. One example of this is shown in Figure 4d, where Arg 24 of one monomer interacts with Asp 85 and Tyr 90 in a separate monomer. Additionally, the fibril structure displays further electrostatic, hydrophobic, and polar interactions. Suppl. Fig. 7 summarizes some examples. We also estimated the fibril solvation energies and found that they are the weakest of AL fibrils available in the Amyloid Atlas to date (Suppl. Fig. 8a, Suppl. Table 3).

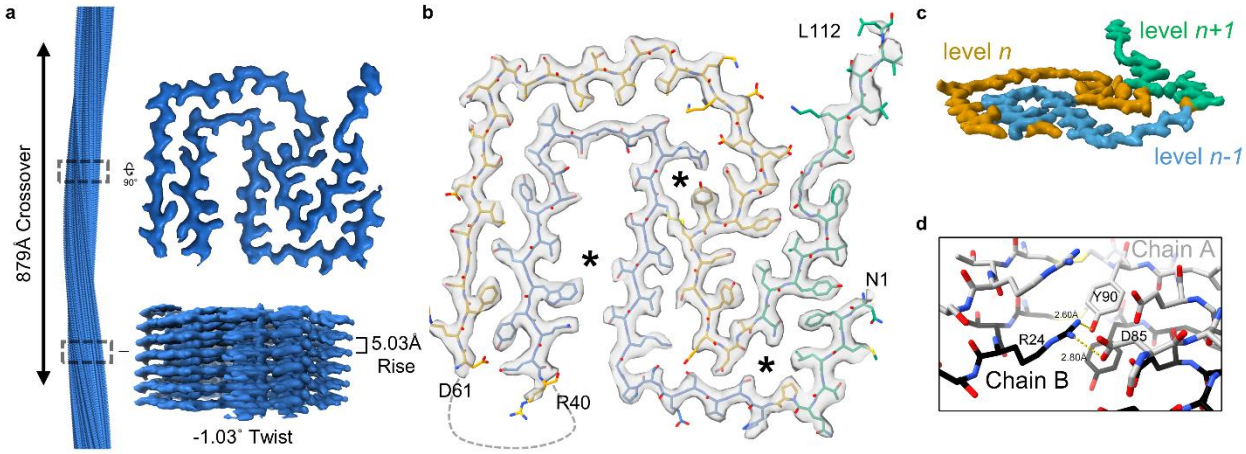

Figure 4. Modelling the dominant morphology in extracted AL fibrils. a) Fibril topology, including helical rise, twist, and crossover. The density map was solved to a resolution of 3.0A. **b)** The fibril model consists of a double layered protein spanning residues 1-40 and 61-112. Asterisks mark unmodelled densities. **c)** A single layer of the fibril is spatially distributed along the fibril axis. Colors match the backbone colors in 4b. **d)** Interactions between monomers occur due to the spatial distribution and stabilize the fibril.

Sequence roles in fibril formation

As part of our analysis, we computationally predicted aggregation-prone regions (APRs) in our LC sequence. We used four APR prediction models (WALTZ, TANGO, Aggrescan, and AmylPred), and calculated an overall APR prediction for each residue with a score from zero to four, reflective of how many models predicted each residue to be part of an APR.[19–25] Our analysis showed eleven regions in the full-length LC protein predicted to be an APR by at least one model (Suppl. Table 4). Of these eleven APRs, seven are located in the VL domain and five reside in the structured core of the fibril (Fig. 5). These APRs encompassed most of the core structure, and the

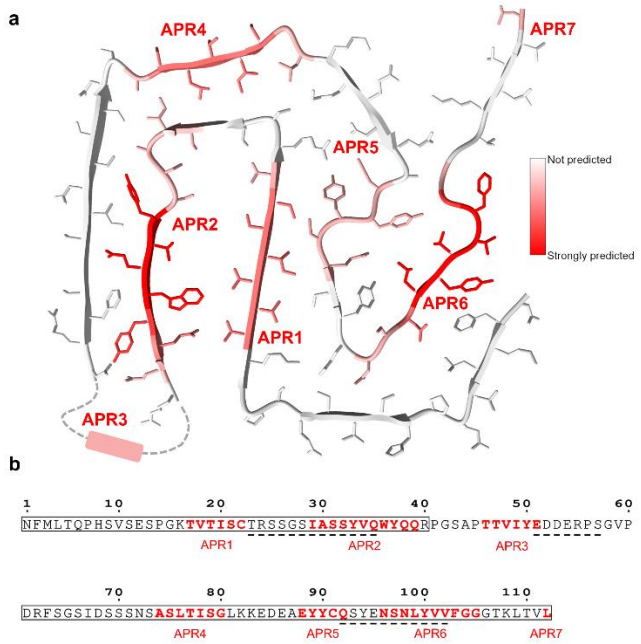

Figure 5. Predicted APRs reside in the structured core of the fibril. a) A heat map showing the predicted APRs in the AL fibril structure. Darker red indicates higher likelihood for aggregation propensity. **b)** Sequence identity of the predicted APRs. Residues in the core of the fibril are boxed, and complementarity-determining regions are underlined.

two APRs most often predicted were from residues Ile 29 to Gln 39 and from residues Asn 96 to Gly 105. Additionally, both Cys 22 and Cys 91 residues were also involved in APRs.

