Primary structural variations in systemic AL amyloidosis

Dissertation

Submitted for the degree of “Doctor rerum naturalium (Dr. rer. nat.)” to the Faculty of Natural Sciences, Ulm University

Submitted by

Julian David Baur

Ulm Germany

Ulm, 2023

Dean of the Faculty of Natural Sciences:	**Prof. Dr. Thorsten Bernhardt**
First reviewer:	**Prof. Dr. Marcus Fändrich**
Second reviewer:	**Prof. Dr. Nadine Schwierz**
Day of doctorate awarded:	**01.06.2023**

Abbreviations

Abbreviation	Meaning
%	Percent
°C	Centigrade
µg	Microgram
µl	Microliter
µm	Micrometer
Ala	Alanine
App	Appendix
Asn	Asparagine
AU	Arbitrary units
Aβ	Amyloid beta
BJP	Bence Jones protein
C	Constant
cDNA	Complementary deoxyribonucleic acid
CDR	Complementarity determining region
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
CID	Collision-induced dissociation
C_L	Constant light
Cryo-EM	Cryo-electron microscopy
CV	Column volumes
Cys	Cysteine
Da	Dalton
DPBS	Dulbecco's phosphate buffered saline
EDTA	Ethylenediaminetetraacetic acid
ESI	Electrospray ionization
Et al.	Et alii
FA	Formic acid
FLC	Free light chain
FR	Framework region
GL	Germline
Gln	Glutamine
Gly	Glycine
Gnd-HCl	Guanidine hydrochloride
GRAVY	Grand average of hydropathy
h	Hour
HC	Heavy chain
HCD	Higher-energy C-trap dissociation
HCl	Hydrochloric acid
His	Histidine
HPLC	High performance liquid chromatography
HS	High salt
IEC	Ion exchange chromatography
IGK	Immunoglobulin kappa
IGKV	Immunoglobulin kappa variable segment
IGLV	Immunoglobulin lambda variable segment

Ile	Isoleucine
J	Joining
k	Kilo
kDa	Kilodalton
kV	Kilovolt
LC	Light chain
LDS-PAGE	Lithium dodecyl sulphate polyacrylamide gel electrophoresis
Leu	Leucine
LS	Low salt
Lys	Lysine
M	Molar [mole per liter]
mAU	Milli absorbance units
m/z	Mass-to-charge ratio
Met	Methionine
mg	Milligram
min	Minute
ml	Milliliters
mm	Millimeter
mM	Millimolar [millimole per liter]
mRNA	Messenger ribonucleic acid
ms	Millisecond
MS	Mass spectrometry
NaCl	Natrium chloride
NaOH	Natrium hydroxide
nm	Nanometer
Phe	Phenylalanine
Pro	Proline
PTM	Post-translational modification
RPC	Reversed-phase chromatography
rpm	Revolutions per minute
s	Seconds
SAP	Serum amyloid P
SD	Standard deviation
SDC	Sodium deoxycholate
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
Ser	Serine
T20	Tween® 20
TCB	Tris calcium buffer
TEB	Tris EDTA buffer
TEM	Transmission electron microscopy
TFA	Trifluoroacetic acid
Thr	Threonine
Tris	Tris(hydroxymethyl)aminomethane
TX-114	Triton™ X-114
Tyr	Tyrosine
UB	Ultracentrifugation buffer

UC	Ultracentrifugation
V	Variable
v/v	Volume per volume
V_0	Void volume
Val	Valine
V_e	Elution volume
V_L	Variable light
w/v	Weight per volume
xg	Gravitational force
α	Alpha
β	Beta
Δ	Delta
κ	Kappa
λ	Lambda

Prologue and Acknowledgement

This PhD thesis was written under the supervision of Prof. Dr. Marcus Fändrich at the Institute of Protein Biochemistry (Ulm University). The project was part of the Research Unit FOR 2969 and dealt with the extraction and subsequent mass spectrometric analysis of amyloidogenic LCs and their fragments. Accordingly, the dissertation was carried out in close collaboration with Heidelberg University Hospital, providing patient samples and cDNA information, and the Core Unit Mass Spectrometry (Medical Department, Ulm University) under the Supervision of Sebastian Wiese, who operated the mass spectrometer.

Parts of this work have already been published in the following scientific article (License: CC BY-NC-ND 4.0, https://creativecommons.org/licenses/by-nc-nd/4.0/, © 2022 The Author(s)):

- **Baur J**, Berghaus N, Schreiner S, Hegenbart U, Schönland SO, Wiese S, Huhn S, Haupt C. Identification of AL proteins from 10 λ-AL amyloidosis patients by mass spectrometry extracted from abdominal fat and heart tissue. Amyloid. 2022 Jul 6:1-11. doi: 10.1080/13506129.2022.2095618. Epub ahead of print. PMID: 35792725.

Following articles were published with the collaboration of Julian Baur but are not part of this thesis:

- Anker S, Hinderhofer K, **Baur J**, Haupt C, Röcken C, Beimler J, Zeier M, Weiler M, Wühl E, Kimmich C, Schönland S, Hegenbart U. Lysozyme amyloidosis-a report on a large German cohort and the characterisation of a novel amyloidogenic lysozyme gene variant. Amyloid. 2022 May 9:1-10. doi: 10.1080/13506129.2022.2072198. Epub ahead of print. PMID: 35533055.
- Heerde T, Rennegarbe M, Biedermann A, Savran D, Pfeiffer PB, Hitzenberger M, **Baur J**, Puscalau-Girtu I, Zacharias M, Schwierz N, Haupt C, Schmidt M, Fändrich M. Cryo-EM demonstrates the in vitro proliferation of an ex vivo amyloid fibril morphology by seeding.

Nat Commun. 2022 Jan 10;13(1):85. doi: 10.1038/s41467-021-27688-5. PMID: 35013242; PMCID: PMC8748726.

- Banerjee S, **Baur J**, Daniel C, Pfeiffer PB, Hitzenberger M, Kuhn L, Wiese S, Bijzet J, Haupt C, Amann KU, Zacharias M, Hazenberg BPC, Westermark GT, Schmidt M, Fändrich M. Amyloid fibril structure from the vascular variant of systemic AA amyloidosis. Nat Commun. 2022 Nov 25;13(1):7261. doi: 10.1038/s41467-022-34636-4. PMID: 36433936.
- Steinebrei M, Gottwald J, **Baur J**, Röcken C, Hegenbart U, Schönland S, Schmidt M. Cryo-EM structure of an ATTRwt amyloid fibril from systemic non-hereditary transthyretin amyloidosis. Nat Commun. 2022 Oct 27;13(1):6398. doi: 10.1038/s41467-022-33591-4. PMID: 36302762; PMCID: PMC9613903.
- Radamaker L, **Baur J**, Huhn S, Haupt C, Hegenbart U, Schönland S, Bansal A, Schmidt M, Fändrich M. Cryo-EM reveals structural breaks in a patient-derived amyloid fibril from systemic AL amyloidosis. Nat Commun. 2021 Feb 8;12(1):875. doi: 10.1038/s41467-021-21126-2. PMID: 33558536; PMCID: PMC7870857.
- Radamaker L, Karimi-Farsijani S, Andreotti G, **Baur J**, Neumann M, Schreiner S, Berghaus N, Motika R, Haupt C, Walther P, Schmidt V, Huhn S, Hegenbart U, Schönland SO, Wiese S, Read C, Schmidt M, Fändrich M. Role of mutations and post-translational modifications in systemic AL amyloidosis studied by cryo-EM. Nat Commun. 2021 Nov 5;12(1):6434. doi: 10.1038/s41467-021-26553-9. PMID: 34741031; PMCID: PMC8571268.

First of all, I would like to thank Marcus Fändrich for offering me this very interesting project, for giving me the opportunity to write my thesis in his team, and for his supervision and support. Furthermore, I would like to thank Christian Haupt for his advice on scientific questions and his introduction to new scientific methods. I could learn a lot from him and had a very helpful and friendly colleague in him. I would also like to thank my other colleagues

Sara Karimi-Farsijani, Sambhasan Banerjee, Tim Moderer, Lynn Radamaker, Matthias Schmidt, Kartikay Sharma, Thomas Heerde, Peter Pfeiffer, Maximilian Steinebrei, and Giada Andreotti for insightful scientific discussions, for listening when I had some scientific problems, and for the great collaboration during the research work. I would like to thank the Medical department V of Heidelberg University Hospital, especially Ute Hegenbart and Stefan Schönland, for the great collaboration in our research work, the preparation of patient tissue, and the support concerning medical questions. I would also like to thank Natalie Berghaus, Sarah Schreiner, and Stefanie Huhn (Medical Department V, Heidelberg University Hospital) for preparing the cDNA sequences and for the great collaboration on our research paper. In addition, I would like to thank the Core Unit Mass Spectrometry and Proteomics at Ulm University, especially Sebastian Wiese, for his excellent mass spectrometric advice and support, and Carolin Weiss for the sample preparation and mass spectrometric measurements. I would like to thank the “Deutsche Forschungsgemeinschaft” for funding my work on AL amyloidosis and all the members of our research group FOR2969 for a very enjoyable time.

I also would like to thank my friends, especially Alexander Heß, for their support and helpful comments on my thesis. A special thanks goes to my parents, Petra and Bernhard Baur, who have been with me all my life and made my studies possible. Also, during my PhD thesis, they supported me in every possible way and made my work much easier. I love you very much! Finally, I would like to thank my beloved girlfriend, Alina Hoffmann, whom I met during my PhD thesis and who has supported me ever since. You are the rock in my life, and I am absolutely grateful to have you!

At last, I would like to thank my uncle Roland Baur, who passed away in 2021 and was always very interested in my research. You spurred me on to continue my work in the best possible way, and I hope you are proud of me up there!

TABLE OF CONTENTS

1. Introduction 1

1.1 Systemic AL amyloidosis 1

1.2 Immunoglobulin LCs 2

1.3 Primary structure elements of LCs promoting amyloid formation 8

1.3.1 Mutations in the primary structure affect amyloidogenicity 9

1.3.2 PTMs and their effect on amyloidogenicity of LCs 9

1.4 Mechanism of LC unfolding and fibril formation 11

1.5 Mass spectrometry 13

2. Aims 17

3. Material and Methods 18

3.1 Materials 18

3.1.1 Chemicals and compounds 18

3.1.2 Solutions and buffers 19

3.1.3 Standards, ready-to-use and other substances 20

3.1.4 Other laboratory materials 21

3.1.5 Chromatography media and columns 22

3.1.6 Other devices 22

3.1.7 Software 23

3.2 Methods 23

3.2.1 Ethics 23

3.2.2 Patient tissue 24

3.2.3 Extraction of AL fibrils from heart tissue 26

3.2.4 Extraction of AL fibrils from adipose tissue 26

3.2.4.1 Fibril extraction with collagenase from C. histolyticum 27

3.2.4.2 Fibril extraction with chloroform 27

3.2.4.3 Fibril extraction with acetone 28

3.2.4.4 Fibril extraction with detergents 28
3.2.4.5 Fibril extraction using ultracentrifugation (UC) and N-lauryl sarcosine 29
3.2.4.6 Mechanical extraction of amyloid fibrils using pestle and mortar 29
3.2.4.7 Mechanical extraction of amyloid fibrils using a self-made fractionator 30
3.2.5 Isolation of LCs from urine 31
3.2.5.1 Ammonium sulfate precipitation of LCs 31
3.2.5.2 Purification of LCs from urine by IEC 32
3.2.5.3 Analysis of the purified LCs by SEC 32
3.2.6 Analysis of blood serum for AL proteins 33
3.2.6.1 Purification of AL protein extracted from heart tissue 33
3.2.6.2 Applying the purification procedure of AL proteins from heart biopsies to serum 34
3.2.7 Liquid chromatography 35
3.2.7.1 Ion exchange chromatography 35
3.2.7.2 Reversed-phase chromatography 36
3.2.7.3 Size exclusion chromatography 36
3.2.8 LDS-PAGE 37
3.2.9 Negative stain TEM 38
3.2.10 Mass spectrometric analyses 38
3.2.10.1 Total mass analysis 39
3.2.10.2 Determination of the amino acid sequence of AL proteins by MS-based protein *de novo* sequencing 40
3.2.11 Primary structure determination of the AL proteins 42
3.2.12 Computational analysis 42
3.2.12.1 GL assignment 42
3.2.12.2 Computational analysis of the protein properties 43
3.2.13 Three-dimensional visualization 44
3.2.14 Statistics 44

4. Results 45

4.1 Characterization of LCs and their derivatives from serum and urine 45
4.1.1 Purification of FLCs from urine and investigating their oligomeric state 45

4.1.2 Analysis of LC truncations in serum 49

4.2 Establishment of a universal protocol to determine the primary structure of AL proteins extracted from abdominal fat and heart tissue 59

