

**Protease-sensitive regions in amyloid light chains: what a common pattern of
fragmentation across organs suggests about aggregation**

Giulia Mazzini¹, Stefano Ricagno², Serena Caminito¹, Paola Rognoni¹, Paolo Milani¹, Mario
Nuvolone¹, Marco Basset¹, Andrea Foli¹, Rosaria Russo³, Giampaolo Merlini¹, Giovanni
Palladini^{1*}, Francesca Lavatelli^{1*}

¹*Amyloidosis Treatment and Research Center, Fondazione IRCCS Policlinico San Matteo
and Università Degli Studi di Pavia, V.le Golgi 19 - 27100 Pavia, Italy;* ²*Dipartimento di
Bioscienze, Università Degli Studi di Milano, Via Celoria 26 - 20133 Milano, Italy;*

³*Dipartimento di Fisiopatologia Medico-Chirurgica e Dei Trapianti, Università Degli Studi
di Milano, Via Francesco Sforza 35 - 20122 Milano, Italy*

Running title: Light chains fragments in AL amyloidosis

Corresponding Authors:

Francesca Lavatelli, Amyloidosis Treatment and Research Center, Fondazione IRCCS
Policlinico San Matteo and University of Pavia, V.le Golgi 19, 27100 Pavia, Italy. Tel: +39
0382502984; Email: francesca.lavatelli@unipv.it

Giovanni Palladini, Amyloidosis Treatment and Research Center, Fondazione IRCCS
Policlinico San Matteo and University of Pavia, V.le Golgi 19, 27100 Pavia, Italy. Tel: +39
0382502995; Email: giovanni.palladini@unipv.it

Abstract

Light chain (AL) amyloidosis is characterized by deposition of immunoglobulin light chains (LC) as fibrils in target organs. Alongside the full-length protein, abundant LC fragments are always present in AL deposits. Defining the cleavage sites and their position on native and fibrillar LC structures is essential to understand the timing and mechanisms of proteolysis, and thus its role in pathogenesis.

Herein, by combining gel-based and mass spectrometry analyses, we identified and compared the fragmentation sites of amyloid LCs from multiple organs of an AL λ amyloidosis patient (AL-55). The positions pinpointed here in kidney and fat, alongside those previously detected in heart of the same patient, were aligned and mapped on the LC's dimeric and fibrillar states. All tissues contain fragmented LCs along with the full-length protein; the fragment pattern is coincident across organs, although microheterogeneity exists. Multiple cleavage positions were detected; some are shared, whereas some are organ-specific, likely due to a complex of proteases. Cleavage sites are concentrated in "proteolysis-prone" regions, common to all tissues. Several proteolytic sites are not accessible on native dimers, while they are compatible with fibrils. Overall, data suggest that LC fragments originate in tissues and are consistent with digestion of preformed fibrils, or with the hypothesis that proteolytic cleavage in the constant domain trigger LCs amyloidogenic potential.

This work provides a unique set of molecular data on proteolysis from *ex vivo* amyloid, which allows discussing hypotheses on role and timing of proteolytic events occurring along amyloid formation and accumulation in AL patients.

Keywords:

Amyloid fibrils, Immunoglobulin Light chains, Proteolysis, Proteomics, Amyloidogenesis.

Nonstandard abbreviations:

ACN: acetonitrile

AL amyloidosis: immunoglobulin light chain amyloidosis

CL: light chain's constant domain

Cryo-EM: cryo-electron microscopy

DTT: dithiothreitol

EA: ethanolamine

FA: formic acid

HCD: Higher-energy collisional dissociation

IAA: iodoacetamide

IEF: isoelectrofocusing

LCs: immunoglobulin light chains (note: LC in the standard hyphenated abbreviation LC-

MS/MS indicates liquid chromatography)

PTM: post-translational modifications

VL: light chain's variable domain

2D-PAGE: two-dimensional polyacrylamide gel electrophoresis

Introduction

The term amyloidosis indicates a group of diseases whose pathognomonic feature is the presence of extracellular protein aggregates with a specific cross β -sheet fibrillar structure [1-3]. Approximately 40 distinct autologous proteins are currently known precursors of localized or systemic amyloid deposits in humans [3]. Among the systemic forms, light chain amyloidosis (AL amyloidosis) is the most frequent disorder in industrialized countries, and it is caused by widespread deposition of misfolding-prone immunoglobulin free light chains (LCs), produced by a plasma cell clone in the bone marrow and circulating in the bloodstream [1, 4]. Light chains are ~215 residues long proteins, whose primary sequence includes a variable and a constant region, which fold into two distinct domains (indicated as V domain, or VL, and C domain, or CL) in the native protein [5, 6]. The combination of germline gene recombination and somatic hypermutation, occurring in the immunoglobulin genes, translates into high variability among LCs at the level of VL, so that the sequence of each patient's monoclonal protein is virtually unique [7].

Fibril deposition is consequent to loss of the precursor's native conformation, but the molecular events associated with LC fibrillogenesis *in vivo* and the tissue mechanisms counteracting the presence of amyloid fibrils are scarcely elucidated. The structures of *ex vivo* amyloid fibrils, recently resolved using cryo-electron microscopy (cryo-EM) [8-10] show that the rigid fibrillar core is composed of the variable region of the LC, and that the conversion from the native to the fibrillar form is associated with complete unfolding of VL [8, 9].

A prominent aspect in *ex vivo* AL fibrils is the fact that, alongside the full-length monoclonal protein, LC fragments are always present and account for a large fraction of all the deposited species. The major fragments contain the complete VL and CL segments of variable length [11-13]. Several reports have indeed shown that fragments containing only portions of CL are also present in tissue fibrils, sometimes representing the major fibril component, especially in

AL κ cases [11, 14-16].

To date, the origin of LC fragmentation is explained by two alternative working hypotheses: one postulates that LC cleavage near the joining region, with release of the amyloidogenic VL domain, triggers amyloidogenesis [17-19]; conversely, the other views the fragments as the result of post-deposition digestion of preformed fibrils [11, 15, 20, 21]. Defining which mechanism is predominant *in vivo* is crucial, since this knowledge is preliminary to the design of potential therapeutic strategies. Protection from proteolysis through stabilization of the LC native dimer, for example, is being explored as a novel way to prevent fibril formation and toxicity [22].

We recently implemented a mass spectrometry (MS)-based approach to identify the *termini* of the LC fragments in tissue fibrils, and utilized this method for identifying the cleavage sites that result in the heterogeneous population of LC proteoforms in the heart of two patients with AL λ amyloidosis [15]. The data showed that the cleavage sites have in common the fact of being located in poorly structured regions of the aggregates, suggesting that the proteolytic events observed in natural cardiac fibrils largely reflect post-deposition remodelling [15]. Other researchers had previously come to the same conclusion, based on the above mentioned invariable presence of the full-length LC, together with fragments containing only portions of the CL [11]. It must be noted, however, that these considerations are not intrinsically in contrast with the hypothesis that proteolysis of native LCs at specific positions triggers fibril formation, followed by recruitment of both truncated and full-length LCs.