Polymorphism and a unique rotational symmetry

We observed two distinct double protofilament structures during 3D classification using RELION (Fig. 3c). The structures each had distinct rotational symmetries, with protofilaments in each class related by rotations around axes either parallel or perpendicular to the fibril axis (Fig. 6a). The less resolved structure, termed DF-A, has a rotational 2-fold symmetry with two identical protofilaments related by a 180° rotation around the axis of the fibril. The more resolved structure, termed DF-B, has a rotational 2-fold symmetry with two identical protofilaments related by a 180° rotation around an axis that is perpendicular to the fibril axis. When compared to the single protofilament structure, DF-A and DF-B had diverse helical rise and twist values (Suppl. Table 2). This resulted in different helical crossover lengths compared to the single protofilament structure (Fig. 6b,c). The DF-A density map was extremely limited and was estimated by RELION to reach 7.29Å. Based on FSC curves, the DF-B density map had a resolution of 3.78Å, (Suppl. Fig. 6c). The limited resolution of the density maps hindered the modelling of a structure for these two double protofilament morphologies. However, since the 3D classes share a structural resemblance with the single protofilament morphology, we used molecular docking to fit the single protofilament model into the double protofilament maps (Fig. 6b,c). The docking resulted in reasonable fits, allowing for preliminary conjectures about the residues involved in connecting the two protofilaments.

DISCUSSION

In this study, we show three unique fibril morphologies for amyloid extracted from the explanted heart of a patient with AL amyloidosis and cardiac involvement (Fig. 3). We determined the structure of the dominant morphology at 3.0Å resolution and observed two double protofilament structures, all of which seem to share the same protofilament fold with differences in the number of protofilaments found in the fibril. *De novo* sequencing by MS enabled us to generate a structure model for the dominant morphology (Fig. 2 and Fig. 4). The depth of coverage varies across the

protein, with some areas better fit to the data than others. As a result, a majority of peptide reads in the constant domain region are nonunique and can be matched to more than one constant domain template sequence.. Nevertheless, the prediction is strongest in the variable domain region where our structural data is focused.

Analysis of the intact mass and WB data reveals that the dominant LC protein in the extraction is ~12.8 KDa. This reflects possible proteolytic activity that is truncating the LC protein, which is consistent with previous observations that AL fibrils predominantly consist of the complete VL domain and varying amounts of the CL domain.[26] This truncation and the absence of any full-length LC bands in the WB data suggests an absence of the LC constant domain, which may be the cause for a large amount of peptide reads to be nonunique and ambiguously matched to more than one constant domain template sequence. Additionally, the high decoy score attributed to COLVI suggests its presence in our extracted fibrils, an observation consistent with previous studies on AL fibrils.[12]

Our present structure displays a few similarities to past AL fibril structures.[11,12,27–31] Consistent with other structures, the AL fibrils have an intact disulfide bond around which the beta strands in the fibril adopt an antiparallel configuration (Suppl. Fig. 9). This contrasts with the native structure, where the beta strands involved in the disulfide bond adopt a parallel orientation. Additionally, the structured core of the fibril includes only the variable and joining region, which is consistent with most of the published AL fibril structures to date.[11,27–31]

Our data showed that AL fibrils extracted from patients with AL amyloidosis can include multiple protofilaments, a novel observation in cryo-EM studies of AL amyloidosis to date. Indeed, multiple morphologies have been observed in cardiac amyloid fibrils from other amyloidoses, but none