4.2.1 Establishment of an extraction method for amyloid fibrils from abdominal fat tissue 59

4.2.2 Establishment of a workflow to determine the primary structure of AL proteins 71

4.2.2.1 Determination of the primary structure of the AL proteins using the monoisotopic mass and the provided cDNA-derived sequence 73

4.2.2.2 Determination of the primary structure using the total mass and MS-based protein sequencing 79

4.2.2.3 Application of MS-based protein sequencing to cases without cDNA sequence 86

4.3 Determination and analysis of AL proteins' primary structures to identify parameters determining amyloidogenicity 89

4.3.1 Characterization of the cases whose primary structure could be determined 89

4.3.1.1 Amyloid fibril extraction in the cases with resolved primary structures 89

4.3.1.2 The primary structure of the AL proteins from the 19 cases 93

4.3.2 Computational analysis of the AL proteins to identify parameters determining amyloidogenicity 102

4.3.2.1 Assignment of the GLs and identification of mutations in the precursor LC sequences 103

4.3.2.2 Comparison of the mutation sites of all 19 precursor LCs 111

4.3.2.3 Mutation sites in the three-dimensional structure of a natively folded LC 116

4.3.2.4 Computational analysis of biochemical parameters from the amyloidogenic LCs in comparison to the GL 118

5. Discussion 130

5.1 Characterization of LCs and their derivatives from serum and urine 130

5.1.1 Purification of FLCs from urine and investigation of their oligomeric state 130

5.1.2 Analysis of LC truncations in serum 132

5.2 Establishment of a universal protocol to determine the primary structure of AL proteins extracted from abdominal fat and heart tissue 133

5.2.1 Extraction of amyloid fibrils for MS analysis 133

5.2.2 Primary structure determination 135

5.3 Determination and analysis of AL proteins' primary structures to identify parameters determining amyloidogenicity 137

5.3.1 Truncations in AL proteins 137

5.3.2 PTMs in AL proteins 138

5.3.3 Mutations affecting amyloidogenicity 140

5.3.3.1 GL assignment and identification of mutations 140

5.3.3.2 Mutations affecting amyloidogenicity 141

5.3.3.3 Junctional diversity and other rare mutations 143

5.3.3.4 GL-specific mutations 144

5.3.4 The effect of biochemical parameters on the amyloidogenicity of LCs 145

5.4 Conclusion 147

6. Summary 149

7. References 152

8. Appendix 170

9. Statutory declaration of independence 225

10. Curriculum vitae 226

1. Introduction

1.1 Systemic AL amyloidosis

Systemic AL amyloidosis is the most common systemic unfolding disease in Western countries, with an annual incidence of 4 to 15 new patients per 1 million inhabitants per year [1–4]. It is caused by the abnormal proliferation of a plasma cell leading to an overproduction of monoclonal light chains (LCs), which are part of intact antibodies [1,5]. Those monoclonal free LCs (FLCs) are secreted into the blood and excreted in the urine as Bence Jones proteins (BJPs) [6–8]. In AL amyloidosis, LCs that are not excreted aggregate and form amyloid fibrils, which deposit in the body. “Amyloid” refers to intra- or extracellular fibrillar deposits of abnormal protein aggregates causing pathological manifestations and diseases [9–12]. The amyloid fibrils are deposited in various organs like the heart, kidneys, liver, peripheral nerves, and gastrointestinal tract but occur most frequently in the kidneys and heart [5,10]. In addition, every AL amyloidosis patient also has fibrillar deposits in the abdominal fat, which has been used for diagnosis since 1973 [13]. Preliminary work has shown by mass spectrometry (MS), infrared spectroscopy, and transmission electron microscopy (TEM) that fibrils and their constituents are the same in different organs of a patient [14]. Thus, AL amyloid fibrils from abdominal fat are representative of all AL amyloid fibrils in the body of a patient [14]. Survival of AL amyloidosis patients mainly depends on fibril deposition in the heart [15,16]. After diagnosis, the median survival rate of AL amyloidosis patients is approximately 2 – 3.8 years [17], whereas patients with advanced heart involvement have the lowest survival rate of only 7 months [18].

One obstacle that complicates the treatment of AL amyloidosis and hampers the elucidation of the disease-causing molecular mechanisms is the patient-specific nature of the disease. A combination of variability-increasing mechanisms that occur naturally during gene recombination and the protein biosynthesis of antibodies results in a unique monoclonal LC for

each patient [19]. The differences in amino acid sequence, post-translational modifications (PTMs), and fibril structure lead to different deposition patterns and a heterogenous clinical manifestation [19]. However, it is unclear which properties make a LC amyloidogenic, or in other words, which properties lead to aggregation and fibril formation.

1.2 Immunoglobulin LCs

Immunoglobulin LCs are part of naturally occurring antibodies and are produced in the immune cells of the adaptive immune response [20]. Two LCs are connected to two heavy chains (HCs) via disulfide bonds and form an intact antibody (Fig. 1) [20].

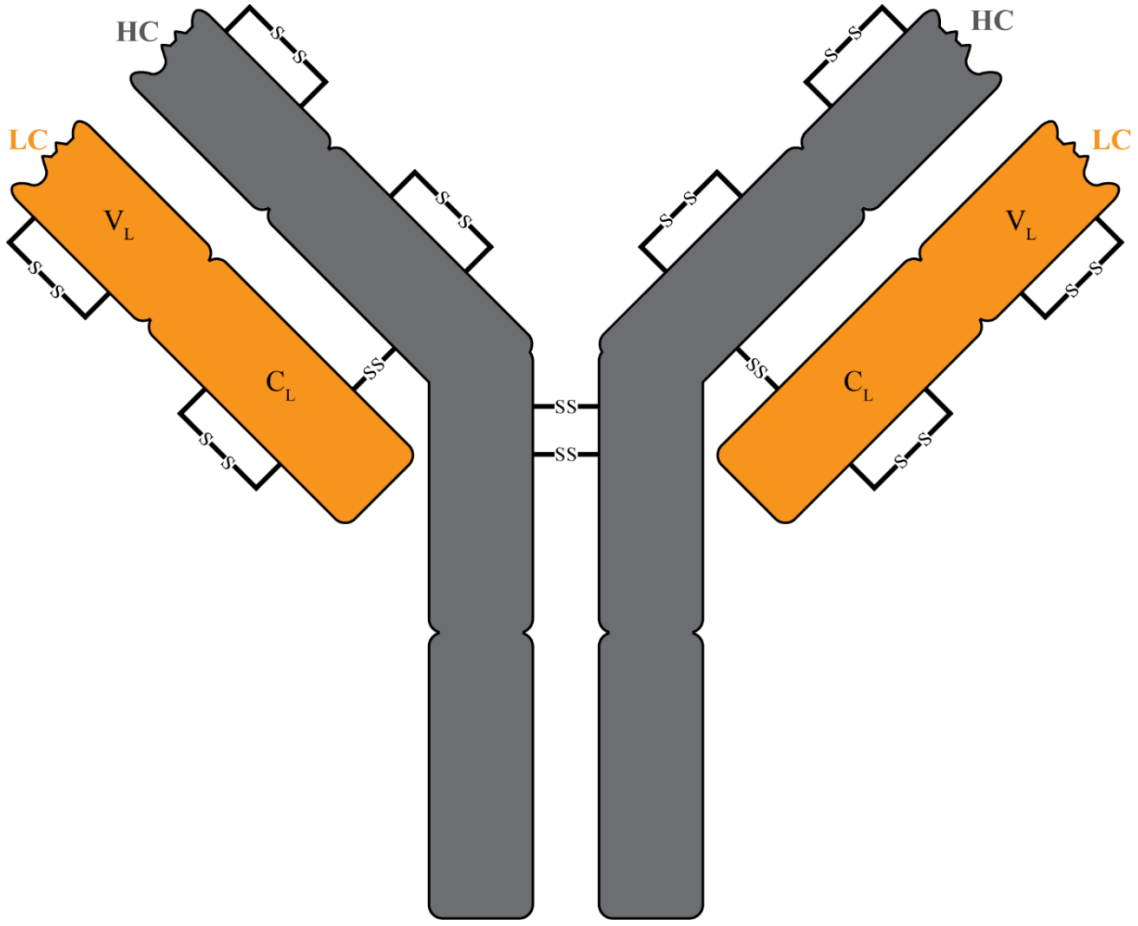


Figure 1: Structure of a Y-shaped antibody in a schematic presentation. Different subunits are marked in orange (LC) and gray (HC). LC: light chain. HC: heavy chain. SS: disulfide bond. V_L: variable light domain. C_L: constant light domain.

A LC has a molecular mass of 25 kDa and consists of two subunits, a variable light (V_L) domain, which specifically interacts with an antigen, and a constant light (C_L) domain (Fig. 1) [20]. Both domains have a characteristic immunoglobulin fold (Fig. 2) [20]. They consist of two β-sheets connected via a disulfide bond, which form a Greek key β-sandwich (Fig. 2) [21]. The β-sheets of the V_L domain are composed of 4 antiparallel β-strands packed tightly against 5 antiparallel β-strands(ABED/C''C'CFG) (Fig. 2) [21–23]. The β-sheets of the C_L domain are composed of 4 antiparallel β-strands, which are tightly packed against 3 antiparallel β-strands (ABED/CFG) (Fig. 2) [21,23]. The V_L domain contains four conserved regions called framework regions (FRs), which form the β-strands [20]. They are interrupted by three hypervariable regions called complementarity determining regions (CDRs), corresponding to three unstructured loops at the tip of the V_L domain that form the antigen binding site (Fig.2) [20,22].

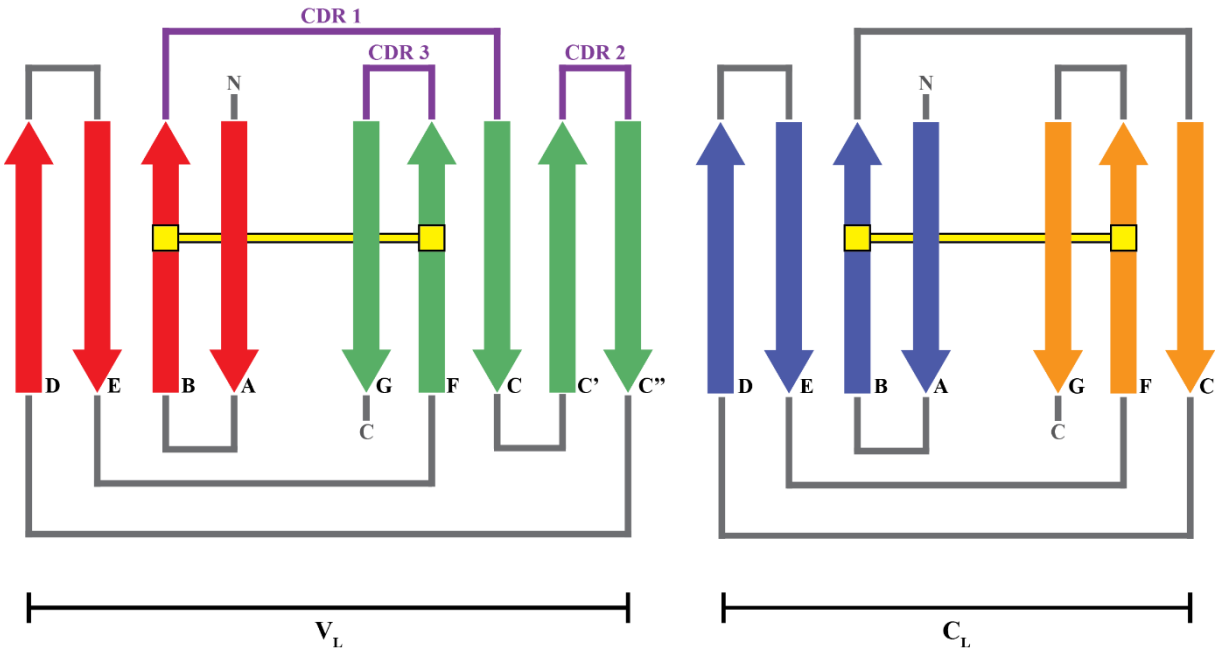


Figure 2: Topology diagram of the variable and constant domain of LCs. β-sheets: indicated by different colors (red, green, blue, orange). β-strand: arrow. Yellow line: disulfide bond connecting two β-strands. Letters: labelling of different β-strands according to their order in the three-dimensional structure. Gray/violet line: connections between β-strands. Violet lines: CDRs. Gray N and C: N- and C-terminus.