An interesting additional observation previously derived from the gel electrophoresis analysis of natural fibrils is the fact that, in patients with multiorgan involvement, the size and pattern of fragments is similar across tissues at the individual level [11]. These seminal biochemical findings have now also a structural parallel: in fact, ultrastructural analyses recently showed

that the fibrils within a tissue can be polymorphic, *i.e.* present in multiple distinguishable morphologies, but that the distribution of polymorphs is similar in different organs from the same patient [23, 24].

However, the finding of similar fragments across distinct organs of a patient prompts additional questions in terms of timing of proteolysis. On one hand, in fact, the presence of analogous fragments in various tissues may suggest their common origin from blood before deposition; on the other side, however, the structural identity of fibrils would translate into exposure of the same protease-sensitive regions across organs.

These considerations, overall, underline the need to precisely define the cleavage sites of amyloid LCs across tissues, in order to assess if the mechanisms that generate the fragments differ between anatomical sites, and to hypothesize involved proteases. This is important missing information in the perspective of understanding the events that generate the LC proteoforms present in mature amyloid fibrils in patients.

The objective of this study was indeed to bring the knowledge on amyloid LC proteolysis a step forward, by providing a detailed comparative description of the fragmentation sites of amyloid LCs in multiple organs of an affected individual, and by analyzing the accessibility of these sites on the native and fibrillar LC structures. To this aim, we combined high resolution gel-based analyses with our recently implemented LC-MS/MS-based approach for characterizing the N- and C-terminal residues of fragments, in order to characterize and compare LC cleavage sites in fibrils from kidney and subcutaneous white fat from a patient (AL-55) deceased due to AL λ amyloidosis with previously reported proteolytic pattern of fibrils from heart of the same patient.

Results

2D-PAGE analysis: different tissues contain analogous populations of fragments. The

presence of amyloid fibrils in tissues was assessed by Congo red staining. The results of the histology analysis of kidney and subcutaneous fat are reported in Figure 1 a (histology data referring to the heart of the same patient had been previously shown [8]). Fibrils were typed as AL λ by immunoelectron microscopy. Kidney and fat tissue were used as sources of amyloid fibrils for the proteomic characterization. Two-dimensional western blot analyses of previously described heart fibrils [15] were also run in parallel herein, in order to compare the population of spots from the three tissues. After repeated cycles of tissue homogenization in Tris-EDTA buffer, the remaining tissue pellets, enriched in amyloid fibrils, were solubilized and subjected to 2D-PAGE separation and anti- λ LCs immunoblotting.

The western blots corresponding to the distinct organs are shown in Figure 1 b. Top figure shows the uncropped 2D WB from adipose tissue, representative of the pattern seen also in the other analyzed samples. The insets in bottom Figure 1 b display, side by side, the region of gels containing LCs in monomeric form in kidney, fat and heart, respectively.

Corresponding spots, or “trains” of spots, are labeled with the same numbers and letters across the panels. As previously reported [8, 12, 15], both the full-length LC and the fragments are visible as horizontal trains of spots, referable to the presence of charge modification due to PTMs, as described later. Inspection of the 2D WB data shows that the populations of fragments in the various tissues are strikingly superimposable in terms of MW and *pI* distribution of the fragments. The full-length LC, with computed MW 23.3 kDa and computed *pI* 5.50, corresponds to the train of spots labeled as 1 (spots labeled from A to G). At least 13 distinct LC fragments can be discerned (numbered from 1b to 13) in each of the three samples; the relative abundance of each fragment, however, varies considerably across tissues (Figure 1 b). We previously observed that sequence coverage from in-gel digested spots is suboptimal and does not allow reliable characterization of the fragment’s *termini* [15]; in addition, several spots contain a microheterogeneous mixture of fragments. These

considerations prompted us to skip the MS characterization of the 2D-PAGE spots, and to perform instead an in-solution LC-MS/MS-based definition of the cleavage sites, as reported later.

The presence of trains of spots with different *pI* reflects the fact that the amyloid LCs contain PTMs that modify their charge. Analysis of the gel spots was not diriment in defining which modifications differentiate the proteoforms, as previously observed [12]. We therefore investigated this aspect using LC-MS/MS data, and performed a targeted bioinformatic analysis to verify the presence of PTMs that change protein charge. Given the fact that the distribution of charge isoforms is analogous in the various tissues, we performed the PTM study only on the adipose tissue dataset, due to the particular purity of the fibrils from this site. The results are reported in Figure 1 c, in which the detected modifications are indicated over the corresponding amino acid residue on AL-55 sequence. Partial oxidation of the only Met residue (Met 3) and of Trp 100 and Trp 153 was present. In addition, deamidation was confirmed to be a prominent phenomenon in amyloid LCs. In particular, all the asparagine residues of the sequence showed partial deamidation, as well as multiple glutamines (at positions 35 in the V region, and 113 and 172 in the C region). Deamidation possibly explains the appearance of several *pI* isoforms on 2D-PAGE. Whereas asparagine deamidation can occur during sample preparation, deamidation of glutamines occurs at a much slower rate and is usually considered an *in vivo* process typical of long-lived proteins [25, 26]. In addition, the extent of LC deamidation is higher than what observed in other proteins in the sample such as β -actin (data not shown), likely indicating amyloid deamidation *in vivo*, on long-lived amyloid fibrils that may persist for years in the body.

Notably, glutamines 54 and 131, at which levels truncation of the chain has been evidenced by the terminomics study, are also modified by conversion to pyroglutamate. This modification, in fact, occurs on glutamines located at the N-terminus of proteins. No evidence

of the other modifications explored (namely acetylation, phosphorylation, nitrosylation, methylation or ubiquitination) was detected.

Investigation of the cleavage sites of light chain fragments across tissues. Taking advantage of the recently implemented LC-MS/MS-based approach for assessing the LCs' cleavage sites in *ex vivo* AL fibrils [15], we have herein characterized the C- and N-terminal residues of the LC fragments in amyloid deposits from adipose tissue and kidney. The analogous characterization in relation to the heart of AL-55 patient had been performed previously and the reported data were used in the present study for comparative purposes [15].

As observed in the heart sample, AL-55 was identified as one of the most abundant proteins also in kidney and fat, with complete MS/MS coverage of the amino acid sequence in both tissues. As previously reported in the case of the heart, other proteins were also identified in the tissue pellets with ≥ 2 unique peptides (1184 in the kidney and 353 in adipose tissue). Compared to heart and fat, the relative abundance of amyloid LCs in kidney with respect to the other tissue proteins was slightly lower, reflecting the uniquely glomerular localization of the amyloid deposits, which occupy a smaller fraction of the total tissue volume. The fewer number of cleavage sites found in kidney extracts likely reflects the lower amount of aggregates present in the tissue. The additional tissue proteins, overall, are not relevant for this study and will not be considered in the following discussion.