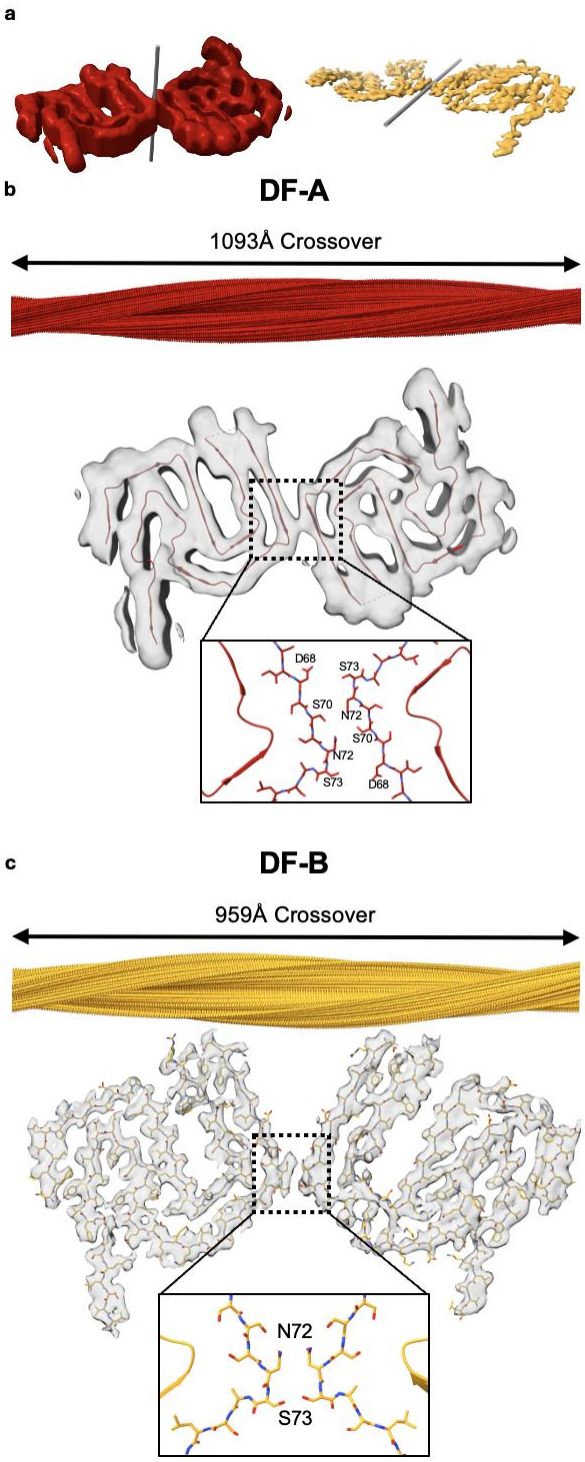

Figure 6. AL fibrils can adopt structures with multiple protofilaments. a) The double protofilament structures can adopt orthogonal modes of symmetry, where the grey rods are the rotational axes. **b,c)** The two double protofilament structures are presented with a rigid-body fit of the single protofilament model. Insets show the residues nearest the possible interaction site.

have been observed with 2-fold rotational symmetry perpendicular to the fibril axis, as described here.[13,32] In fact, across the 506 publicly available amyloid fibril structures in the Amyloid Atlas (https://people.mbi.ucla.edu/sawaya/amyloidatlas/), we only found 9 with a similar rotational symmetry perpendicular to the fibril axis. In this orientation, the two protofilaments are oriented in an antiparallel fashion. The individual protofilaments expose a unique head and tail on each side, much like the plus and minus ends of a microtubule. Then, the association of the protofilaments to form the double protofilament structure creates ends of the fibril with both a "plus" and a "minus" end (Suppl. Fig. 10) Molecular docking allowed us to visualize the nature of these orientations (Fig. 6b,c). This unique arrangement raises questions about how these protofilaments assemble and elongate. A possible mechanism for the assembly of these fibrils could be that single protofilaments form and elongate individually, then come together and interact favorably to form the double protofilament structure. Our current study is limited by the resolution of the double protofilament data, though these observations encourage further study into polymorphism in AL and the process by which these antiparallel fibrils elongate.

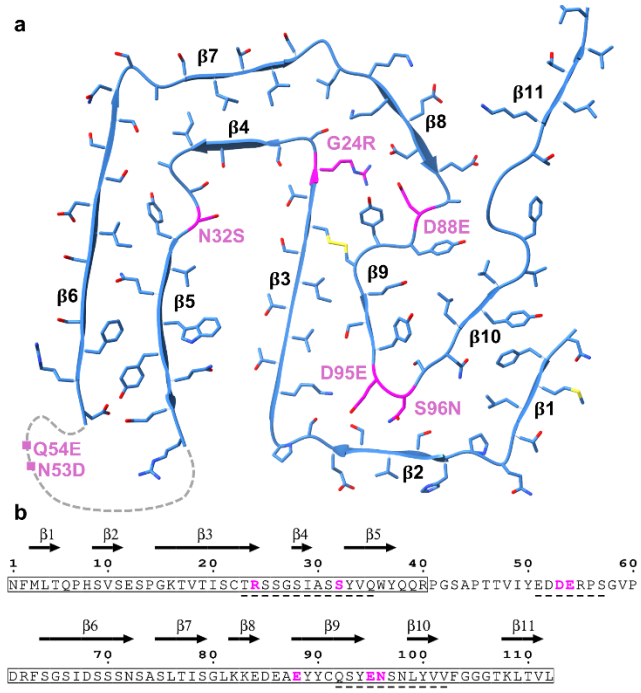

Figure 7. Most of the mutations in the AL fibril reside in the structured core. Mutation sites are shown on the model **(a)** and in the sequence **(b)**. The secondary structure is also displayed. Residues in the core of the fibril are boxed, and complementarity-determining regions are underlined.