In healthy individuals, LCs are mainly bound to HCs [24]. However, under pathological conditions, FLCs occur either as monomers, dimers, or multimers and interact via the conserved amino acids necessary for binding the HCs (Fig. 3) [24,25]. FLCs can be linked covalently by an interchain disulfide bond of the C-terminal cysteine (Cys) of the C_L domain (Fig. 3 a + c) or non-covalently by hydrophobic interactions and hydrogen bonds of the V_L domain [24]. The three conserved amino acids tyrosine (Tyr) 38, glutamine (Gln) 40, and Tyr 89 are involved in interchain hydrogen bond formation (Fig. 3) [25]. Leucine (Leu) 48 and phenylalanine (Phe) 102 form hydrophobic interactions (Fig. 3) [25].

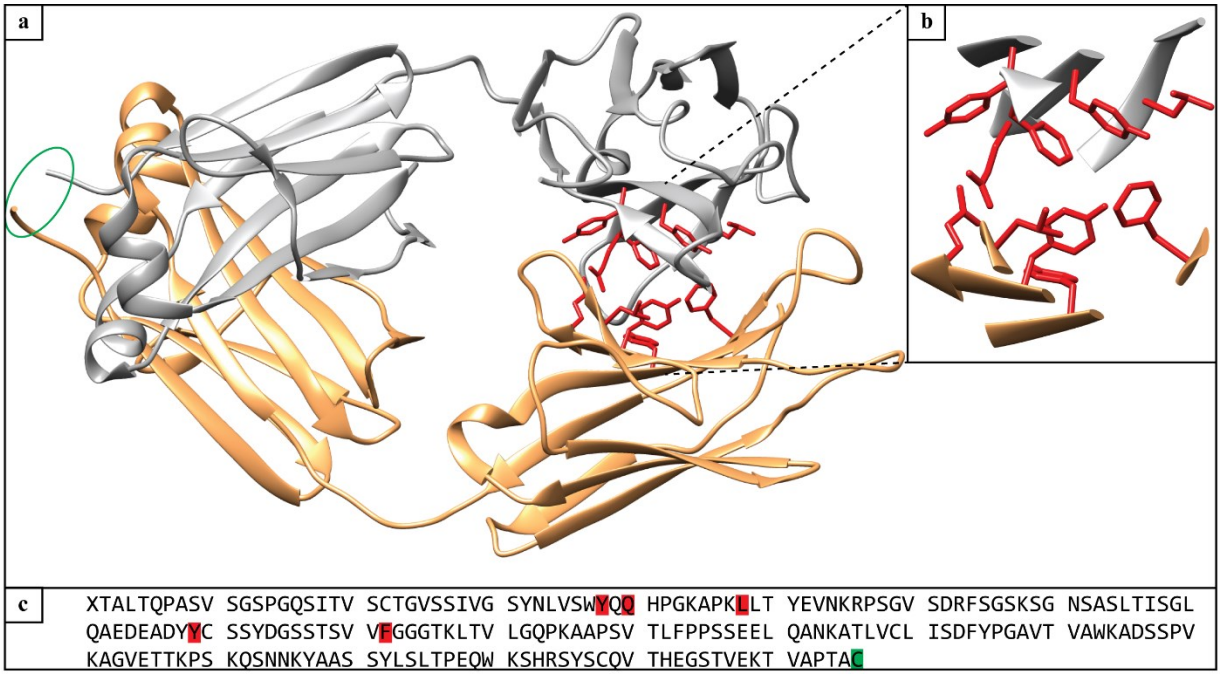


Figure 3: Three-dimensional structure of a LC dimer (PDB:1JVK) [26]***. a) Two LCs illustrated as ribbon diagram in orange and gray. Arrows: β-strands. Green circle: Cys forming covalent dimers. Red atoms: important amino acids of a non-covalent dimer. b) Close-up of the non-covalent dimer interface. c) Amino acid sequence of the LC (PDB:1JVK). Green background: Cys of covalent dimers. Red background: important amino acids of non-covalent dimers.***

LCs are highly variable, allowing for specific antigen binding and immune response [20]. One variability-increasing factor is the combinatorial diversity of gene segments encoding a LC [20]. LCs are divided into two isotypes called kappa (κ) and lambda (λ) LCs, which have their own sets of gene segments [20]. The V_L domain is encoded by two germline (GL) gene

segments, the variable (V) and joining (J) segment, whereas the C_L domain is encoded by one constant (C) segment (Fig. 4) [20]. There are roughly 40 V segments, 5 J segments, and 1 C segment that encode for κ LCs and form three separate clusters on chromosome 2 [27]. About 30 V segments, 7 J segments, and 7 C segments encode for λ LCs on chromosome 22 [28]. Whereas the λ V segments are located in one independent locus, the J and C segments are organized into pairs (Fig. 4) [29]. According to sequence similarities, V segments of κ LCs are classified into 7 families (IGKV1 to 7) and λ LCs into 11 families (IGLV1 to 11) [27,28]. A random combination of V, J, and C segments results in a large repertoire of LCs.

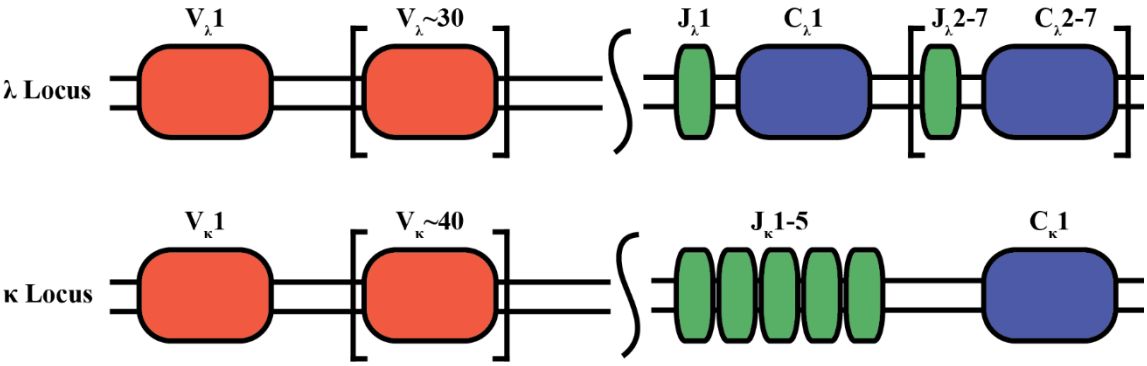


Figure 4: Genomic organization of κ and λ gene loci. λ Locus: chromosome 22. κ Locus: chromosome 2. Different gene segments are colored in red (V), green (J) and blue (C). A repetitive arrangement of gene segments is represented by square brackets.

Additional variability results from the imprecise joining of V and J segments during gene recombination and is referred to as junctional diversity (Fig. 5) [30]. Initially, an endonuclease protein complex removes the DNA between V and J segments by introducing DNA double-strand breaks (Fig. 5 a) [31]. As a consequence, the free single strands of the V and J segments interact with their complementary counterpart and form hairpin loops (Fig. 5 a) [31]. Another protein complex opens these loops by cutting at a random site (Fig. 5 b) [31]. If that cleavage occurs at a different location than the first double-strand break, bases from the non-encoded strand or inverted bases of the coding strand add onto the coding strand, changing the DNA sequence (Fig. 5 b) [31]. While it is also possible that nucleotides are lost during recombination,

additional nucleotides are referred to as P nucleotides (Fig. 5 b) [31]. In some LCs, another enzyme binds after opening the hairpin loops and adds random nucleotides at the single-stranded overlaps [30]. These nucleotides also change the final DNA sequence and are called N nucleotides [30]. DNA repair enzymes complement the single-stranded overlaps on both sides to a double strand, connect both double strands, and complete V and J recombination (Fig. 5 c) [31].

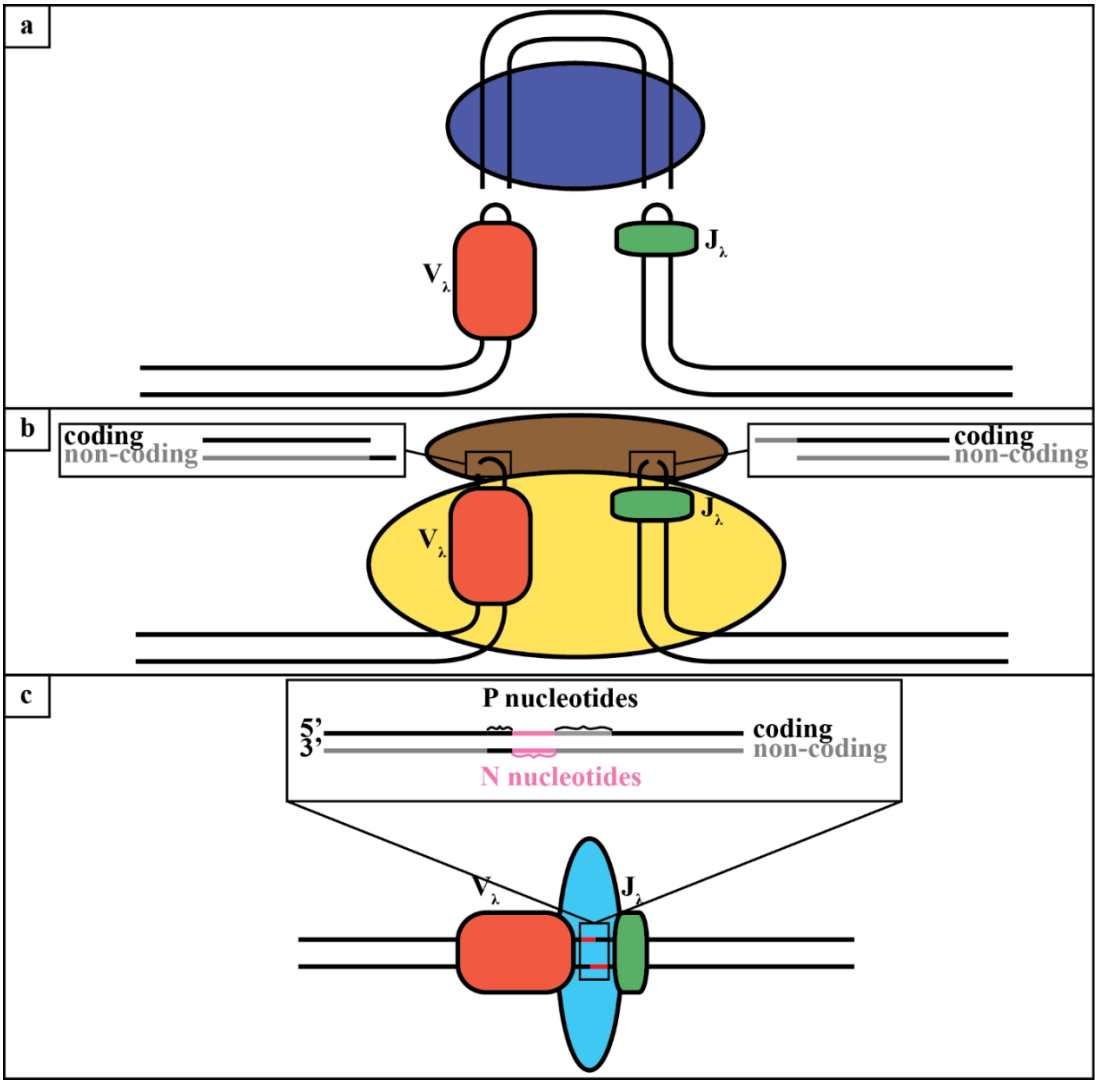


Figure 5: Schematic illustration of V and J segment recombination. a) An endonuclease complex (dark blue) creates a double strand break after the V and in front of the J segment resulting in hairpin loops between complementary DNA single strands. b) Another enzyme complex (yellow and brown) opens the hairpins at a random site leaving single stranded overlaps. c) DNA repairing enzymes (light blue) connect both DNA double strands containing non-encoded N and P nucleotides.

Somatic hypermutation further increases the variability of LCs and occurs after the recombination of one V, J, and C segment in activated B cells. [20,32]. During the transcription of primary antibodies, the double-stranded DNA is unraveled and becomes accessible for an enzyme that randomly converts cytosine residues to uracil (Fig. 6) [20]. The DNA repair machinery recognizes the uracil and replaces it with the correct base (Fig. 6). However, the enzyme leads to a large number of conversions from cytosine to uracil, increasing the probability of errors in DNA repair. If the wrong base pair is inserted during the replacement of the uracil, transition or transversion mutations occur (Fig. 6). Sometimes, the uracil is not recognized by the DNA repair machinery and remains in the DNA strand until the DNA is replicated. In DNA replication, the uracil is treated as thymine, resulting in a transition from cytosine to thymine (Fig. 6). By chance, those mutations increase the affinity to antigens and result in a more diverse repertoire of LCs.