In adipose tissue, excluding the canonical *N-terminus* of the sequence, 14 internal amino acids were confidently identified as dimethylated, indicating that they contain a free amino group on the α carbon, and are therefore the *N-termini* of corresponding LC fragments (Fig. 2a, residues labeled with letter "F" and symbol \perp). In addition, 7 internal residues were found to be derivatized with ethanolamine on the α carboxyl group, indicating that they are the *C-termini* of corresponding LC fragments (Fig. 2 a, residues labeled with letter "F" and symbol \lrcorner). The position of all cleavage sites found in adipose tissue are colored in blue and shown in

the primary sequence, on the native structure of a dimeric LCs and on the structure of fibrillar AL-55 in Fig. 2 a-c.

Eight of the 14 N-terminally labeled positions in fat are coincident with positions identified also in the heart and/or kidney (Fig. 3 a and b). The majority of the remaining N-terminally labeled cleavage sites (Thr 47, Leu 48, Ile 49, Glu 51, Ser 66, Tyr 182, Ser 184 and Ser 195), despite not coinciding with those detected in the other organs, are however located in specific regions in AL-55 sequence, previously demonstrated to be “cleavage-sensitive” also in the heart (in particular regions 47-66 and 170-195) [15] (Fig. 2 a). Two novel N-termini (Leu 4 and Thr 17) are unique for adipose tissue and are localized in the proximal portion of the V domain. All the C-terminally labeled residues identified in fat are concentrated in a restricted region of the sequence that spans residues 115 to 139, where fragmentation sites were concentrated in heart fibrils (Fig. 2 a). Five of these 7 C-terminally labeled residues coincide with those found in the heart; an additional one is shared with kidney (Fig. 3 a and b).

In fibrils from kidney, 8 amino acids were dimethylated on the α amino group (Fig. 2 a). Six of these are in common with heart and/or fat (Fig. 3 a and b); the remaining 2 (Cys 139 and Ser 192) fall within previously mentioned “cleavage-sensitive” segments. Two amino acids were labeled on the α carboxyl group. One position (Ser 119) is labeled also in fat, and both Ser 119 and Thr 121 fall within the “cleavage-sensitive” region 115-139 (Fig. 2 a and d). An overview of the cleavage sites identified overall is provided in figure 3 a, which also indicates whether a site was present in a single organ or was found in multiple tissues.

We then considered each identified fragmentation point in the LC sequence, regardless of whether labeled at the amino or the carboxyl side, and we computed the frequency with which different amino acids occurred at either side of the bond. The results of this analysis are shown in Figure 3 c. Overall, several distinct amino acids are found at both sides of the

breakage points; however, predominance of lysine residues at the C-side and of serine residues at the N-side was observed.

The structure of cardiac amyloid fibrils from patient AL-55, previously obtained, and that of JTO LC [19], whose primary sequence is highly homologous to that of AL-55, were used as surrogates to map the LC cleavage sites identified in fat and kidney (Fig. 2 b and c).

Overall, a global look at the proteomic data obtained for the three organs (Figures 2 and 3) shows that: 1) several labeled positions in each organ are shared by at least another organ (Fig. 3 a and b). Three fragmentation sites are identical in all the tested tissues (position 53 in the VL, 135 and 177 in CL); 2) the vast majority of all fragmentation sites in the three organs are concentrated in three specific “fragmentation-rich” segments of the LC (Fig. 2a and d): residues 47 to 66 (VL domain), 115 to 139 (proximal part of the CL domain), and 170 to 195 (distal part of the CL domain). In the structure of AL-55 heart fibrils previously determined by cryo-EM [15] (Fig. 2 c), these regions are not structured and therefore are plausible targets of proteolytic degradation. Exceptions to this observation are the two residues Leu 4 e Thr 17 identified as labeled only in fat, which fall in the N-terminal segment of the V region. This zone, in heart fibrils, belongs to the structured core [8, 15]. This observation may be due to alternatively folded fibrils in fat tissue or, more likely, to partially aggregated species co-purified with amyloid deposits.

Notably, the microheterogeneity in fragmentation sites across organs is not in contrast with the substantial coincidence of 2D-PAGE spots. In fact, gel separation of fragments with very close MW and *pI* is not complete, and the same protein spot can contain multiple proteoforms differing for a minimal number of amino acids. Indeed, as mentioned above, the fact that each 2D gel protein spot contains a microheterogeneous ensemble of proteoforms had already been reported [15].

Discussion

In this work we performed a combined gel-based and gel-free proteomic characterization of the amyloid LCs extracted from adipose tissue and kidney from patient AL-55, deceased due to AL λ amyloidosis. The cleavage sites in cardiac LC fibrils from AL-55 had been previously defined and mapped on the native and fibrillar LC structures [15], and have been considered herein for comparison purposes. Overall, all the detected cleavage sites in cardiac fibrils shared the feature of being located in non-structured regions of the fibrillar aggregates: this had led to hypothesize that most of the observed cleavage events that generate the fragment visible in mature heart LC fibrils occur after deposition, on proteolysis-sensitive structural segments [15].

The present analyses extend the amyloid LC characterization to additional affected tissues, with the aim to investigate if the fragmentation mechanisms are similar across organs with distinct physiology and extracellular environments. This study provides important novel information for elucidating the mechanisms of LC processing. In particular, our data show that: I) all tissue deposits contain fragmented LCs along with the full-length protein; II) by gel analysis, the pattern of LC fragments is substantially coincident across tissues, although the relative abundance of the various proteoforms varies slightly from site to site; III) multiple cleavage positions can be detected by LC-MS/MS in all tissues, consistently with the presence of many fragments. Some of the cleavage positions are shared by various tissues, whereas some are organ-specific; IV) the cleavage sites are not uniformly distributed along the AL-55 LC sequence, but they are instead concentrated in specific “cleavage-prone” regions, which appear to be conserved in all the three tested tissues (Fig. 2 d).

The fact that all tissues contain a mixture of full-length LC and fragments is a known phenomenon; in addition, previous analyses based on custom antibodies directed against distinct regions of the light chains had shown that the fragments of each patient are similar

across organs [11, 14]. From a methodological point of view, the 2D electrophoresis used herein provides increased resolution compared to one dimensional SDS-PAGE, because it separates also proteoforms with the same MW but different charge. Indeed, our 2D-PAGE data confirm and further reinforce the concept that the overall proteolytic pattern is very similar in fat, kidney and heart.

The LC-MS/MS-based data here reported provide additional crucial information: the precise mapping of LC fragmentation sites, an aspect which was hard to define using traditional gel electrophoresis analysis. Our high-resolution data indicate that several fragmentation points are identical across tissues and a large proportion of the overall detected sites are shared by at least two of the tested organs; importantly, they also suggest that microheterogeneity exists in fragmentation positions.

The presence of similar LC fragments in different tissues could be explained by two alternative hypotheses. Such fragments may have an identical origin, *i.e.* proteolysis occur in blood and proteolysed LCs circulate before reaching different organs. An opposite possibility is that LC fragments originate locally in different tissues, from digestion of the same regions/positions. Our data are more consistent with the second hypothesis, for several reasons. First, although the fragmentation sites are concentrated in the same regions of AL-55 across different tissues (residue 47 to 66 in VL, aa 115-139 and aa 170 to 195 in CL), the population of LC fragments in mature AL fibrils is complex and rather diversified from tissue to tissue. No predominance of a specific fragment or subset of fragments can be observed. It is unlikely that such complex pattern may be so precisely reproduced across tissues from aggregation of common circulating proteolysed proteoforms. Secondly, a relevant number of proteolytic sites are not accessible on the native circulating LCs dimers while are fully compatible with aggregated LCs. Finally, the full-length LC is an important constituent of the

fibrils, as well as fragments containing portions of the C domain, indicating that the non-fragmented LC participates in the formation of amyloid fibrils.