Our present structure is the second published AL fibril structure derived from the IGLV6-57 germline sequence.[30] In the native form of IGLV6 LC, it has been previously reported that mutations drive amyloid aggregation by destabilizing the native fold.[9] In our study, the structure of a IGLV6 light chain contains five mutations from the germline sequence that reside in the core of the fibril (Fig. 7). Analysis of the structure suggests that none of these mutations contribute to the fibril solvation energy (Suppl. Fig. 8a). However, they may contribute to the formation of stabilizing interactions with neighboring residues. For instance, the germline Gly25Arg mutation creates the formation of a salt bridge within the structure, which may play a role in fibril stability despite no predicted effect on solvation energy. The Asp95Glu mutation results in a longer and more flexible sidechain that may create a more favorable interaction with Lys 16. The Asp88Glu mutation enables a possible interaction with Arg 25 by relaxing the electrostatic repulsions between the Asp 88 residues above and below. The Ser96Asn mutation forms polar interactions along the fibril axis with Asn 96 residues above and below, which would not be as favorable with a serine sidechain. Finally, the Asn32Ser likely has the least stabilizing effect on the fibril, where the sidechain is oriented away from hydrophobic regions. In sum, the mutations in our structure may stabilize the structure of the fibril overall, indicating the importance of somatic mutations in amyloidogenic LCs.

In conclusion, we have determined the structure of AL fibrils extracted from the explanted heart of a patient with AL amyloidosis. Additionally, we observed the presence of multiple

morphologies with unique and rare rotational symmetries, which raises new questions about the fibril formation process in AL amyloidosis.

METHODS

Extraction of AL fibrils. Amyloid fibrils were extracted from flash-frozen tissue as previously described[13]. Briefly, ~200 mg of frozen tissue was thawed at room temperature and cut into small piece with a scalpel. The minced tissue was resuspended into 0.4 mL Tris-calcium buffer (20 mM Tris, 138 mM NaCl, 2 mM $CaCl_2$, 0.1% NaN_3, pH 8.0) and centrifuged for 5 min at 5000 × g and 4 °C. The pellet was washed in Tris-calcium buffer three additional times. After the washing, the pellet was resuspended in 0.5 mL of 5 mg mL^{-1} collagenase solution (collagenase stock prepared in Tris-calcium buffer) and incubated overnight at 37 °C, shaking at 700 rpm. The resuspension was centrifuged for 30 min at 5000 × g and 4 °C and the pellet was resuspended in 0.4 mL Tris–ethylenediaminetetraacetic acid (EDTA) buffer (20 mM Tris, 140 mM NaCl, 10 mM EDTA, 0.1% NaN_3, pH 8.0). The suspension was centrifuged for 5 min at 5000 × g and 4 °C, and the washing step with Tris–EDTA was repeated nine additional times. All the supernatants were collected for further analysis. After the washing, the pellet was resuspended in 150 µL of ice-cold water supplemented with 5 mM EDTA and centrifuged for 5 min at 5000 × g and 4 °C. This elution step was repeated three additional times.

Negative-stained transmission electron microscopy. Amyloid fibril extraction was confirmed by transmission electron microscopy as described.[13] Briefly, a 2 µL sample was spotted onto a freshly glow-discharged carbon film 300 mesh copper grid (Electron Microscopy Sciences), incubated for 1 min, and gently blotted onto a filter paper to remove the solution. The grid was negatively stained with 2 µL of 2% uranyl acetate for 1 min and gently blotted to remove the solution. Another 5 µL uranyl acetate was applied onto the grid and immediately removed. An FEI Tecnai 12 electron microscope at an accelerating voltage of 120 kV was used to examine the specimens.

Western Blot. 1 µg of fibrils or control was combined with loading dye and heated at 95°C for 10 minutes. The samples were then subjected to electrophoresis on a SurePAGE Bis-Tris 4-12% gel. Recombinant human lambda light chain protein (Abcam) served as the control. Following electrophoresis, the proteins were transferred onto a 0.2 µm nitrocellulose membrane. The membrane was then probed with a polyclonal rabbit anti-human antibody targeting human AL light chain protein (Abcam, 1:500). A horseradish peroxidase-conjugated goat anti-rabbit IgG (Invitrogen, 1:1000) was used as the secondary antibody. The AL protein was detected using Promega Chemiluminescent Substrate, following the protocol provided by the manufacturer.

Intact mass spectrometry analysis. 20 µL (5 µg) of extracted AL fibrils were spun at 21000 x g for 1hr at 4 °C. The 15 µL of supernatant was removed and 5 µL of 8M GuHCl was added to dissociate the fibrils. Mixture was incubated at 37 °C for 1 h with continuous shaking at 1000 rpm (Eppendorf ThermoMixer). Then, an additional 5 µL of 8M GuHCl was added and incubated at 37 °C for 1 at with continuous shaking at 1000 rpm. Finally, 5 µL of MilliQ water containing reducing agent (DTT) was added. Samples were desalted and analyzed by LC/MS, using a Sciex X500B QTOF mass spectrometer coupled to an Agilent 1290 Infinity II HPLC. Samples were injected onto a POROS R1 reverse-phase column (2.1 x 30 mm, 20 µm particle size, 4000 Å pore size) and desalted. The mobile phase flow rate was 300 µL/min, and the gradient was as follows: 0-3 min: 0% B, 3-4 min: 0-15% B, 4-16 min: 15-55% B, 16-16.1 min: 55-80% B, 16.1-18 min: 80% B.

The column was then re-equilibrated at initial conditions prior to the subsequent injection. Buffer A contained 0.1% formic acid in water and buffer B contained 0.1% formic acid in acetonitrile.