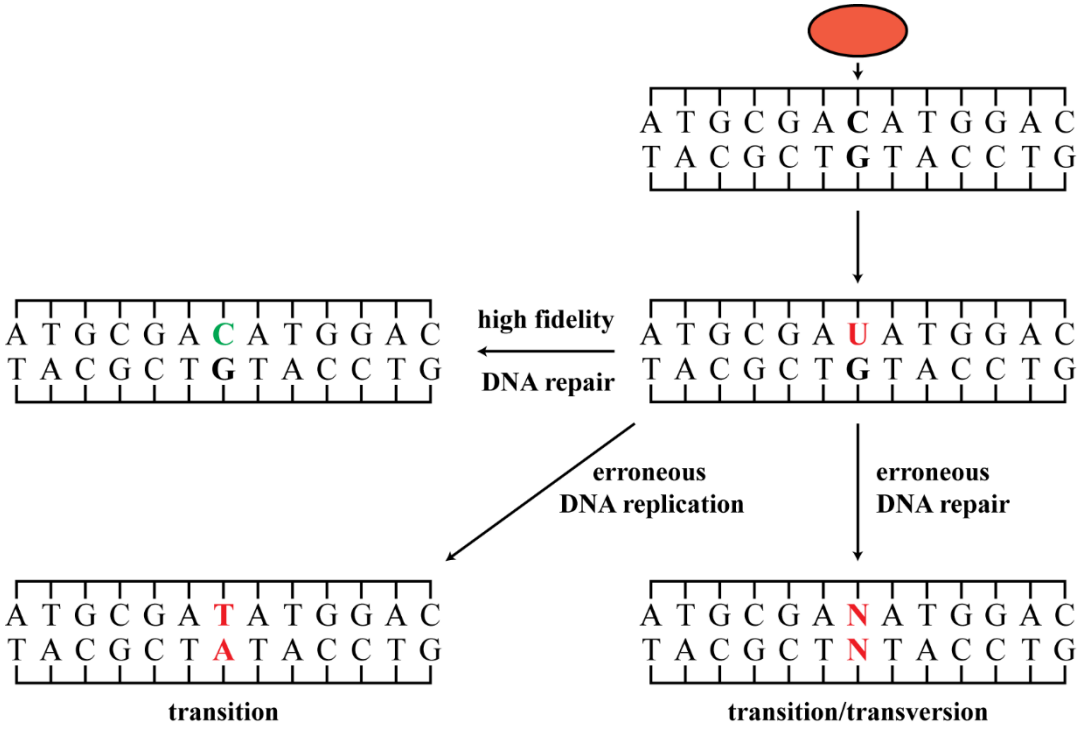


Figure 6: Schematic illustration of somatic hypermutations. Orange ellipse: Enzyme converting cytosine to uracil. A: adenine. T: thymine. G: guanine. C: cytosine. N: any base (A, T, G, C). Green: correct DNA repair. Red: incorrect nucleotide.

The combination of combinatorial diversity, junctional diversity, and somatic hypermutation leads to an enormous expansion of LC repertoire, complicating the pathogenicity of antibody-associated diseases.

1.3 Primary structure elements of LCs promoting amyloid formation

Out of all patients with a plasma cell disorder, only a minor fraction develops AL amyloidosis. For example, in monoclonal gammopathy of undetermined significance, only 9 % of the patients develop AL amyloidosis [33]. That observation raised the question of what causes fibril formation. Comparing amyloidogenic and non-amyloidogenic LCs indicates that the primary structure of LCs determines amyloidogenicity, such as the sequence differences of the LC isotypes. Evidence in support is that two-thirds of the antibodies in healthy individuals belong to the κ isotype [34], while three-quarters of amyloidogenic LCs belong to the λ isotype [35]. Sequence differences among GL V segments also appear to influence the amyloidogenicity of LCs. Of all GL V segments represented in the normal B cell repertoire, only five GL gene segments (IGLV1-44, IGLV2-14, IGLV3-1, IGLV6-57, IGKV1-33) account for approximately 60 % of all expressed amyloidogenic LCs [36–41]. Some of these GLs even influence the disease manifestation and lead to fibril deposition in specific organs, which is called organ tropism [34,35]. IGLV1-44- and IGLV2-14-derived amyloid fibrils mainly deposit in the heart and peripheral nerves [36–38,41]. IGLV6-57-derived amyloid fibrils are predominantly found in the kidneys [36–39,41], and IGKV1-33-derived fibrils mainly deposit in the liver [41].

1.3.1 Mutations in the primary structure affect amyloidogenicity

The prerequisite for more detailed analyses of the amyloidogenicity of LCs was the discovery that amyloid fibrils can be formed *in vitro* [42]. That finding was the basis for *in vitro* studies showing that mutations in the GL sequence affect fibril formation and are thus essential for determining whether patients develop AL amyloidosis. Evidence was provided by the recombinant expression of amyloidogenic V_L domains and their corresponding GL. Comparing both proteins showed that amyloidogenic V_L domains form amyloid fibrils, while GLs do not [43,44] or form fibrils much more slowly [45,46]. Introducing point mutations in the GL sequences showed that some mutations accelerate fibril formation [43,44,47,48] and destabilize the proteins on multiple levels. They reduce the thermodynamic stability of the proteins [43,44,49,50] and increase conformational dynamics [22,43,44]. They also disrupt intra- and intermolecular interactions, changing the orientation of the V_L and C_L domain [33,49], destabilizing the dimer interface of LCs [45], and making the LCs more susceptible to endoproteolysis [51]. In addition, there is evidence that mutations affect amyloidogenicity by creating aggregation-prone regions [52] and glycosylation sites [53]. The fact that single-point mutations increase the amyloidogenicity of LCs suggests that the decisive factor is the location rather than the number of mutations [22,43,44,53,54]. However, no mutation is known, affecting the amyloidogenicity of the majority of LCs. Instead, destabilizing mutations are distributed over the FRs [43,47–50,55,56], CDRs [44,57], and the linker region between the V_L and C_L domain [57].

1.3.2 PTMs and their effect on amyloidogenicity of LCs

It is suggested that PTMs influence amyloid formation, as many studies found PTMs in amyloidogenic LCs [19,53,58–60]. An extensively studied PTM is glycosylation, which is

mainly found in κ LCs [19,53] and increases the risk of AL amyloidosis [61]. However, it is unknown how glycosylation affects fibril formation [19]. Other PTMs found in amyloidogenic LCs are pyroglutamate formation [58], oxidation [58], deamidation [58], and disulfide bonds [59,60]. It is suggested that some PTMs alter the charge of the LCs [58], but the relevance to fibril formation is unclear.

Another PTM that is discussed to influence fibril formation is the truncation of the LCs. There is evidence that it is not the full-length LCs but fragments, the so-called AL proteins, that form amyloid fibrils [14,62–65]. Most fragments consist of the V_L domain and small parts of the C_L domain [19,66], while some κ AL proteins are exclusively composed of the C_L domain [67,68]. However, the fragmentation was never systematically investigated, and most primary structures of AL proteins and precursor LCs are derived from cDNA [22,37,38,66,69–72], which does not provide information on the proteins present in the fibrils. Even the Amyloid Light Chain Database contains more than 500 precursor LC sequences based solely on cDNA, while only about 20 sequences are confirmed by protein information. Therefore, very little is known about the actual constituents of the amyloid fibrils. Some studies took fragmentation for granted and sequenced only the cDNA of the V_L domain without checking what the amyloid fibrils consist of [22,69,70]. Others did peptide digestion to demonstrate that full-length LCs constitute amyloid fibrils. They prepared a homogenate from fibril-containing patient tissue, digested the proteins, and found peptides covering the complete LC sequence [58,73]. However, peptide digestion makes it impossible to identify whether the peptides originate from proteins of the deposited amyloid fibrils or non-fibril proteins present in the tissue. It is clear that fragments with a molecular mass of 12,096 to 12,168 Da were found in fibril deposits [14,63]. Whether proteolysis influences fibril formation and occurs before or after deposition has not been well investigated.

1.4 Mechanism of LC unfolding and fibril formation

The fact that amyloidogenic LCs exist in the native state or amyloid fibrils raised the question of what conformational changes need to occur before fibril formation. According to the concepts of protein folding, it is suggested that LCs in the native state need to unfold to adopt the conformation of aggregates [74]. However, more detailed investigations on the fibril formation mechanism were only possible after discovering that fibrils can be formed *in vitro* [42]. Monitoring fibril formation revealed that it comprises three phases: The lag phase, where LCs misfold; The elongation phase, where fibrils form; And the plateau phase, where the reaction reaches a steady state [10]. It became clear that various biochemical factors, divided into intrinsic and extrinsic factors, lead to misfolding [75,76]. Intrinsic factors comprise mutations and PTMs that alter hydrophobic patterns, thermodynamic stability, and charge, while extrinsic factors are pH, solutes, proteases, and protein components [75,76]. Investigation of LC unfolding in the lag phase suggested early structural rearrangements, changes in the secondary structure, and the formation of oligomers serving as seeds for the amyloid fibrils [77].

The improvement of helical reconstruction using cryo-electron microscopy (cryo-EM) [78] enabled the structural elucidation of *ex vivo* AL amyloid fibrils and revealed the structure of the misfolded LCs (Fig. 7 a + b) [59,60,79]. The amyloid fibrils show a helical twist and are composed of one protofilament consisting of one monomer per layer (Fig. 7 a+ b) [59,60,79]. The monomers stack on top of each other forming parallel in-register cross-β-sheets (Fig. 7 a) [59,60,79]. The fibrils are stabilized via backbone hydrogen bonds and interactions of stacked polar and aromatic amino acids [59,60,79]. The investigation of the AL proteins revealed substantial differences between natively-folded LCs and the fibril structure. The fibril core only consists of the V_L domain with a different number and position of β-strands compared to the native fold [59,60,79]. The fibril conformation is more flattened than the native state and

enables the stacking of monomers in the amyloid fibrils (Fig. 7 a + b) [59,60,79]. In addition, there is evidence of substantial structural rearrangements during misfolding near the disulfide bond of a natively folded V_L domain [59,60,79]. Although the disulfide bond is also present in the AL proteins (Fig. 7 b), the connected segments show a 180° switched orientation relative to the native state [59,60,79]. Those findings support the observations that misfolding is accompanied by unfolding, profound structural rearrangements, and changes in the secondary structure.

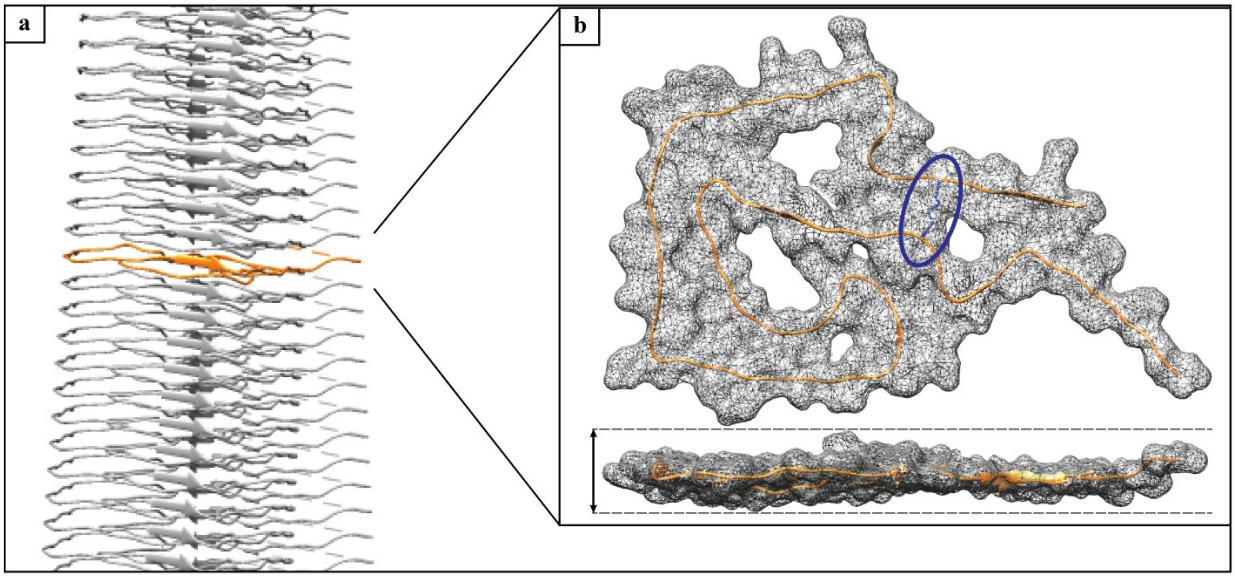


Figure 7: Three-dimensional structure of an IGLV1-derived amyloid fibril and the fibril forming monomer. a) Illustration of the amyloid fibril in a ribbon diagram. Arrows: β-strands. Lines: loops. Orange: one monomer. The amyloid fibril has been published previously and was modified according to* [59], *CC BY 4.0, https://creativecommons.org/licenses/by/4.0/. b) Illustration of one monomer in top and side view. Gray mesh: surface of the monomer. Orange: ribbon diagram. Dark blue circle: marks the dark blue disulfide bond. The figure was created from the PDB 6IC3.