It is noteworthy that the presence of LC fragments in AL patients' urines is a well-known phenomenon, while the existence of significant populations of fragments in blood has never been consistently demonstrated [12, 27-29]. Nevertheless, MS analyses evidenced differences between urinary and deposited fragments, which may thus be generated through distinct agents, such as tissue and urinary proteases [30]. Therefore, the LC fragments present in amyloid are more likely to originate *in loco* at the level of affected tissues. The above-mentioned microheterogeneity in cleavage points across organs likely reflects the action of different proteases in any given tissue. The overall pattern of proteolysis is thus consistent with the combined action of a complex mixture of proteolytic enzymes, as previously suggested in relation to heart fibrils [15]. Analysis of amino acids at each side of the proteolysed peptide bonds (Figure 3 c) does not suggest a clear prevalence of specific enzymes, although a marginal predominance of lysines at the C- side was present, possibly related to the action of proteases with trypsin-like specificity. Taken together, our data and the above considerations suggest that proteolysis of LCs occurs once these proteins are localised at the target organs rather than when circulating in blood.

Even though the present data only provide a snapshot of the end products of LC aggregation and do not carry any direct evidence on the mechanisms of amyloid formation, we can nevertheless use the information on the proteolytic pruning in *ex vivo* fibrils to cast hypotheses on the fibrillogenesis pathways. Indeed, models should be compatible with the features of natural aggregates, here characterised in detail.

VL domains are typically more amyloidogenic than the respective full-length LC [17-19, 31, 32], and it has been proposed that a proteolytic cleavage in the linker region is necessary to unleash the amyloidogenic potential of VL. In this context, VL domain is viewed as the real

responsible species of LC aggregation. Such hypothesis is not in agreement with our data (no cleavage sites have been detected in the joining region). Although possible, we deem it unlikely that such an aggregation prone species (*i.e.* isolated VL domain) is present only in such low amount that cannot be detected in our experiments.

Our data are compatible with the idea that no proteolysis occurs before aggregation. In the structure of LC fibrils, CL domains are invisible and thus disordered [8-10], however were CL domains totally unfolded in mature fibrils, proteolysis would be homogenous along the CL sequence, while we observed a concentration of cleavage points in “proteolysis-rich” segments (Fig. 2 d). To explain the specific proteolysis-rich segments observed on CL domains, we should therefore assume that such domains may not be natively structured, but anyway folded in a specific conformation, in which except the two proteolysis-rich segments the rest of the CL domain is protected from protease attack. Protection from the proteolytic attack may otherwise be conferred by association of specific CL segments with other molecules, such as the extracellular chaperones that are known ubiquitous components of the amyloid deposits *in vivo*.

However, another alternative explanation is also possible. The general view is that CL domains exert a stabilising effect on VL domains [19, 31, 33, 34]. Nevertheless, recent data indicate that full-length LCs with destabilised CL domains or non-optimal interdomain interactions are highly amyloidogenic [31, 35, 36]. It is possible that proteolysis on exposed loops (aa 120-131 and 181-190) may destabilise the CL domain, which in turns renders the whole native LC more aggressively amyloidogenic. This may also help VL domains to rapidly unfold and misfold to form the mature fibrillar conformation, which is readily protected from proteolysis [8, 9]. Further proteolysis of the CL domain may occur preferentially in the 115-139 and 170 to 195 regions because these regions are highly flexible and exposed due to the first proteolytic events. The CL proteolytic pattern is broadly

conserved in three different tissues of AL-55 patient, and also in the heart of another patient (AL-H7) [15], whose fibrils were formed by a LC belonging to a different germline. Such conservation may suggest a common chain of molecular events. As far as the proteolytic pattern of the VL domain is concerned, the proteolysis observed in the AL-55 and AL-H7 tissues is best viewed as post-aggregation remodelling.

Along with the above considerations, few caveats shall be taken in account. This study does not provide quantitative information on each cleavage product; even in the WB analyses, the relative abundance of the fragments cannot be rigorously compared. In the future, complementing the descriptive data provided herein with quantitative assessment of fragment abundance will be instrumental to precisely define the importance of each cleavage site and to shed additional light on the mechanisms of fibrillogenesis *in vivo*. Moreover, although the proteolytic patterns are clearly well conserved, the profile of proteolysis has so far been analyzed only in four different tissues from two different patients (AL-55 and in previously described AL-H7 [15]); therefore, additional analyses may be necessary to generalize such considerations to the whole population of AL patients.

In summary, the present data, overall, provide a precise molecular description of natural AL amyloids *ex vivo* from different organs of an individual patient. Such molecular understanding allows to critically discuss available hypotheses on the specific proteolytic events occurring upon amyloid deposition and accumulation in AL patients.

Methods

Enrichment of amyloid fibrils from tissues of patient AL-55. Amyloid fibrils were enriched from three distinct tissues (heart, kidney, subcutaneous abdominal fat) acquired during autopsy from patient AL-55, affected by AL λ amyloidosis with clinical involvement of heart and kidney and positivity of fat for amyloid during the diagnostic workup. The clinical features of patient AL-55 have been described in detail in previous publications [8]. Immediately after acquisition, unfixed tissue specimens were stored at -80 °C, and kept frozen until use. The sequence of the monoclonal λ LC from patient AL-55 derives from the IGLV6-57 germline gene and has been previously deposited in Gene Bank with code **MH670901** [8, 15]. The cryo-EM structure of *ex vivo* fibrils extracted from the heart of patient AL-55 was previously defined and has been deposited in Protein Data Bank with code **6HUD** [8].

All the procedures for enrichment of fibrils for proteomics were performed on ice in presence of a cocktail of protease inhibitors (Complete, Roche, Basel, Switzerland), as previously described in detail [15]. Briefly, each tissue specimen (approximately 250 mg) was washed 5 times in cold PBS/protease inhibitors, and then homogenized in Tris-EDTA/protease inhibitors buffer (20 mM Tris, 140 mM NaCl, and 10 mM EDTA, pH 8.0) with a glass potter pestle. The sample was centrifuged at 3,100 x g after homogenization and the supernatant was discarded to remove soluble tissue and blood proteins. The overall homogenization procedure in Tris-EDTA was repeated 10 times, after which the remaining pellet was rinsed once with ultrapure water and then used for the proteomic analyses. Use of human samples for research purposes has been approved by the Ethical Committee of Fondazione IRCCS Policlinico San Matteo and was performed in accordance with the Declaration of Helsinki.

Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) and western blotting. For protein quantification, the protein pellet was incubated for 1 h with 8 M urea in order to

dissolve protein aggregates, followed by centrifugation to remove non-solubilized material. Protein samples were quantified using micro BCA assay (Thermo Fisher Scientific, Waltham, MA, USA). For 2D-PAGE and western blotting, 100 μ g of proteins were diluted with DeStreak buffer (GE Healthcare, Chicago, IL, USA) and 0.1 M DTT. Isoelectrofocusing (IEF) was performed using 11 cm 3-10 NL pH gradient strips (Biorad, Hercules, CA, USA); the second dimension was performed using precast 8-16% polyacrylamide gradient gels (Biorad), under denaturing and reducing conditions. The detailed protocol for 2D-PAGE has been previously described [8, 15]. For immunoblotting, a rabbit polyclonal antibody against human lambda light chains (Agilent Dako, Santa Clara, CA, USA) was employed [15].

Labeling of the free N- and C-termini of the amyloid LC fragments. The procedures for covalently labeling the C- and N-terminal residues of the LC proteoforms have been previously described in detail [15]. Briefly, after reduction of disulfide bonds and alkylation of cysteines, primary amines in the reconstituted samples were dimethylated using formaldehyde and NaBH₃CN, resulting in covalent modification of the free primary amino groups, including those bound to the α carbon of N-terminal amino acids. For labeling of the C-terminal amino acids, charge-reversal derivatization of the carboxyl groups (including those bound to the α carbon of C-terminal amino acids) was obtained by amidation with ethanolamine (EA), performed on the proteins previously subjected to amine protection (i.e. through the primary amine derivatization procedure mentioned above). Prior to LC-MS/MS analysis, all samples (unlabeled protein extracts, samples subjected to amine dimethylation for study of the N-termini and samples subjected to labeling of the carboxyl groups) were digested with trypsin (Sequencing Grade Modified Trypsin, Promega, Madison, WI), as described [15].

LC-MS/MS analysis and database search. LC-MS/MS analyses were performed on a Dionex Ultimate 3000 nano-UHPLC RSLC system coupled to a Q Exactive Plus mass spectrometer

equipped with an EASY-spray ion source, as previously described in detail[15]. After washing on a trap column (PepMap100 C18, 0.3 x 5mm, 5 μ m, 100 Å), peptides were separated on an analytical column (PepMap RSLC C18, 75 μ m x 50 cm, 2 μ m, 100 Å), at a flow rate of 250 nl/min, using a linear gradient from 2% of ACN / FA 0.1% to 35% in 61 min and from 35% to 95% in 12 min. Acquisition of mass spectra in 300-2000 m/z range was operated in positive ion mode. The ten most intense precursor ions were selected for higher-energy collisional fragmentation (HCD). Data were processed using the Sequest HT-based search engine contained in the Thermo Scientific Proteome Discoverer software, version 2.0. In a preliminary database search step, the MS/MS data were searched against the human proteome database (downloaded from UniProt in August 2019 and containing 74190 entries) additioned of the sequence of AL-55; full tryptic cleavage and carbamidomethylation of cysteines as fixed modification were specified. This step allowed creating a sub-database (target database) containing the entries identified in the sample. The target database was used in the second step of spectra alignment, in which semi-tryptic cleavage and additional PTMs were indicated. Specifically, these modifications included: a) in samples labeled as described above: dimethylation of lysines and *N-terminus* (+28.031) and addition of EA on aspartate, glutamate and *C-terminus* (+43.042); b) in unlabeled samples: oxidation of methionines and triptophans (+15.995), cyclization of N-terminal glutamines (-17.027), deamidation of glutamines and asparagines (+0.984), acetylation on lysines and *N-terminus* (+42.011), ubiquitination on lysines (by investigating the presence of a di-glycine peptide bound to the ϵ -amino group of lysines (K ϵ -GG), +114.043), phosphorylation of serines, threonines and tyrosines, (+79.966), nitrosylation of triptophans and tyrosines, (+44.985), methylation of aspartic acids, glutamic acids, histidines, lysines, arginines, serines and threonines, (+14,016). A maximum of 5 dynamic modification per peptide was allowed. All spectral matches from peptide ions putatively labeled or containing PTMs were manually checked

before being confirmed.

Data availability

The mass spectrometry proteomics data have been deposited to the Proteome Xchange Consortium via the PRIDE partner repository [37, 38] with the dataset identifier PXD025185.

Other raw data will be shared upon request (contact Francesca Lavatelli, E-mail: francesca.lavatelli@unipv.it).