The mass spectrometer was controlled by Sciex OS v.3.0 using the following settings: Ion source gas 1 30 psi, ion source gas 2 30 psi, curtain gas 35, CAD gas 7, temperature 300 °C, spray voltage 5500 V, declustering potential 135 V, collision energy 10 V. Data was acquired from 400-2000 Da with a 0.5 s accumulation time and 4 time bins summed. The acquired mass spectra for the proteins of interest were deconvoluted using Bio Tool Kit within Sciex OS in order to obtain molecular weights. Peaks were deconvoluted over the entire mass range of the mass spectra, with an output mass range of 7000-9000 Da, using low input spectrum isotope resolution.

Protein digestion and enzymatic mass spectrometry. Proteins were precipitated in 23% trichloroacetic acid (Sigma-Aldrich, T0699) at 4 °C overnight. Proteins were pelleted by centrifugation 16100 x g for 30 min at 4 °C. Pelleted proteins were rinsed with 200 μL cold acetone. Air-dried proteins were solubilized in 8 M urea 100 mM Triethylammonium bicarbonate (TEAB) (Sigma-Aldrich, T7408) pH 8.5 and reduced with 5 mM Tris(2-carboxyethyl)phosphine hydrochloride (Sigma-Aldrich, C4706) and alkylated with 55 mM 2-Chloroacetamide (Sigma-Aldrich, 22790). Proteins were digested for both 3 hr and 18 hr at 37 °C in 100 mM TEAB pH 8.5 and varying concentrations of urea (0.5-1.0 M), with 0.5 ug trypsin (Promega, Madison, WI, product V5111) or chymotrypsin (Promega, V106A). Digestion was stopped by addition of formic acid to 5%. (Thermo Scientific). For elastase analysis, MS analysis, 0.5 µg of extracted fibrils were dissolved in a tricine SDS sample buffer, boiled for 2 minutes at 85 °C, and run on a Novex™ 16% tris-tricine gel system using a Tricine SDS running buffer. Gel was stained with Coomassie dye, destained, and the protein smear was cut from the gel. Samples were digested overnight with elastase (Pierce) following reduction and alkylation with DTT and iodoacetamide (Sigma–Aldrich).

Single phase analysis was performed using a Q Exactive HF Orbitrap Mass Spectrometer coupled to an Ultimate 3000 RSLC-Nano liquid chromatography system. Samples were injected onto a 75 µm ID, 15-cm long EasySpray column (Thermo) and eluted with a gradient from 0-28% buffer B over 90 min. Buffer A contained 2% (v/v) ACN and 0.1% formic acid in water, and buffer B contained 80% (v/v) ACN, 10% (v/v) trifluoroethanol, and 0.1% formic acid in water. The mass spectrometer operated in positive ion mode with a source voltage of 2.5 kV and an ion transfer tube temperature of 300 °C. MS scans were acquired at 120,000 resolution in the Orbitrap and up to 20 MS/MS spectra were obtained in the ion trap for each full spectrum acquired using higher-energy collisional dissociation (HCD) for ions with charges 2-8. Dynamic exclusion was set for 20 s after an ion was selected for fragmentation.

Raw File Processing. The raw data files were converted to .mgf format using MSConvert[33] GUI. The following filters were applied: 1) “peakPicking vendor msLevel=2-”, 2) “msLevel 2-”, 3) “scanSumming precursorTol=0.01 scanTimeTol=30 ionMobilityTol=5 sumMs1=0”.

Enabling scanSumming is an important feature of this data processing pipeline since the data was acquired with dynamic exclusion disabled. Without dynamic exclusion, the mass spectrometer can repeatedly fragment analyte ions, producing multiple spectra from the same precursor ions. While this ensures that analyte ions are more likely to be fragmented when they are most abundant, resulting in higher quality spectra, it also leads to data redundancy. ScanSumming merges these duplicate spectra, reducing data complexity and improving the signal-to-noise ratio.

Full MS Convert Command. "msconvert.exe" --mgf --32 --zlib --filter "peakPicking vendor msLevel=2-" --filter "msLevel 2-" --filter "scanSumming precursorTol=0.01 scanTimeTol=30 ionMobilityTol=5 sumMs1=0" --filter "titleMaker <RunId>.<ScanNumber>.<ScanNumber>.<ChargeState> File:"""^<SourcePath^>""", NativeID:"""^<Id^>""" "QEX3_1170178.raw" " QEX3_1170180.raw" "QEX2_1181293.raw" "QEX2_1181294.raw" "QEX2_1185163.raw"

De Novo *Search.* The .mgf files were searched using Casanovo v4.2.0.[16–18] The trypsin-digested samples were analyzed using the default Casanovo model (*casanovo_massivekb.ckpt*) provided with the v4.0.0 release. Similarly, the chymotrypsin- and elastase-digested samples were analyzed using the non-tryptic model (*casanovo_nontryptic.ckpt*) from the same release. Post-search, each of the result files was individually aligned to ensure a median mass error of 0 ppm for the top 10% of high-scoring peptides. The alignment script used can be found in the GitHub repository.