In addition to structural biology studies, fibril formation was also investigated at the cellular level. There is evidence that amyloidogenic LCs are natively folded in plasma cells because they must pass the endoplasmic reticulum quality control to be secreted into the blood [80,81]. That assumption was supported by resolving the crystal structure of patient-derived

amyloidogenic LCs with a native fold [14,43]. What happens to the LCs after they are secreted into the blood is still under investigation. It is suggested that the soluble LCs are internalized by the cells at the fibril deposition site and are delivered to the lysosomal system [82]. In the cell, LCs form amyloid fibrils, which are released by lysosomes abutting the cell membrane or after cell death [82,83]. External amyloid fibrils attach to the cells and serve as seeds for soluble LCs [82]. However, no postulated cellular mechanism of fibril formation provides information on the timing of LC truncation, and as a consequence requires further investigations. That time point is particularly important for whether truncation proceeds and initiates fibril formation, as is the case for Alzheimer's Disease [84,85].

1.5 Mass spectrometry

MS is an excellent method to study proteins and is also applied in AL amyloidosis. It is used for diagnosis [86], investigation of PTMs [58,66,73], and determination of the total mass of AL proteins [14]. However, the opportunities arising from MS are increasing year by year, and MS-based protein *de novo* sequencing, in particular, could dramatically expand knowledge of the primary structures of amyloidogenic LCs. Regardless of the application, MS is used to determine the molecular mass of sample constituents [87]. For this purpose, MS generates ions that are separated in the gas phase according to their mass-to-charge ratio (m/z) [87]. A mass spectrometer consists of an ion source, a mass analyzer, and a detector [87]. In life science, electrospray ionization (ESI) is a common ionization method because of its mildness and ability to ionize large intact proteins [88]. In fact, ESI does not generate ions but transfers multiply charged ions from the solution to the gas phase [89]. For this purpose, charged molecules are sent through a spray capillary located in front of a cylindrical counter electrode (Fig. 8) [90]. A high voltage at the tip of the capillary generates an electric field that initiates a spray of small droplets containing charged ions (Fig. 8) [90]. As the aerosol expands towards the counter

electrode, the solvent evaporates, and the desolvated ions enter the mass analyzer through a small orifice of the counter electrode (Fig. 8) [90,91]. ESI results in multiply charged ions, shifting the m/z values of ions in an accessible range for most mass analyzers.

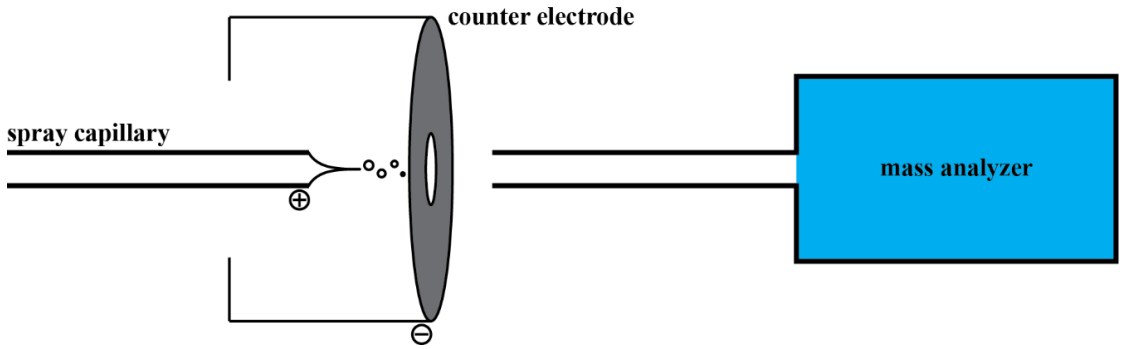


Figure 8: Schematic illustration of ESI. A dissolved analyte is sent through a spray capillary with an applied voltage at its tip. The resulting electric field initiates a spray of droplets accelerating toward the counter electrode. After evaporation of the solvent, the charged analytes enter the mass analyzer through an orifice of the counter electrode.

There are several mass analyzers, such as quadrupole and orbitrap, used individually or in combination to achieve better results. The simplest quadrupole mass analyzer consists of four hyperbolic or cylindrical rod electrodes arranged in a square configuration (Fig. 9 a) [92]. The two opposite rods have the same applied potential composed of a DC and an AC voltage [92]. For a given DC and AC voltage, only ions of a specific m/z value can pass through the quadrupole while other ions are retained [92,93]. A mass spectrum is acquired by monitoring the ions passing through the quadrupole while varying the voltages [93]. The quadrupole thus acts as a mass filter, which is often used in front of an Orbitrap to increase the versatility of the mass spectrometer [92–94]. The Orbitrap is an ion trap with high resolving power and accurate mass determination [95]. It consists of a spindle-shaped central electrode set at an attractive electric potential and a barrel-shaped outer electrode (Fig. 9 b) [96]. Ions approach the central electrode in a decentralized, tangential manner and are trapped in a rotational motion along the axes of the central electrode due to an equilibrium between the attractive electrostatic and opposing centrifugal force (Fig. 9 b) [95,96]. The m/z values of the analytes are calculated from

the oscillation frequency of the ions detected at the outer electrode [95,96]. For a suitable instrument, accurate ion injection in terms of injection angle, timing, velocity, and the number of injected ions must be ensured [94,96]. Therefore, an adapted quadrupole is employed in front of every orbitrap, accumulating, storing, and thermalizing the ions before injection [94]. Due to its bent C-shape, this quadrupole is called a C-trap (Fig. 9 b) [94,95].

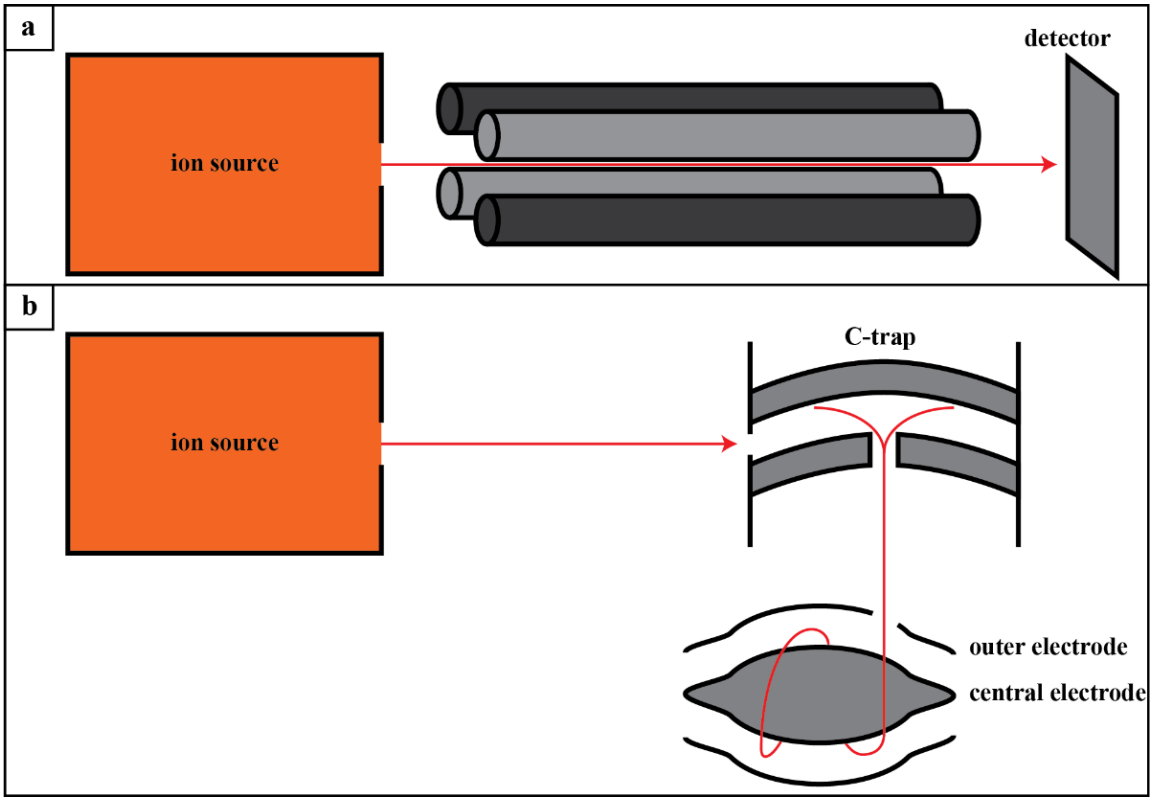


Figure 9: Illustration of different mass analyzers. a) Quadrupole: An ion trap that is open for ions with a specific m/z ratio by adjusting the potential of the four cylindrical rods. b) Orbitrap: An ion trap that uses a preceding C-trap to collect the analytes and inject them in a specific way into the actual orbitrap consisting of a spindle-shaped central electrode and a barrel-shaped outer electrode.

When investigating heterogeneous samples or determining MS-based protein *de novo* sequences, the combination of at least two mass analyzers has become very useful, called tandem mass spectrometry (tandem MS) or mass spectrometry/mass spectrometry (MS/MS) [97,98]. The advantage of MS/MS is the ability to fragment selected ions [97]. A common fragmentation technique is collision-induced dissociation (CID), where the ion beam passes

through a collision cell [99]. Inside that cell, the ions collide with a neutral gas like N_2 resulting in the fragmentation of the analyte [99]. If low-mass fragment ions are needed, higher-energy C-trap dissociation (HCD) can be used [100]. HCD is limited to tandem MS with Orbitrap instruments because it uses the C-trap as a collision cell [100]. This work benefits from the versatility of MS and uses it for a detailed study of the primary structure of AL proteins.

2. Aims

There is overwhelming evidence that the primary structural features of the LCs are crucial for the development of AL amyloidosis. The decisive factors are mutations that change biochemical properties and destabilize the LCs or their dimer interface. The destabilized LCs undergo structural rearrangements and possibly fragmentation and form the AL proteins that assemble into the disease-causing amyloid fibrils. However, the primary structure of LCs or their fragments that form amyloid fibrils is rarely known, and most primary structural information is based on cDNA. Therefore, essential information about PTMs and truncations is frequently missing. It is still unclear whether proteolysis occurs before fibril formation and if PTMs affect amyloidogenicity. Furthermore, only some mutations could be related to increased amyloidogenicity, and it is unclear whether general parameters affect fibril formation. To address these issues, this thesis aimed at the following objectives:

Aim 1: Characterization of LCs and their derivatives from serum and urine.

Aim 2: Establishment of a universal protocol to determine the primary structure of AL proteins extracted from abdominal fat and heart tissue.

Aim 3: Determination and analysis of AL proteins' primary structures to identify parameters determining amyloidogenicity.

The results of the present work should expand the repertoire of known primary structures of fibril-forming AL proteins, provide a better understanding of the mechanism of amyloid fibril formation, and thus help to find new ways to inhibit fibril formation.

3. Material and Methods

3.1 Materials

3.1.1 Chemicals and compounds

Table 1: Chemicals and compounds used throughout this thesis with the corresponding manufacturer.

Product	Manufacturer
Acetic acid (glacial)	VWR International
Acetone	Sigma-Aldrich
Acetonitrile	Sigma-Aldrich
Ammonium bicarbonate ($(NH_4)HCO_3$)	Applichem
Ammonium sulphate	Carl Roth GmbH
Brilliant blue R250	Carl Roth GmbH
Calcium chloride dihydrate	Carl Roth GmbH
3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS)	Carl Roth GmbH
Chloroform	VWR
Chymotrypsin	Thermo Fisher Scientific
Collagenase from *Clostridium histolyticum*	Sigma-Aldrich
Di-sodium hydrogen phosphate	Carl Roth GmbH
Dispase II	Sigma-Aldrich
Dithiothreitol	AppliChem
Elastase	Promega
Ethanol absolute	Sigma-Aldrich
Ethanolamine	Carl Roth GmbH
Formic acid (FA)	VWR
Guanidine hydrochloride (Gnd-HCl)	Carl Roth GmbH
Hydrochloric acid (HCl) (fuming)	Carl Roth GmbH
HCl (37 %)	Applichem
Iodoacetamide	Sigma-Aldrich
Isopropyl alcohol	Sigma-Aldrich
LysC	Thermo Fisher Scientific
Methanol	Carl Roth GmbH
N-lauryl sarcosine sodium salt	Sigma-Aldrich
Pepsin	Merck
Sodium azide	Carl Roth GmbH
Sodium chloride	Carl Roth GmbH
Sodium deoxycholate (SDC)	Sigma-Aldrich
Sodium dihydrogen phosphate monohydrate (NaH_2PO_4)	Carl Roth GmbH
Sodium dodecyl sulfate (SDS)	Carl Roth GmbH
Sodium hydroxide	Sigma-Aldrich
Sucrose	Sigma-Aldrich
Titriplex® III (disodium dihydrogen ethylenediaminetetraacetate dihydrate)	Sigma-Aldrich
Trifluoroacetic acid (TFA)	Carl Roth GmbH
Triton™ X-114 (TX-114)	Sigma-Aldrich

Trizima® base/Tris(hydroxymethyl)aminomethane (Tris)	Sigma-Aldrich
Trypsin	Thermo Fisher Scientific
Tween® 20 (T20)	Carl Roth GmbH
Uranyl acetate	Polyscience Inc.
Urea	Carl Roth GmbH
Xylene	VWR

3.1.2 Solutions and buffers

Table 2: Solutions and buffers used in this thesis with ingredients and corresponding concentrations.