Figures

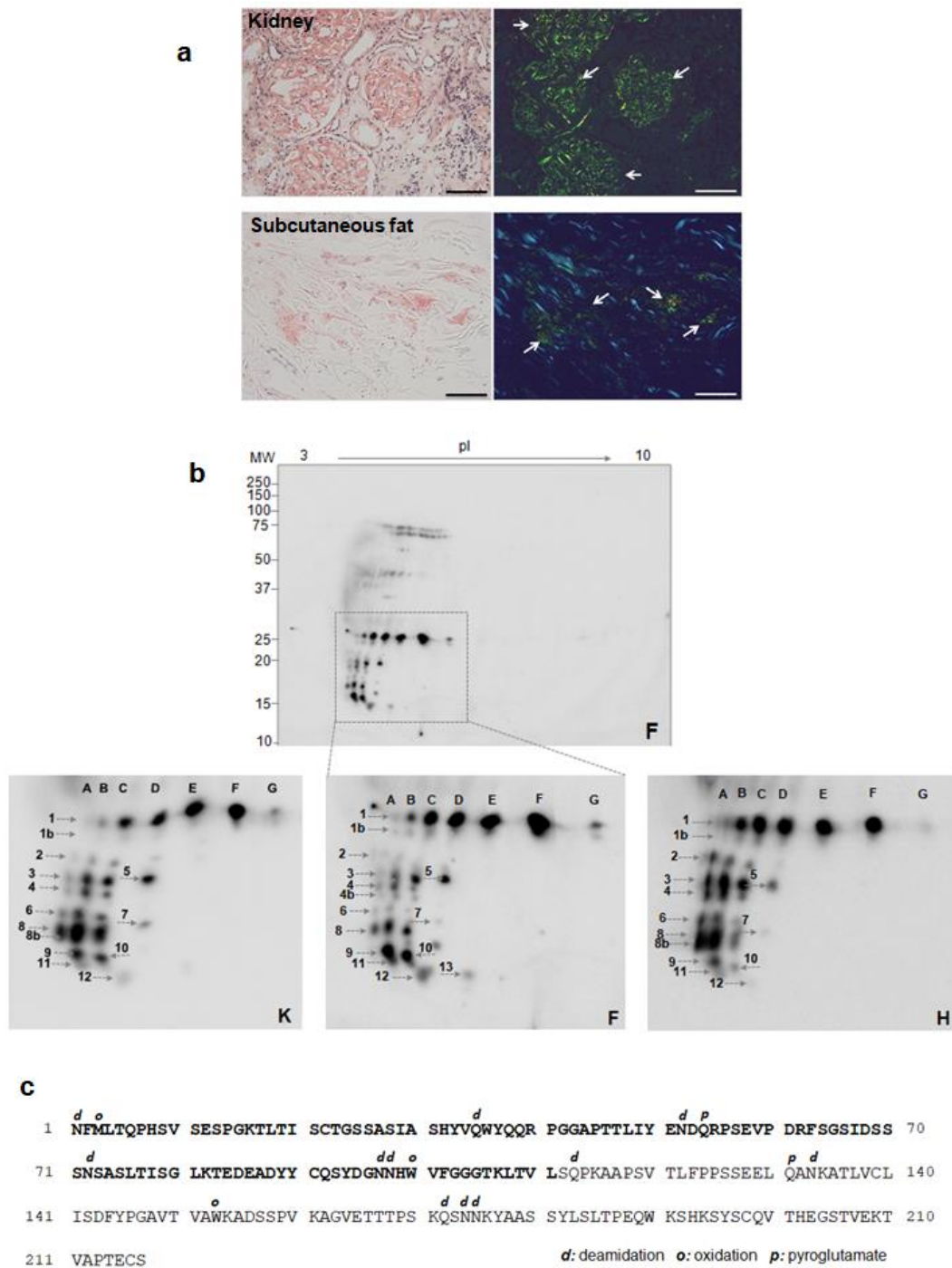


Figure 1. Amyloid LCs fibrils from tissues of patient AL-55. (a): Congo red-stained amyloid deposits in kidney and subcutaneous fat, visualized under non polarized (left panels) and polarized light (right panels). Amyloid deposits in kidney glomeruli and in adipose tissue interstitium show the pathognomonic apple-green birefringence upon light polarization (arrows). Scale bar: 100 μ m. (b): Western blotting analysis of amyloid light chains, enriched from adipose tissue, kidney and heart. The fibrillar LCs were solubilised in IEF buffer,

separated by 2D-PAGE and detected using a polyclonal anti- λ LC antibody. Top panel: representative western blot referring to the entire 2D gel from adipose tissue. The region where monomeric full-length LCs and fragments migrate is boxed (MW 25 to 10 kDa). Bottom panels: enlarged insets from the western blots of the three tissues (F: subcutaneous fat; K: kidney; H: heart), corresponding to the boxed region in the top panel. Matched pI isoforms of the full length LC across the three panels are labelled with the same letters; matched trains of fragments spots are labeled with the same numbers. (c): Post-translational modifications detected in amyloid LCs from adipose tissue, indicated above each modified amino acid (d: deamidation; o: oxidation; p: pyroglutamate). Abbreviations: F: subcutaneous fat; K: kidney; H: heart; pI: isoelectric point; MW molecular weight.

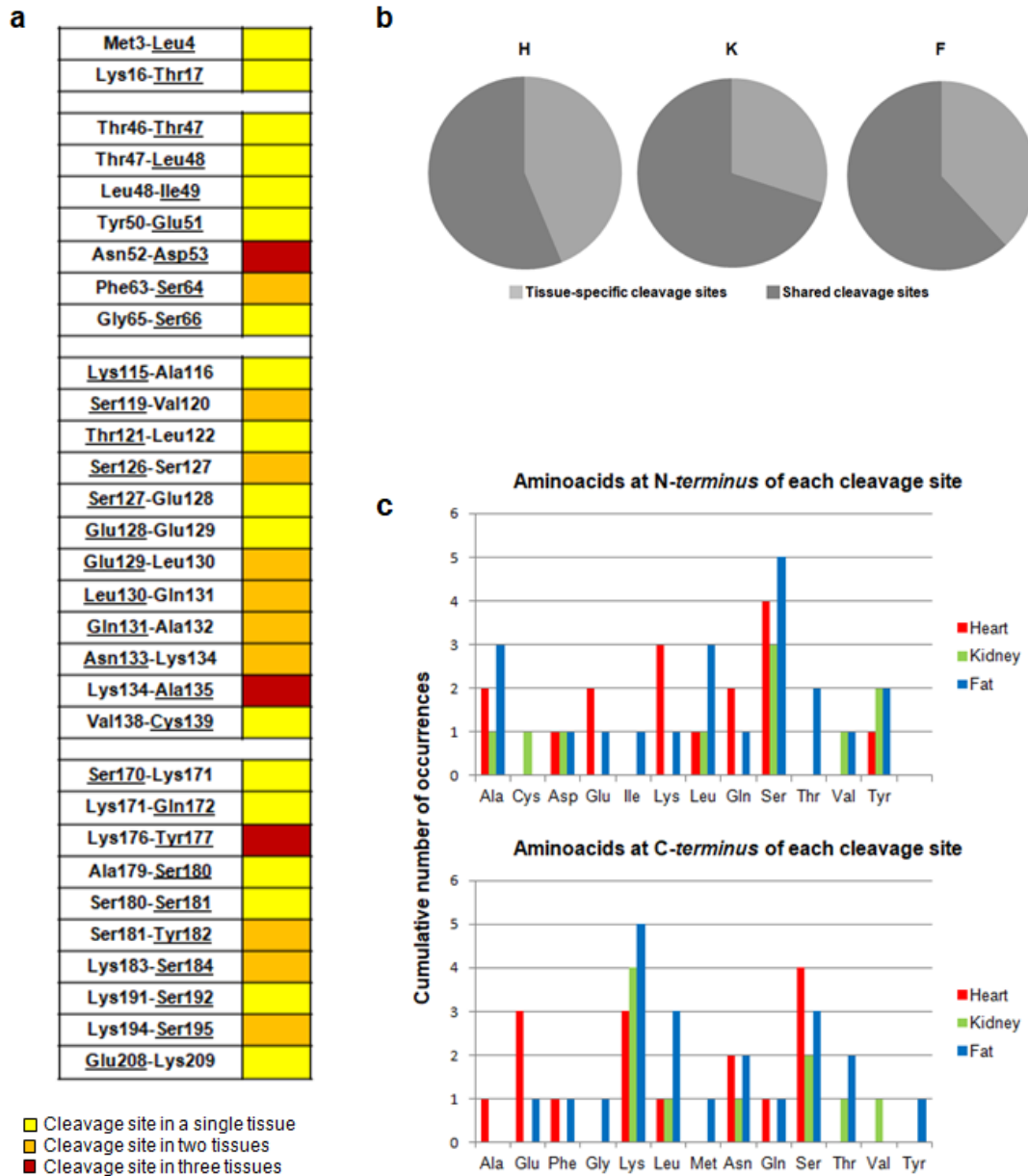


Figure 3. Analysis of the fragmentation sites in AL-55 amyloid light chains from heart, subcutaneous fat and kidney. (a): Cleaved peptide bonds, based on the fact that the amino acid at the N- or C-terminal side of the bond was identified as labelled (corresponding, respectively, to the C-terminal amino acid or to the N-terminal amino acid of the corresponding fragment). The two amino acids involved in each bond are indicated; the labelled residues are underlined. Red: cleavage sites identified in all the three tissues; orange: cleavage sites identified in two tissues; yellow: cleavage sites identified in a single tissue (further details are displayed in Figure 2a). (b): Shared and tissue-specific cleavage sites in fat, heart and kidney. (c): Cumulative occurrence of different amino acid residues at the N- or C-terminal side of the cleaved bonds in AL-55 LC, divided by organ in which cleavage was detected.