Antibody Assembly. Assembly of the antibody sequence from the *de novo* results was done with Stitch v1.5.0.[15] A customized batch file was created based on the provided monoclonal batch file. All Casanovo input files utilized a CutOffScore of 0.9 and FilterPPM of 5, except for the elastase sample, which used a CutOffScore of 0.85. The FilterPPM and CutOffScore values were determined by analyzing the observed ppm error and score distributions from the top 10% highest scoring peptides for each file. Template matching parameters were configured as follows: EnforceUnique=True, CutOffScore=10, and AmbiguityThreshold=0.9.

An additional joining region template (*>IGLJ2c IGLJ3c*) was added to improve consistency when identifying and filling the gap between the joining and variable regions. This custom template was based on the best matching joining region and several high-scoring peptides which bridged the joining and variable regions. Also, because of the presence of contaminant proteins in the samples, several of the more abundant contaminants identified during the database search were added to the Stitch contaminants file.

Cryo-EM of AL fibrils. Freshly extracted fibril samples were applied to glow-discharged R1.2/1.3, 300 mesh, Cu grids (Quantifoil), blotted with filter paper to remove excess sample, and plunge-frozen in liquid ethane using a Vitrobot Mark IV (FEI/Thermo Fisher Scientific). Cryo-EM samples were screened on Talos Arctica at the Cryo-Electron Microscopy Facility (CEMF) at The University of Texas Southwestern Medical Center (UTSW). Images were collected using a 300 keV Titan Krios microscope (FEI/Thermo Fisher Scientific) with Falcon4i detector operated at slit width of 10 eV at the Stanford-SLAC Cryo-EM Center (S^2C^2). Further details are provided in Suppl. Table 2.

Helical Reconstruction. The raw movie frames were gain-corrected, aligned, motion-corrected and dose-weighted using RELION's own implemented motion correction program.[34] Contrast transfer function (CTF) estimation was performed using CTFFIND 4.1.[35] All steps of helical reconstruction, three-dimensional (3D) refinement, and post-process were carried out using RELION 4.0.[36,37] The filaments were picked automatically using Topaz in RELION 4.0.[38,39] Particles were extracted using a box size of 512, 320, and 256 pixels with an inter-box distance of 3 asymmetrical units at helical rise of 4.75 Å. 2D classification of 512-pixel particles was only used to estimate the fibril crossover distance. 2D classifications of 320-pixel particles were used to select suitable particles for further processing. An initial 3D reference model was generated from a subset of 2D class averages using relion_helix_inimodel2d program, as previously

described.[40] Fibril helix was assumed left-handed for 3D reconstruction. 2D classifications of 256-pixel particles were used for the following refinement steps. 3D auto-refinement was followed by 3D classification without image alignment to filter out suboptimal classes, after which selected particles underwent further 3D auto-refinement. Multiple rounds of CTF refinement, 3D classification (without image alignment), auto-refinement, and post-processing were performed sequentially to reach the highest possible resolution. Final maps were further refined with 3D auto-refinement and were sharpened through post-processing. Overall resolutions were calculated from Fourier shell correlations at 0.143 between two independently refined half-maps using a soft-edged solvent mask.[41] Local resolutions were estimated using RELION's Local Resolution tool. Additional processing details are available in Suppl. Table 2.

Atomic model building and structural refinement. An initial backbone was built using alanine residues in the software COOT v0.9.8.93.[42] From there, the disulfide bond was positioned according to the structural data and residues were altered according to the predicted sequence. We made real space refinements in both COOT and Phenix v1.20.1 (*phenix.real_space_refine*).[43] We then used MolProbity to validate the model and identify areas for further refinement. The final model statistics are listed in Suppl. Table 2.

Rigid body model fitting. Maps were trimmed to just a few layers using the Map Eraser tool in ChimeraX v1.7.1 for ease of fitting.[44] Fitting the single protofilament structure into the double protofilament maps was done using Rigid Body Fit Molecule in COOT v0.9.8.93 for each individual protofilament.[42]

APR Analyses. APRs were predicted using four common prediction models: TANGO[23–25], WALTZ[19], Aggrescan[22], and AmylPred[21]. The predicted AL sequence was input to a single model. Any residues predicted to be part of an APR according to the output of the prediction model ("hit") were given a score of 1, while residues not part of an APR were given a score of 0. This process was repeated for each of the four models, and the per-residue score was calculated as the sum of hits from the four models. Criteria for calling hits were based on the requirements set by the creators of each model. Briefly, they are as follows: For TANGO, hits were counted for any stretch of 5 residues or more with a 5.0 or higher aggregation score, according to the TANGO usage guidelines. For WALTZ, hits were counted for any residue with a score greater than 0. For Aggrescan, hits were counted as any stretch of 5 or more residues with an aggregation score of 0.02 or higher. For AmylPred, hits were counted as any residues with a consensus score of 3 or more. The total sum of hits was tabulated and written into a PDB file as the B-factor for generating the heatmap model figure. The heatmap model was generated using ChimeraX v1.7.1.[44]

Solvation energy calculations. The stabilization energy per residue was calculated by the sum of the products of the area buried for each atom and the corresponding atomic solvation parameters.[45] The overall energy was calculated by the sum of energies of all residues, and assorted colors were assigned to each residue, instead of each atom, in the solvation energy map.