Solution/buffer	Constituents	Concentration
Alkylation buffer	$(NH_4)HCO_3$ (pH = 8)	0.05 M
	Iodoacetamide	0.005 M
DTT buffer	$(NH_4)HCO_3$ (pH = 8)	0.05 M
	Dithiothreitol	0.055 M
High salt (HS) extraction buffer	Tris-HCl (pH = 8)	0.02 M
	NaCl	0.4 M
	$CaCl_2$	0.002 M
High performance liquid chromatography (HPLC) buffer A	FA	0.1 % (v/v)
HPLC buffer B	FA	0.1 % (v/v)
	Acetonitrile	86 % (v/v)
Ion exchange chromatography (IEC) elution buffer (serum/AL protein)	Ethanolamine (pH = 10)	0.02 M
	NaCl	1 M
IEC elution buffer (urine)	Ethanolamine (pH = 9)	0.02 M
	NaCl	1 M
IEC running buffer (serum/AL protein)	Ethanolamine (pH=10)	0.02 M
IEC running buffer (urine)	Ethanolamine (pH = 9)	0.02 M
Lithium dodecyl sulphate polyacrylamide gel electrophoresis (LDS-PAGE) destaining solution	Ethanol	20 % (v/v)
	Acetic acid	10 % (v/v)
LDS-PAGE staining solution	Ethanol	30 % (v/v)
	Acetic acid	10 % (v/v)
	Brilliant blue R250	2.5 g/L
Low salt (LS) extraction buffer	Tris-HCl (pH = 8)	0.02 M
	NaCl	0.138 M
	$CaCl_2$	0.002 M
MS dissociation buffer	Tris-HCl (pH = 8)	0.05 M
	Gnd-HCl	6 M
Protease buffer (Chymotrypsin)	Tris-HCl (pH = 8)	0.05 M
Protease buffer (Elastase)	Tris-HCl (pH = 9)	0.05 M
	$CaCl_2$	0.01 M
Protease buffer (LysC)	$(NH_4)HCO_3$ (pH = 8)	0.05 M

Protease buffer (Pepsin)	HCl (pH = 1.5)	0.04 M
Protease buffer (Trypsin)	$(NH_4)HCO_3$ (pH = 8)	0.05 M
Reversed-phase chromatography (RPC) running buffer	TFA	0.1 % (v/v)
RPC elution buffer	TFA	0.1 % (v/v)
	Acetonitrile	86 % (v/v)
Size exclusion chromatography (SEC) buffer	NaH_2PO_4 (pH = 6.8)	0.2 M
	NaCl	0.1 M
Tris Calcium Buffer (TCB)	Tris-HCl (pH = 8)	0.02 M
	NaCl	0.138 M
	$CaCl_2$	0.002 M
	NaN_3	0.1 % (v/v)
Tris EDTA Buffer (TEB)	Tris-HCl (pH = 8)	0.02 M
	NaCl	0.14 M
	EDTA	0.01 M
	NaN_3	0.1 % (v/v)
Ultracentrifugation buffer (UB)	Tris-HCl (pH = 7,4)	0.01 M
	NaCl	0.8 M
	Sucrose	10 % (w/v)
	EDTA	0.005 M

3.1.3 Standards, ready-to-use and other substances

Table 3: Standards and ready to use substances

Substance	**Manufacturer**
Aprotinin (LMW Kit)	Cytiva
Blue dextran 2000 (LMW Kit)	Cytiva
BlueEasy Prestained Protein Marker	Nippon Genetics Europe
Carbonic anhydrase (LMW Kit)	Cytiva
cOmplete Tablets, Mini EDTA-free, EASYpack	Roche
cOmplete Tablets, Mini, EASYpack	Roche
Conalbumin (LMW Kit)	Cytiva
Dulbecco's Phosphate Buffered Saline (DPBS) (1x) [-] $CaCl_2$ [-] $MgCl_2$	Thermo Fisher Scientific
NuPAGE™ Antioxidant	Thermo Fisher Scientific
NuPAGE™ LDS Sample Buffer (4x)	Thermo Fisher Scientific
NuPAGE™ MES SDS Running Buffer (20x)	Thermo Fisher Scientific
NuPAGE™ Sample Reducing Agent (10x)	Thermo Fisher Scientific
Pierce™ LTQ ESI Positive Ion Calibration solution	Thermo Fisher Scientific
Ribonuclease A (LMW Kit)	Cytiva

3.1.4 Other laboratory materials

Table 4: Other laboratory materials used

Product	Manufacturer
Amicon® Ultra-0.5 Centrifugal Filter Device (3k)	Merck Millipore
Amicon® Ultra-15 Centrifugal Filter Device (3k)	Merck Millipore
Cellstar® tubes, PP, conical bottom, 15 ml	Greiner bio-one
Cellstar® tubes, PP, conical bottom, 50 ml	Greiner bio-one
Cutfix disposable scalpel, stainless steel, 22	B. Braun
Disposable pasteur pipettes, graduated	Carl Roth GmbH
Disposal bags	Brand
Dumont tweezer, style N5, Biological Grade, Dumoxel®	Dumont
Eppendorf pipette, Research® plus (10 µl)	Eppendorf
Eppendorf pipette, Research® plus (100 µl)	Eppendorf
Eppendorf pipette, Research® plus (1000 µl)	Eppendorf
Eppendorf pipette, Research® plus (2.5 µl)	Eppendorf
Eppendorf tubes (5.0 ml)	Eppendorf
Formvar/Carbon Coated Copper Grids, 200 mesh	Electron Microscopy Science
Injekt® syringe, 5 ml	B. Braun
Injekt®-F, single-use fine dosage syringe	B. Braun
MF-Millipore™ membrane (47 mm, 0.22 µm, MCE membrane)	Merck Millipore
Micro-centrifuge tubes for high g-force (1.5 ml)	VWR
Nitrile Powder Free Examination Gloves, Atlantis	Amadex
NuPAGE™ 4-12% Bis-Tris Gel (12 well)	Invitrogen
NuPAGE™ 4-12% Bis-Tris Gel (17 well)	Invitrogen
NuPAGE™ 4-12% Bis-Tris Midi Gel (26 well)	Invitrogen
Omnifix syringe, 10 ml	B. Braun
Parafilm PM-996	Bemis
PD-10 Desalting Columns	Cytiva
Pipette tips (100 µl – 200 µl)	Nerbe plus
Pipette tips (1250 µl)	Nerbe plus
Pipette tips, super slim (10 µl)	Nerbe plus
Rapid-Flow™ sterile bottle top filter	Nalgene
Reaction tube (2.0 ml)	Sarstedt
Reusable bottle top filter, thread size 45 mm	Nalgene
SafeSeal reaction tube (1.5 ml)	Sarstedt
Serological pipettes, with cotton plug, polystyrene, sterile, non-pyrogenic/endotoxin-free, non-cytotoxic (10 ml)	Sarstedt
Serological pipettes, with cotton plug, polystyrene, sterile, non-pyrogenic/endotoxin-free, non-cytotoxic (50 ml)	Sarstedt
Sterican®, GR. 17, 24 G, Ø 0.55 x 25 mm, disposable hypodermic needle	B. Braun
Union Luerlock female/M6 male	Cytiva
Union Luerlock female/M6 female	Cytiva
Whatman™ Filter Paper Circle, 110 mm	Cytiva

3.1.5 Chromatography media and columns

Table 5: Information on chromatographic techniques, media and columns used in this work

Medium	Chromatography technique	Manufacturer	Column
C18 (2 µm)	RPC	Thermo Fisher Scientific	Acclaim™ PepMap™ RSLC nanoViper (0.075 mm x 500 mm)
C18 (5 µm)	RPC	Thermo Fisher Scientific	Acclaim™ PepMap™ 100 (0.3 mm x 5 mm)
Q Sepharose™ Fast Flow	IEC	Cytiva	XK 16/20 (10 ml)
Source 15RPC	RPC	Cytiva	Resource RPC (3 ml)
Superdex™ 75 Increase	SEC	Cytiva	Tricorn 5/150 Column

3.1.6 Other devices

Table 6: Devices used for laboratory work

Device	Manufacturer
Accu-jet® pro pipette controller	Brand
ÄKTA Pure 150 purification system	Cytiva
ÄKTApurifier purification system	GE Healthcare
Alpha 2-4 LD plus lyophilization system	Martin Christ
Avanti® J-26 XP	Beckman Coulter
BD 53 Incubator	Binder
Block heater SBH130D	Stuart
Centrifuge 5418R	Eppendorf
Concentrator 5301	Eppendorf
Epson Perfection V850 Pro	Epson
GBox Chemi XL 1.4	Syngene
JEM-1400 Plus electron microscope	Jeol
Laboport N86 membrane vacuum pump	KNF
LTQ Orbitrap Elite mass spectrometer	Thermo Fisher Scientific
MSE3.6P-000-DM microbalance	Sartorius
MTS 2/4 digital shaker	IKA
PELCO easiGlow™ glow discharge system	Ted Pella
Pellet pestle, motor cordless	Sigma-Aldrich
Polymax 1040 shaker	Heidolph
PowerPac™ HV High-Voltage Power Supply	Bio-Rad
RET basic, safety control, magnetic stirrer	IKA
SevenEasy pH meter	Mettler Toledo
Sonorex super 10 P	Bandelin
TemCam-F216	TVIPS
U3000 RSLCnano	Thermo Fisher Scientific
Ultrasonic processor GEX500	Sonics
Vortex-mixer VV3	VWR
VZ 14 vacuum chamber	Heraeus

XCell SureLock™ Electrophoresis Mini-Cell	Invitrogen
XCell4 SureLock™ Electrophoresis Midi-Cell	Invitrogen

3.1.7 Software

Table 7: Software employed for analysis, imaging processes, writing and calculations

Software	Distributor
Adobe Illustrator CS6	Adobe Systems Incorporated
Adobe Photoshop CS6	Adobe System Incorporated
EMMENU	TVIPS
GeneSys 1.3.3.0 image acquisition software	Syngene
MASH Explorer	Ying Ge research group
Microsoft Office Professional Plus 2019	Microsoft
mMass 5.5.0	Martin Strohalm
Origin 2021 Pro	Origin Lab Corporation
Peaks AB 2.0	Bioinformatics Solutions Inc.
SilverFast 8.0	LaserSoft Imaging AG
TEM Center	JEOL
Thermo Xcalibur 2.2 SP1.48	Thermo Fisher Scientific
Thermo Xcalibur 3.0.63	Thermo Fisher Scientific
UCSF Chimera 1.15	University of California
Unicorn™ 5.11	General Electric Company
Unicorn™ 6.4.1	General Electric Company

3.2 Methods

3.2.1 Ethics

This study includes human subjects comprising cardiac and abdominal fat biopsies as well as urine and blood serum samples. Sample collection has been approved by the ethics committee in Heidelberg (S-123/2006 – 14 May 2018). The investigation of the tissue in the context of the proposal of the research group FOR2969 regarding amyloid fibrils and amyloid precursor proteins has been approved by the ethics committee in Ulm (203/18 – 30 May 2018). The ethics approval documents are attached to the appendix (App. Fig. 64 – 67).

3.2.2 Patient tissue

Abdominal fat aspirates, heart tissue, serum, and urine were used in this work. Patient samples were collected at Heidelberg University Hospital by Ute Hegenbart. The characterization of the samples was carried out under the supervision of Ute Hegenbart at Heidelberg University Hospital. The characterization included the determination of the amount of amyloid in abdominal fat using a semi-quantitative method [101] and the determination of the amount of albumin and the free LCs in urine. All cases examined in this work are listed below with patient information and sample type (Tab. 8).

Table 8: List of cases processed in this study with tissue information and patient information (sex and age).