Funding and additional information

This work was supported by Fondazione Cariplo Grants 2018-0257, 2016-0489 and 2015-0591, Italian Ministry of Health Grants GR-2018-12368387, RF-2016-02361756 and IZS PLV 10/19 RC, the Italian Medicines Agency Grant AIFA-2016-02364602, the European Joint Programme on Rare Diseases E-RARE JTC 2016 Grant ReDox, Fondazione ARISLA project TDP-43-STRUCT, Fondazione Telethon GGP17036, and by Cancer Research UK, FCAECC and AIRC under the Accelerator Award 2017 Program “Early detection and intervention: Understanding the mechanisms of transformation and hidden resistance of incurable haematological malignancies”. We thank Dr. Laura Verga for the histology images. Organs drawings in graphical abstract were modified from Servier Medical Art (<http://servier.com/Powerpoint-image-bank>), licensed under a Creative Commons Attribution 3.0 Unported License.

Conflict of interest

The authors declare that they have no conflicts of interest with the contents of this article.

Author contribution

F. L., G. Mazzini, and S. R. conceptualization; F. L., G. Mazzini, S. R., P. R., and S. C. data acquisition and curation; F. L., G. Mazzini and S. C. formal analysis; F. L., S. R., G. Merlini, and G. P. supervision; F. L., S. R., M. N. and G. P. funding acquisition; F. L., G. Mazzini, S. R., P. R., and G. P. writing-original draft; F. L. and G. P. project administration; F. L., G. Mazzini, S. R., P. R., P. M., M. N., M. B., A. F., R. R., G. Merlini, and G. P. writing-review and editing; G. P. and G. Merlini resources; P. M., M. B., A. F., M. N. acquisition and clinical characterization of patients' samples; S. R. and R. R. structural mapping.

References

1. Merlini, G., Dispenzieri, A., Sanchorawala, V., Schönland, S. O., Palladini, G., Hawkins, P. N. & Gertz, M. A. (2018) Systemic immunoglobulin light chain amyloidosis, *Nat Rev Dis Primers*. **4**, 38.
2. Ke, P. C., Zhou, R., Serpell, L. C., Riek, R., Knowles, T. P. J., Lashuel, H. A., Gazit, E., Hamley, I. W., Davis, T. P., Fändrich, M., Otzen, D. E., Chapman, M. R., Dobson, C. M., Eisenberg, D. S. & Mezzenga, R. (2020) Half a century of amyloids: past, present and future, *Chem Soc Rev*. **49**, 5473-5509.
3. Benson, M. D., Buxbaum, J. N., Eisenberg, D. S., Merlini, G., Saraiva, M. J. M., Sekijima, Y., Sipe, J. D. & Westermark, P. (2020) Amyloid nomenclature 2020: update and recommendations by the International Society of Amyloidosis (ISA) nomenclature committee, *Amyloid*. **27**, 217-222.
4. Palladini, G., Milani, P. & Merlini, G. (2020) Management of AL amyloidosis in 2020, *Blood*. **136**, 2620-2627.
5. Oberti, L., Rognoni, P., Barbiroli, A., Lavatelli, F., Russo, R., Maritan, M., Palladini, G., Bolognesi, M., Merlini, G. & Ricagno, S. (2017) Concurrent structural and biophysical traits link with immunoglobulin light chains amyloid propensity, *Sci Rep*. **7**, 16809.
6. Maritan, M., Romeo, M., Oberti, L., Sormanni, P., Tasaki, M., Russo, R., Ambrosetti, A., Motta, P., Rognoni, P., Mazzini, G., Barbiroli, A., Palladini, G., Vendruscolo, M., Diomedede, L., Bolognesi, M., Merlini, G., Lavatelli, F. & Ricagno, S. (2020) Inherent Biophysical Properties Modulate the Toxicity of Soluble Amyloidogenic Light Chains, *J Mol Biol*. **432**, 845-860.
7. Bellotti, V., Mangione, P. & Merlini, G. (2000) Review: immunoglobulin light chain amyloidosis--the archetype of structural and pathogenic variability, *J Struct Biol*. **130**, 280-9.

8. Swuec, P., Lavatelli, F., Tasaki, M., Paissoni, C., Rognoni, P., Maritan, M., Brambilla, F., Milani, P., Mauri, P., Camilloni, C., Palladini, G., Merlini, G., Ricagno, S. & Bolognesi, M. (2019) Cryo-EM structure of cardiac amyloid fibrils from an immunoglobulin light chain AL amyloidosis patient, *Nat Commun.* **10**, 1269.
9. Rademaker, L., Lin, Y. H., Annamalai, K., Huhn, S., Hegenbart, U., Schönland, S. O., Fritz, G., Schmidt, M. & Fändrich, M. (2019) Cryo-EM structure of a light chain-derived amyloid fibril from a patient with systemic AL amyloidosis, *Nat Commun.* **10**, 1103.
10. Rademaker, L., Baur, J., Huhn, S., Haupt, C., Hegenbart, U., Schönland, S., Bansal, A., Schmidt, M. & Fändrich, M. (2021) Cryo-EM reveals structural breaks in a patient-derived amyloid fibril from systemic AL amyloidosis, *Nat Commun.* **12**, 875.
11. Enqvist, S., Sletten, K. & Westermark, P. (2009) Fibril protein fragmentation pattern in systemic AL-amyloidosis, *J Pathol.* **219**, 473-80.
12. Lavatelli, F., Perlman, D. H., Spencer, B., Prokaeva, T., McComb, M. E., Théberge, R., Connors, L. H., Bellotti, V., Seldin, D. C., Merlini, G., Skinner, M. & Costello, C. E. (2008) Amyloidogenic and associated proteins in systemic amyloidosis proteome of adipose tissue, *Mol Cell Proteomics.* **7**, 1570-83.
13. Eulitz, M. (1992) Amyloid formation from immunoglobulin chains, *Biol Chem Hoppe Seyler.* **373**, 629-33.
14. Olsen, K. E., Sletten, K. & Westermark, P. (1998) Fragments of the constant region of immunoglobulin light chains are constituents of AL-amyloid proteins, *Biochem Biophys Res Commun.* **251**, 642-7.
15. Lavatelli, F., Mazzini, G., Ricagno, S., Iavarone, F., Rognoni, P., Milani, P., Nuvolone, M., Swuec, P., Caminito, S., Tasaki, M., Chaves-Sanjuan, A., Urbani, A., Merlini, G. & Palladini, G. (2020) Mass spectrometry characterization of light chain fragmentation sites in

cardiac AL amyloidosis: insights into the timing of proteolysis, *J Biol Chem.* **295**, 16572-16584.

16. Engvig, J. P., Olsen, K. E., Gislefoss, R. E., Sletten, K., Wahlström, O. & Westermark, P. (1998) Constant region of a kappa III immunoglobulin light chain as a major AL-amyloid protein, *Scand J Immunol.* **48**, 92-8.