Figure Panels. Figures were created with ChimeraX v1.7.1, COOT v0.9.8.93, Stitchv1.5.0, and ESPript3.0 (https://espript.ibcp.fr).[15,42,44,46]

DATA AVAILABILITY

The mass spectrometry proteomics data have been deposited to MassIVE (ProteomeXchange). MassIVE accession number: . The electron density maps for the single and double protofilament structures can be accessed using EMDB codes EMD-48160 and EMD-48161. The model for the single and double protofilament structure can be accessed using PDB codes 9ELS and 9ELT.

CONTRIBUTIONS

Conceptualization: L.S.; Project Administration: L.S.; Supervision: L.S.; Funding acquisition: L.S.; Methodology: L.S., J.J.M., P.T.G., P.B.; Investigation: P.B., V.S., J.J.M., S.A.; Analysis: P.B., B.A.N., P.T.G.; Software: P.T.G.; Validation: P.B., B.A.N.; Writing; original draft: P.B.; Visualization: P.B.; Writing; review and editing: L.S., P.B., B.A.N.

ACKNOWLEDGEMENTS

We are very grateful to the patient who donated their tissue, who made this research possible. We would like to thank the UTSW Proteomics Core facility for assistance with proteomics experiments, as well as the UTSW Cryo-Electron Microscopy Facility, the UTSW Electron Microscopy Core Facility, and the national cryo-EM facilities Stanford-SLAC (project CA60) for instrumentation, technical support, and data collection. Additionally, we'd like to thank Michael Sawaya for his valuable feedback and insights on our structures.

CONFLICTS OF INTEREST

L.S. consults for Intellia Therapeutics Inc. and Attralus Inc., and Advisory Board member for Alexion Pharmaceuticals and Pfizer. The remaining authors declare no competing interests.

REFERENCES

1. Merlini, G. *et al.* Systemic immunoglobulin light chain amyloidosis. *Nat Rev Dis Primers* **4**, 38 (2018).
2. Palladini, G. & Merlini, G. What is new in diagnosis and management of light chain amyloidosis? *Blood* **128**, 159–168 (2016).
3. Del Pozo-Yauner, L. *et al.* Role of the mechanisms for antibody repertoire diversification in monoclonal light chain deposition disorders: when a friend becomes foe. *Front Immunol* **14**, 1203425 (2023).
4. Sabinot, A. *et al.* State-of-the-art review on AL amyloidosis in Western Countries: Epidemiology, health economics, risk assessment and therapeutic management of a rare disease. *Blood Reviews* **59**, 101040 (2023).

5. Charles A Janeway, J., Travers, P., Walport, M. & Shlomchik, M. J. The structure of a typical antibody molecule. in *Immunobiology: The Immune System in Health and Disease. 5th edition* (Garland Science, 2001).
6. Lefranc, M.-P. Nomenclature of the Human Immunoglobulin Lambda (IGL) Genes. *Exp Clin Immunogenet* **18**, 242–254 (2001).
7. Charles A Janeway, J., Travers, P., Walport, M. & Shlomchik, M. J. The generation of diversity in immunoglobulins. in *Immunobiology: The Immune System in Health and Disease. 5th edition* (Garland Science, 2001).
8. Berghaus, N. *et al.* Analysis of the complete lambda light chain germline usage in patients with AL amyloidosis and dominant heart or kidney involvement. *PLoS One* **17**, e0264407 (2022).
9. del Pozo Yauner, L. *et al.* Influence of the germline sequence on the thermodynamic stability and fibrillogenicity of human lambda 6 light chains. *Proteins* **72**, 684–692 (2008).
10. Kourelis, T. V. *et al.* Clarifying immunoglobulin gene usage in systemic and localized immunoglobulin light-chain amyloidosis by mass spectrometry. *Blood* **129**, 299–306 (2017).
11. Karimi-Farsijani, S. *et al.* Light chain mutations contribute to defining the fibril morphology in systemic AL amyloidosis. *Nat Commun* **15**, 5121 (2024).
12. Schulte, T. *et al.* Helical superstructures between amyloid and collagen in cardiac fibrils from a patient with AL amyloidosis. *Nat Commun* **15**, 6359 (2024).
13. Nguyen, B. A. *et al.* Structural polymorphism of amyloid fibrils in ATTR amyloidosis revealed by cryo-electron microscopy. *Nat Commun* **15**, 581 (2024).
14. Nguyen, B. A. *et al.* ATTRv-V30M amyloid fibrils from heart and nerves exhibit structural homogeneity. *Structure* (2024) doi:10.1016/j.str.2024.09.021.
15. Schulte, D., Peng, W. & Snijder, J. Template-Based Assembly of Proteomic Short Reads For De Novo Antibody Sequencing and Repertoire Profiling. *Anal Chem* **94**, 10391–10399 (2022).