Case	Tissue	Sex	Age
P2/16	Abdominal fat	Male	54
P6/16	Abdominal fat	Female	63
P139/17	Urine (U1)	Female	74
P232/15	Abdominal fat	Male	73
P2691501	Urine (U2)	Female	73
FOR005	Heart	Female	50
	Serum		
FOR007	Heart	Female	51
FOR009	Heart	Male	51
FOR010	Heart	Male	64
	Serum		
FOR012	Heart	Male	56
	Serum		
FOR101	Abdominal fat	Male	72
FOR102	Abdominal fat	Female	34
FOR103	Abdominal fat	Female	80
FOR104	Abdominal fat	Female	76
FOR105	Abdominal fat	Female	69
FOR106	Abdominal fat	Male	68
FOR107	Abdominal fat	Male	55
FOR108	Abdominal fat	Female	50
FOR109	Abdominal fat	Female	41
FOR112	Abdominal fat	Male	62
FOR113	Abdominal fat	Male	61
FOR116	Abdominal fat	Male	60
FOR117	Heart	Female	52
FOR120	Abdominal fat	Male	71
FOR122	Abdominal fat	Female	75
FOR124	Abdominal fat	Female	72
FOR127	Abdominal fat	Female	68
FOR128	Abdominal fat	Male	58

FOR130	Heart	Male	57
FOR131	Abdominal fat	Female	80
FOR132	Abdominal fat	Male	57
FOR133	Abdominal fat	Female	67
FOR135	Abdominal fat	Female	38
FOR136	Abdominal fat	Female	59
FOR137	Abdominal fat	Female	49
FOR138	Abdominal fat	Male	70
FOR139	Abdominal fat	Male	63
FOR140	Heart	Female	57
	Serum		
FOR142	Abdominal fat	Male	50
FOR143	Abdominal fat	Female	68
FOR144	Abdominal fat	Male	64
FOR146	Abdominal fat	Female	59
FOR147	Abdominal fat	Male	54
FOR148	Abdominal fat	Male	57
FOR149	Abdominal fat	Female	60
FOR150	Abdominal fat	Male	51
FOR153	Abdominal fat	Male	64
FOR154	Abdominal fat	Female	54
FOR156	Abdominal fat	Female	60
FOR157	Abdominal fat	Male	71
FOR159	Abdominal fat	Female	77
FOR161	Heart	Female	62
FOR162	Abdominal fat	Male	60
FOR164	Abdominal fat	Female	63
FOR165	Abdominal fat	Male	69
FOR169	Abdominal fat	Female	54
FOR171	Abdominal fat	Male	70
FOR176	Abdominal fat	Male	64
FOR179	Abdominal fat	Male	74
FOR182	Abdominal fat	Female	66
FOR183	Abdominal fat	Male	76
FOR185	Abdominal fat	Female	68
FOR186	Abdominal fat	Male	53
FOR189	Abdominal fat	Male	83
FOR1001	Abdominal fat	Female	53
FOR1006	Abdominal fat	Male	80
FOR1007	Heart	Female	53
FOR1010	Abdominal fat	Female	52
FOR1015	Abdominal fat	Female	78
FOR1021	Abdominal fat	Female	65

3.2.3 Extraction of AL fibrils from heart tissue

AL amyloid fibrils were extracted according to a previously established protocol [63], including various washing steps of the homogenized tissue. Washing consisted of resuspending the pelleted tissue in the respective buffer, subsequent centrifugation, and removing the supernatant. Unless otherwise stated, all related centrifugation steps were performed at 3100 xg for 5 min at 4 °C and resuspending the pellet was conducted using a pellet pestle (motor cordless). An initial amount of 250 mg of heart tissue, minced with a scalpel, was washed 5-9 times with 0.5 ml TCB, depending on blood contamination. The resulting pellet was resuspended in 1 ml TCB containing cOmplete EDTA-free protease inhibitor and 5 mg collagenase from *Clostridium histolyticum*. The sample was incubated overnight at 37 °C and 750 rpm shaking using an MTS 2/4 digital shaker. The next day, the sample was centrifuged at 3100 xg for 30 min at 4°C, and the resulting supernatant was removed. The remaining pellet was washed 4-6 times with 0.5 ml TEB. Afterwards, the pellet was washed 7-9 times in 0.5 ml of ice-cold water, and the supernatant containing the fibrils was stored for further use at 4 °C.

3.2.4 Extraction of AL fibrils from adipose tissue

The following fibril extraction approaches include various methods aimed at successfully extracting amyloid fibrils and keeping the fibrils intact. To this end, all centrifugation steps were performed gently at 3100 xg and 4 °C for 5 min unless otherwise stated. Washing of a sample consisted of resuspending the pelleted tissue in the respective buffer, subsequent centrifugation, and removing the supernatant. A sample for LDS-PAGE analysis was stored after each step in the protocols.

3.2.4.1 Fibril extraction with collagenase from C. histolyticum

The protocol using collagenase from *C. histolyticum* to extract fibrils from adipose tissue is derived from the previously described protocol of Annamalai et al. [14]. Abdominal fat tissue was washed 5-9 times with TCB to remove contaminants. The amount of TCB used in each washing step depended on the initial amount of abdominal fat, with 0.2 ml of TCB used per 100 mg of tissue. After each washing step, the supernatant was removed using an Injekt®-F single-use fine dosage syringe with a Sterican® needle (GR. 17) to avoid destroying the fat layer on the top. Per 100 mg of initial abdominal fat tissue, the resulting sample was mixed with 0.4 ml of TCB, containing cOmplete EDTA-free protease inhibitor and 2 mg collagenase from *C. histolyticum*. The mixture was incubated overnight at 37 °C and 750 rpm using an MTS 2/4 digital shaker. The next day, the sample was centrifuged, and the supernatant was removed using an Injekt®-F single-use fine dosage syringe with a Sterican® needle (GR. 17). The solid layer of fat adhering to the side of the tube was also removed using a spatula. The remaining pellet was washed 3-4 times with 0.2 ml TEB per 100 mg starting tissue. For fibril extraction, the resulting pellet was resuspended 5-8 times in 0.1 ml of ice-cold water per 100 mg of initial tissue. The supernatant containing the fibrils was stored at 4 °C.

3.2.4.2 Fibril extraction with chloroform

This extraction approach is based on a previously invented methanol-chloroform-water extraction method [102]. A solution of 0.25 ml chloroform and 0.25 ml TCB was mixed with approximately 40 mg of abdominal fat tissue. After thoroughly inverting the tube and incubating it for 1 min at 4 °C, the chloroform phase was removed. The TCB phase was centrifuged, and the supernatant was discarded. The remaining pellet was washed with 0.5 ml TEB to remove contaminants. Subsequently, fibrils were extracted by resuspending the

remaining tissue three times with 0.5 ml ice-cold water and storing the supernatant at 4 °C. For LDS-PAGE, the chloroform sample was concentrated using concentrator 5301 to evaporate the organic solvent. The remaining lipids were dissolved in 8 M urea, and NuPAGE™ LDS Sample Buffer (4x) was added.

3.2.4.3 Fibril extraction with acetone

This approach is derived from a previously used protocol by Olsen et al. [103], which was combined with the water extraction procedure of Annamalai et al. [14]. 40 mg of abdominal fat tissue was washed with 0.5 ml TCB and then with 0.5 ml TEB. Lipids were removed from the remaining tissue by mixing with 0.5 ml acetone, centrifuging the mixture, and discarding the supernatant. After air-drying the tissue pellet, fibril extraction was performed by resuspending the pellet in 0.5 ml ice-cold water and collecting the supernatant. For LDS-PAGE, the acetone sample was concentrated using concentrator 5301 to evaporate the organic solvent. The remaining lipids were dissolved in 8 M urea, and NuPAGE™ LDS Sample Buffer (4x) was added.

3.2.4.4 Fibril extraction with detergents

Extraction with detergents is based on several previously published extraction procedures [66,104–108]. Buffers containing either 0.04 % TX-114, 24 mM SDC, or 24 mM CHAPS were prepared by mixing the detergents with cOmplete EDTA-free protease inhibitor and either LS or HS extraction buffer. That resulted in one high-salt and one low-salt solution for each detergent. 0.5 ml of each detergent solution was mixed with 40 mg of abdominal fat tissue. Samples were centrifuged at 8,000 xg for 20 min at 4 °C. The pellet was expected to contain

the fibrils and was stored at 4 °C. For LDS-PAGE and negative stain TEM, the pellet was resuspended in water.

3.2.4.5 Fibril extraction using ultracentrifugation (UC) and N-lauryl sarcosine

This fibril extraction approach is based on a protocol described by Goedert et al. 1992 [109]. 100 mg of abdominal fat tissue was dissolved in 1 ml of UB and centrifuged at 16900 xg for 20 min at 4 °C. After collecting the supernatant, the pellet was resuspended in 0.5 ml UB and centrifuged again under the same conditions. Both supernatants were combined, adjusted to 1 % N-lauryl sarcosine, and incubated at 20 °C for 1 h with shaking at 500 rpm. Subsequently, the sample was centrifuged at 100,000 xg for 1 h at 4 °C, and the resulting pellet should contain the fibrils. Since no pellet was visible, the supernatant was removed, and 0.01 ml of NuPAGE™ LDS Sample Buffer (4x) was added to the sample tube. The tube containing the LDS sample buffer was heated to 90 °C to dissolve any remaining proteins, which were then analyzed with the supernatant by LDS-PAGE.

3.2.4.6 Mechanical extraction of amyloid fibrils using pestle and mortar

This extraction approach is based on a method described by Lavatelli et al. in 2020 [73], in which amyloid fibrils were extracted from cardiac tissue using a mortar and pestle. Therefore, it should be verified whether that protocol is more efficient than using collagenase to extract amyloid fibrils from cardiac tissue before applying it to adipose tissue. 250 mg of heart tissue was diced with a scalpel and washed with 1 ml of DPBS. The tissue was then homogenized with a pestle and mortar and resuspended in 0.5 ml TEB containing cOmplete protease inhibitor. After centrifugation and removal of the supernatant, homogenization was repeated nine times.

The resulting sample was resuspended thrice with 0.5 ml ice-cold water, and the supernatant, which should contain the fibrils, was stored at 4 °C. The supernatant was stored at 4 °C.

3.2.4.7 Mechanical extraction of amyloid fibrils using a self-made fractionator

The following extraction approach is based on a previous study using a fractionator to dissociate fat aspirates mechanically [110]. The fractionator of that study consisted of a Luer-to-Luer transfer with three holes in the center [110]. 400 mg of abdominal fat was washed five times with 0.8 ml TCB. Then the resulting tissue was mixed with 1.6 ml TCB containing EDTA-free protease inhibitor and added to a homemade fractionator. In this work, the fractionator consisted of two Omnifix syringes (10 ml) coupled with a Union Luerlock female/M6 male and a Union Luerlock female/M6 female. Dissociation of the adipose tissue was achieved by continuously pushing the sample from one side of the fractionator to the other for 1 min and 50 s, followed by incubation for 15 min at 20 °C with shaking at 400 rpm using an MTS 2/4 digital shaker. Dissociation and incubation were repeated 4 times in total, and the resulting homogenate was centrifuged for 30 min. The resulting sample contained a top fat layer, which was removed with the supernatant. The pellet was washed thrice with 0.15 ml TEB. Fibril extraction was performed by resuspending the remaining tissue in 0.15 ml ice-cold water six times and storing the supernatant at 4 °C. The remaining lipids were dissolved in 8 M urea, and NuPAGE™ LDS Sample Buffer (4x) was added.

3.2.5 Isolation of LCs from urine

To isolate LCs from urine, other protein species, primarily albumin, had to be removed. Therefore, LCs were precipitated from urine by ammonium sulfate precipitation and subsequently purified by IEC.

3.2.5.1 Ammonium sulfate precipitation of LCs

Ammonium sulfate precipitation was performed at 4°C and in two steps bringing the urine sample first to 25 % (v/v) ammonium sulfate and then to 80 % (v/v). After preparing a saturated ammonium sulfate solution, 1.666 ml of this solution was added dropwise to 5 ml of urine with constant stirring to reach a concentration of 25 % (v/v). The mixture was incubated for 6 h at 4 °C and then centrifuged at 10,000 xg for 30 min at 4 °C to remove precipitated proteins. An additional 18.315 ml of the saturated ammonium sulfate solution was added to the supernatant to reach a final concentration of 80 % (v/v), and the resulting mixture was incubated overnight at 4 °C with constant stirring. The next day, the sample was centrifuged at 5,000 xg for 25 min at 4 °C to collect the precipitated proteins as a pellet containing the free LCs. The pellet was stored at -80 °C for further use.