17. Brumshtein, B., Esswein, S. R., Sawaya, M. R., Rosenberg, G., Ly, A. T., Landau, M. & Eisenberg, D. S. (2018) Identification of two principal amyloid-driving segments in variable domains of Ig light chains in systemic light-chain amyloidosis, *J Biol Chem.* **293**, 19659-19671.

18. Brumshtein, B., Esswein, S. R., Landau, M., Ryan, C. M., Whitelegge, J. P., Phillips, M. L., Cascio, D., Sawaya, M. R. & Eisenberg, D. S. (2014) Formation of amyloid fibers by monomeric light chain variable domains, *J Biol Chem.* **289**, 27513-25.

19. Morgan, G. J. & Kelly, J. W. (2016) The Kinetic Stability of a Full-Length Antibody Light Chain Dimer Determines whether Endoproteolysis Can Release Amyloidogenic Variable Domains, *J Mol Biol.* **428**, 4280-4297.

20. Bohne, S., Sletten, K., Menard, R., Bühling, F., Vöckler, S., Wrenger, E., Roessner, A. & Röcken, C. (2004) Cleavage of AL amyloid proteins and AL amyloid deposits by cathepsins B, K, and L, *J Pathol.* **203**, 528-37.

21. Röcken, C., Stix, B., Brömme, D., Ansorge, S., Roessner, A. & Bühling, F. (2001) A putative role for cathepsin K in degradation of AA and AL amyloidosis, *Am J Pathol.* **158**, 1029-38.

22. Morgan, G. J., Yan, N. L., Mortenson, D. E., Rennella, E., Blundon, J. M., Gwin, R. M., Lin, C. Y., Stanfield, R. L., Brown, S. J., Rosen, H., Spicer, T. P., Fernandez-Vega, V., Merlini, G., Kay, L. E., Wilson, I. A. & Kelly, J. W. (2019) Stabilization of amyloidogenic immunoglobulin light chains by small molecules, *Proc Natl Acad Sci U S A.* **116**, 8360-8369.

23. Annamalai, K., Liberta, F., Vielberg, M. T., Close, W., Lilie, H., Gührs, K. H., Schierhorn, A., Koehler, R., Schmidt, A., Haupt, C., Hegenbart, U., Schönland, S., Schmidt, M., Groll, M. & Fändrich, M. (2017) Common Fibril Structures Imply Systemically Conserved Protein Misfolding Pathways In Vivo, *Angew Chem Int Ed Engl.* **56**, 7510-7514.
24. Annamalai, K., Gührs, K. H., 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 Engl.* **55**, 4822-5.
25. Li, X., Lin, C. & O'Connor, P. B. (2010) Glutamine deamidation: differentiation of glutamic acid and gamma-glutamic acid in peptides by electron capture dissociation, *Anal Chem.* **82**, 3606-15.
26. Lam, Y. P. Y., Chiu, C. K. C., Wootton, C. A., Hands-Portman, I., Li, M., Barrow, M. P. & O'Connor, P. B. (2020) Does deamidation affect inhibitory mechanisms towards amyloid protein aggregation?, *Chem Commun (Camb).* **56**, 9787-9790.
27. Kaplan, B., Ramirez-Alvarado, M., Dispenzieri, A., Zeldenrust, S. R., Leung, N., Livneh, A. & Gallo, G. (2008) Isolation and biochemical characterization of plasma monoclonal free light chains in amyloidosis and multiple myeloma: a pilot study of intact and truncated forms of light chains and their charge properties, *Clin Chem Lab Med.* **46**, 335-41.
28. Eulitz, M. & Linke, R. P. (1993) The precursor molecule of a V lambda II-immunoglobulin light chain-derived amyloid fibril protein circulates precleaved, *Biochem Biophys Res Commun.* **194**, 1427-34.
29. Lavatelli, F., Brambilla, F., Valentini, V., Rognoni, P., Casarini, S., Di Silvestre, D., Perfetti, V., Palladini, G., Sarais, G., Mauri, P. & Merlini, G. (2011) A novel approach for the purification and proteomic analysis of pathogenic immunoglobulin free light chains from serum, *Biochim Biophys Acta.* **1814**, 409-19.

30. Odani, S., Komori, Y. & Gejyo, F. (1999) Structural analysis of the amyloidogenic kappa Bence Jones protein (FUR), *Amyloid*. **6**, 77-88.
31. Weber, B., Hora, M., Kazman, P., Göbl, C., Camilloni, C., Reif, B. & Buchner, J. (2018) The Antibody Light-Chain Linker Regulates Domain Orientation and Amyloidogenicity, *J Mol Biol*. **430**, 4925-4940.
32. Brumshtein, B., Esswein, S. R., Salwinski, L., Phillips, M. L., Ly, A. T., Cascio, D., Sawaya, M. R. & Eisenberg, D. S. (2015) Inhibition by small-molecule ligands of formation of amyloid fibrils of an immunoglobulin light chain variable domain, *Elife*. **4**, e10935.
33. Klimtchuk, E. S., Gursky, O., Patel, R. S., Laporte, K. L., Connors, L. H., Skinner, M. & Seldin, D. C. (2010) The critical role of the constant region in thermal stability and aggregation of amyloidogenic immunoglobulin light chain, *Biochemistry*. **49**, 9848-57.
34. Morgan, G. J., Usher, G. A. & Kelly, J. W. (2017) Incomplete Refolding of Antibody Light Chains to Non-Native, Protease-Sensitive Conformations Leads to Aggregation: A Mechanism of Amyloidogenesis in Patients?, *Biochemistry*. **56**, 6597-6614.
35. Rennella, E., Morgan, G. J., Kelly, J. W. & Kay, L. E. (2019) Role of domain interactions in the aggregation of full-length immunoglobulin light chains, *Proc Natl Acad Sci U S A*. **116**, 854-863.
36. Weber, B., Hora, M., Kazman, P., Pradhan, T., Rührnöbl, F., Reif, B. & Buchner, J. (2020) Domain Interactions Determine the Amyloidogenicity of Antibody Light Chain Mutants, *J Mol Biol*. **432**, 6187-6199.
37. Deutsch, E. W., Bandeira, N., Sharma, V., Perez-Riverol, Y., Carver, J. J., Kundu, D. J., García-Seisdedos, D., Jarnuczak, A. F., Hewapathirana, S., Pullman, B. S., Wertz, J., Sun, Z., Kawano, S., Okuda, S., Watanabe, Y., Hermjakob, H., MacLean, B., MacCoss, M. J., Zhu, Y., Ishihama, Y. & Vizcaíno, J. A. (2020) The ProteomeXchange consortium in 2020: enabling 'big data' approaches in proteomics, *Nucleic Acids Res*. **48**, D1145-D1152.

38. Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., Pérez, E., Uszkoreit, J., Pfeuffer, J., Sachsenberg, T., Yilmaz, S., Tiwary, S., Cox, J., Audain, E., Walzer, M., Jarnuczak, A. F., Ternent, T., Brazma, A. & Vizcaíno, J. A. (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data, *Nucleic Acids Res.* **47**, D442-D450.
39. McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. (2011) Presenting your structures: the CCP4mg molecular-graphics software, *Acta Crystallogr D Biol Crystallogr.* **67**, 386-94.