16. Yilmaz, M., Fondrie, W., Bittremieux, W., Oh, S. & Noble, W. S. De novo mass spectrometry peptide sequencing with a transformer model. in *Proceedings of the 39th International Conference on Machine Learning* 25514–25522 (PMLR, 2022).
17. Yilmaz, M. *et al.* Sequence-to-sequence translation from mass spectra to peptides with a transformer model. 2023.01.03.522621 Preprint at https://doi.org/10.1101/2023.01.03.522621 (2024).
18. Melendez, C. *et al.* Accounting for digestion enzyme bias in Casanovo. 2024.05.16.594602 Preprint at https://doi.org/10.1101/2024.05.16.594602 (2024).
19. Maurer-Stroh, S. *et al.* Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat Methods* **7**, 237–242 (2010).
20. Garbuzynskiy, S. O., Lobanov, M. Yu. & Galzitskaya, O. V. FoldAmyloid: a method of prediction of amyloidogenic regions from protein sequence. *Bioinformatics* **26**, 326–332 (2010).
21. Frousios, K. K., Iconomidou, V. A., Karletidi, C.-M. & Hamodrakas, S. J. Amyloidogenic determinants are usually not buried. *BMC Struct Biol* **9**, 44 (2009).
22. Conchillo-Solé, O. *et al.* AGGRESCAN: a server for the prediction and evaluation of ‘hot spots’ of aggregation in polypeptides. *BMC Bioinformatics* **8**, 65 (2007).
23. Rousseau, F., Schymkowitz, J. & Serrano, L. Protein aggregation and amyloidosis: confusion of the kinds? *Curr Opin Struct Biol* **16**, 118–126 (2006).
24. Linding, R., Schymkowitz, J., Rousseau, F., Diella, F. & Serrano, L. A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins. *J Mol Biol* **342**, 345–353 (2004).
25. Fernandez-Escamilla, A.-M., Rousseau, F., Schymkowitz, J. & Serrano, L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotechnol* **22**, 1302–1306 (2004).

26. Enqvist, S., Sletten, K. & Westermark, P. Fibril protein fragmentation pattern in systemic AL-amyloidosis. *J Pathol* **219**, 473–480 (2009).
27. Radamaker, L. *et al.* Role of mutations and post-translational modifications in systemic AL amyloidosis studied by cryo-EM. *Nat Commun* **12**, 6434 (2021).
28. Radamaker, L. *et al.* Cryo-EM reveals structural breaks in a patient-derived amyloid fibril from systemic AL amyloidosis. *Nat Commun* **12**, 875 (2021).
29. Radamaker, L. *et al.* Cryo-EM structure of a light chain-derived amyloid fibril from a patient with systemic AL amyloidosis. *Nat Commun* **10**, 1103 (2019).
30. Swuec, P. *et al.* Cryo-EM structure of cardiac amyloid fibrils from an immunoglobulin light chain AL amyloidosis patient. *Nat Commun* **10**, 1269 (2019).
31. Puri, S. *et al.* The Cryo-EM STRUCTURE of Renal Amyloid Fibril Suggests Structurally Homogeneous Multiorgan Aggregation in AL Amyloidosis. *Journal of Molecular Biology* **435**, 168215 (2023).
32. Ahmed, Y. *et al.* Amyloid fibril polymorphism in the heart of an ATTR amyloidosis patient with polyneuropathy attributed to the V122Δ variant. 2024.05.09.593396 Preprint at https://doi.org/10.1101/2024.05.09.593396 (2024).
33. Chambers, M. C. *et al.* A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* **30**, 918–920 (2012).
34. Zivanov, J., Nakane, T. & Scheres, S. H. W. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6**, 5–17 (2019).
35. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216–221 (2015).
36. Kimanius, D., Dong, L., Sharov, G., Nakane, T. & Scheres, S. H. W. New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochem J* **478**, 4169–4185 (2021).

37. He, S. & Scheres, S. H. W. Helical reconstruction in RELION. *Journal of Structural Biology* **198**, 163–176 (2017).
38. Bepler, T. *et al.* Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat Methods* **16**, 1153–1160 (2019).
39. Lövestam, S. & Scheres, S. H. W. High-throughput cryo-EM structure determination of amyloids. *Faraday Discuss* **240**, 243–260 (2022).
40. Scheres, S. H. W. Amyloid structure determination in RELION-3.1. *Acta Crystallogr D Struct Biol* **76**, 94–101 (2020).
41. Chen, S. *et al.* High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).
42. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501 (2010).
43. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861–877 (2019).
44. Meng, E. C. *et al.* UCSF ChimeraX: Tools for structure building and analysis. *Protein Science* **32**, e4792 (2023).
45. Sawaya, M. R., Hughes, M. P., Rodriguez, J. A., Riek, R. & Eisenberg, D. S. The expanding amyloid family: Structure, stability, function, and pathogenesis. *Cell* **184**, 4857–4873 (2021).
46. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research* **42**, W320–W324 (2014).