The amount of saturated ammonium sulfate solution, required to achieve a given concentration of ammonium sulfate in the urine sample, was calculated according to the following formula:

$$V[ml] = X\ [ml] \times \frac{(S2 - S1)}{(1 - S2)}$$

V [ml] = volume of saturated ammonium sulfate solution needed to obtain the desired ammonium sulfate concentration of the sample

S1 = initial ammonium sulfate concentration of the sample before the addition of the saturated ammonium sulfate solution

S2 = desired ammonium sulfate concentration of the sample after the addition of the saturated ammonium sulfate solution

X = initial volume of the sample before the addition of the saturated ammonium sulfate solution

3.2.5.2 Purification of LCs from urine by IEC

The stored pellet of the 80 % (v/v) ammonium sulfate precipitation was resuspended in 1 ml of IEC running buffer and loaded onto the IEC column. The purification was conducted with an ÄKTA pure 150 operated with IEC running buffer and IEC elution buffer, and the column was equilibrated with IEC running buffer. Proteins were separated by a linear gradient from 0 % to 100 % (v/v) IEC elution buffer over 20 column volumes (CV). The concentration stayed at 100 % (v/v) IEC elution buffer for 4 CV and was reduced to 0 % (v/v) over 2 CV. The eluted proteins were fractionated over the entire gradient, and the LC-containing fractions were identified by LDS-PAGE analysis.

3.2.5.3 Analysis of the purified LCs by SEC

The purified LCs from IEC were analyzed for their oligomerization state by SEC. Therefore, the LCs were concentrated 10-fold using an Amicon® Ultra-15 Centrifugal Filter Device (3k). 0.05 ml was applied to the SEC column. The analysis was conducted using an ÄKTA purifier operated with the SEC buffer. The column was also equilibrated with the SEC buffer. Separation of the proteins was monitored over 2 CV. LC-containing fractions were identified by LDS-PAGE analysis. The molecular mass of the eluted proteins was calculated using a calibration curve of a previous marker run.

3.2.6 Analysis of blood serum for AL proteins

Heart biopsies and blood serum from the same patient were mandatory to search for AL proteins in blood serum. In the first step, amyloid fibrils were extracted from heart tissue and disrupted by 6 M Gnd-HCl. An FPLC purification procedure was established for the fibril-forming AL proteins, and purification was validated by MS. In a second step, the FPLC columns were cleaned thoroughly, and a sample of 0.5 ml IEC running buffer was subjected to the same purification protocol established for the AL protein from heart-extracted fibrils. The same fractions that previously contained AL protein were collected blindly to check whether AL proteins from the first step remained on the column. Finally, 0.5 ml of blood serum from the same patient was applied to the same purification procedure, and the fractions that previously contained AL protein were again collected blindly.

3.2.6.1 Purification of AL protein extracted from heart tissue

Water extracts of fibrils from heart tissue were adjusted to 0.05 M ethanolamine (pH = 10) and 6 M Gnd-HCl and incubated overnight at 4 °C. The next day, the Gnd-HCl was removed using a PD-10 desalting column. An eluate sample was taken for later LDS-PAGE analysis, and the remaining eluate was subjected to IEC. Purification was conducted with an ÄKTA pure 150, operated with IEC running buffer and IEC elution buffer. The IEC column was equilibrated with IEC running buffer. The proteins were separated by a linear gradient from 0 % to 100 % (v/v) IEC elution buffer over 20 CV. The concentration stayed at 100 % (v/v) IEC elution buffer for 4 CV and was reduced to 0 % (v/v) over 2 CV. The eluted proteins were fractioned over the entire gradient, and the AL protein-containing fractions were identified by LDS-PAGE analysis. After pooling these fractions, the sample was adjusted to 0.2 % (v/v) TFA and subjected to RPC. For this purpose, the ÄKTA pure 150 was operated with RPC running buffer

and RPC elution buffer, and the column was equilibrated with RPC running buffer. Proteins were separated by a linear gradient from 0 % to 58 % (v/v) RPC elution buffer over 20 CV, followed by a linear gradient from 58 % to 100 % (v/v) RPC elution buffer over 4 CV. The concentration stayed at 100 % (v/v) RPC elution buffer for 4 CV and was reduced back to 0 % (v/v) over 3 CV. The eluted proteins were fractionated over the entire gradient, and the AL protein-containing fractions were identified by LDS-PAGE analysis. These fractions were pooled, lyophilized, and stored for further use. When purification resulted in low AL protein concentrations (estimated from low signal intensities at 215 nm in the chromatograms), samples were concentrated 10-fold for LDS-PAGE analysis using an Amicon® Ultra-0.5 Centrifugal Filter Device (3k).

3.2.6.2 Applying the purification procedure of AL proteins from heart biopsies to serum

After thoroughly cleaning the IEC and RPC columns, a sample of 0.5 ml IEC running buffer (serum/AL protein) was subjected to the same purification protocol. The same fractions that previously contained AL protein were collected blindly and analyzed by MS. That step should verify that no AL proteins from previous purifications remained on the IEC or RPC column. Finally, the same purification procedure was applied to 0.5 ml of blood serum from the same patient, and the fractions that previously contained AL protein were collected blindly. The final sample of each RPC purification was then lyophilized and subjected to MS. Identical mass peaks in the heart and serum MS spectra indicated AL protein in serum. When purification resulted in low AL protein concentrations (estimated from low signal intensities at 215 nm in the chromatograms), samples were concentrated 10-fold for LDS-PAGE analysis using an Amicon® Ultra-0.5 Centrifugal Filter Device (3k).

3.2.7 Liquid chromatography

Protein purification was performed using either the ÄKTApurifier for SEC or the ÄKTA Pure 150 for IEC and RPC. Both purification systems included two pumps (A, B), an injection valve, a wavelength detector, a flow restrictor with a pressure sensor, and a fraction collector. A sample pump was unique to the ÄKTA pure 150, allowing sample applications without using sample loops or super loops. Since LCs and AL proteins contain very few aromatic amino acids, which are used for the photometric detection of proteins at 280 nm, these proteins could only be detected based on the absorption of their peptide bonds at 215 nm. Therefore, the wavelength detector recorded absorption at 280 nm and 215 nm. The ÄKTApurifier was operated with the Unicorn™ 5.11 software and the ÄKTA Pure 150 with Unicorn™ 6.4.1. All buffers used were previously filtered with an MF-Millipore™ membrane (47 mm, 0.22 µm, MCE membrane) using a Laboport N86 membrane vacuum pump and reusable bottle top filters. RPC buffers were degassed for 10 min in an ultrasonic bath.

3.2.7.1 Ion exchange chromatography

Proteins were purified using an anion exchange chromatography column with a Q Sepharose™ Fast Flow medium (10 ml). Before use, the column was equilibrated with 5 CV of IEC running buffer. Elution was performed by displacing the bound proteins with high concentrations of NaCl (1 M) in the elution buffer. To ensure a strong interaction of protein and medium, the pH values of the buffers used always differed by at least one pH unit from the pI of the protein of interest. The column was cleaned with 5 CV 1 M NaOH, 5 CV 70 % (v/v) ethanol, 5 CV 1 M acetic acid in combination with 0.5 % (v/v) T20, and 2 CV 20 % ethanol after each use.

3.2.7.2 Reversed-phase chromatography

A Resource RPC column with Source 15RPC medium (3 ml) was used to purify LCs or AL proteins. The running buffer and elution buffer contained TFA, which served as an ion pair reagent and increased the retention time of the protein of interest. Before use, the column was equilibrated with 4 CV of RPC running buffer followed by a linear gradient from 0 % to 100 % (v/v) RPC elution buffer over 4 CV, a reduction from 100 % to 0 % (v/v) over 4 CV and another 3 CV of RPC running buffer. The column was cleaned with 5 CV 1 M NaOH, 5 CV 90 % (v/v) acetic acid, 5 CV RPC running buffer, 5 CV RPC elution buffer, and 3 CV 20 % (v/v) ethanol after each use.

3.2.7.3 Size exclusion chromatography

Purifications were performed using a Tricorn 5/150 Column in combination with Superdex™ 75 Increase medium (3 ml). Before use, the column was equilibrated with 6 CVs of SEC buffer. The void volume (V_0), as well as elution volumes (V_e) of proteins with known molecular masses, were determined according to the instruction of the manufacturer Cytiva. A freshly filtered solution of Blue dextran 2000 (1 mg/ml) in SEC buffer was prepared, and 0.05 ml was applied to the column. V_0 was equal to the elution volume of Blue dextran 2000. A mixture of 1.5 mg/ml conalbumin (75 kDa), 1.5 mg/ml carbonic anhydrase (29 kDa), 3 mg/ml ribonuclease A (13.7 kDa), and 1 mg/ml aprotinin (6.5 kDa) in SEC buffer was prepared, and 0.01 ml was applied to the column. V_e of the proteins was determined by measuring from the point of injection to the center of a corresponding elution peak. Using V_0, V_e, and the geometric column volume V_c (which is 3 ml for a Superdex 75 increase 5/150 GL column), the partition coefficient (K_{av}) could be calculated using the formula:

$$K_{av} = \frac{V_e - V_0}{V_c - V_0}$$

Plotting K_{av} on the X-axis and the logarithmic molecular weight on the Y-axis yielded a calibration line used to determine the unknown molecular masses of eluted proteins. The column was cleaned with 6 CV 1M NaOH, 6 CV 1M acetic acid, 6 CV 6M Gnd-HCl, 4 CV 1 % (w/v) SDS, 4 CV 30 % (v/v) acetonitrile, and 4 CV 70 % (v/v) ethanol after each use.

3.2.8 LDS-PAGE

Unless otherwise stated, LDS-PAGE was conducted under reduced conditions. Samples were mixed with NuPAGE™ LDS Sample Buffer (4x) as well as NuPAGE™ Sample Reducing Agent (10x) and incubated at 95 °C for 10 min on an SBH130D block heater. Meanwhile, a NuPAGE™ 4-12% Bis-Tris Gel was prepared by placing it in an XCell SureLock™ Electrophoresis Mini-Cell or Midi-Cell and filling the reservoirs with NuPAGE™ MES SDS Running Buffer (1x). After adding 0.5 ml of NuPAGE™ Antioxidant (10x) to the reservoir facing the wells of the gel, samples were loaded on the gel, and a BlueEasy Prestained Protein Marker was used as a reference. The gel was run at 180 V for 35 min using a PowerPac™ HV high-voltage power supply, followed by staining with an LDS-PAGE staining solution for 30 min and then incubating in LDS-PAGE destaining solution until protein bands were visible. For the first half of this work, LDS-PAGE images were acquired using a GBox Chemi XL 1.4 and the GeneSys 1.3.3.0 software. In the second half, an Epson Perfection V850 Pro scanner with SilverFast 8.0 software was purchased, acquiring color images. For non-denatured LDS-PAGE, no NuPAGE™ Sample Reducing Agent (10x) and NuPAGE™ Antioxidant (10x) were added.

3.2.9 Negative stain TEM

Before sample application, a Formvar/Carbon Coated Copper Grid (200 mesh) was glow discharged with the dark side up using a PELCO easiGlow™ glow discharge system. Then, 3 µl of the sample was incubated on the dark side of the grid for 1 min and removed using a Whatman™ filter paper circle (110 mm) held to the edge of the grid. The grid was washed with water thrice to remove buffer substances and other contaminants. Then it was stained by washing with 2 % uranyl acetate thrice. For this purpose, three droplets of water (0.05 ml) and three droplets of 2 % uranyl acetate (50 µl) were placed on a piece of Parafilm PM-996, and the dark side of the grid was immersed in it. After each washing step, the grid was dried using a Whatman™ filter paper circle (110 mm) held to the edge of it. Before storing the grid for electron microscopic examination, it was dried for an additional 10 min at room temperature. During the complete grid preparation, it was fixed in a Dumont tweezer (style N5, Biological Grade, Dumoxel®). Negative stain grids were observed in a JEM-1400 Plus electron microscope operated at 120 kV using the EMMENU software. Images were acquired at a 20,000-fold magnification with a TemCam-F216 using the TEM Center software.

3.2.10 Mass spectrometric analyses

The operation of the mass spectrometer and the recording of all mass spectra was performed by employees of the Core Unit Mass Spectrometry and Proteomics (Medical Faculty, Ulm University) under the direction and supervision of Sebastian Wiese. Mass spectrometric measurements were performed using a U3000 RSLCnano HPLC system online coupled to an LTQ Orbitrap Elite mass spectrometer, consisting of two mass analyzers, namely a quadrupole and an Orbitrap. Ionization was performed by ESI using a nanoelectrospray ion source with distal-coated SilicaTips. The mass spectrometer was operated with Thermo Xcalibur 2.2 SP1.48

software, a spray voltage of 1.5 kV, a capillary temperature of 250 °C, and an S-Lens radio frequency level of 68.9 %. Data were acquired in the Orbitrap at a resolution of 30,000 (at m/z 400) with automatic gain control enabled (10^6 ions; maximum fill time: 500 ms), and the resulting m/z scans ranged from 370 to 1700. Calibration of the mass spectrometer was performed using Pierce™ LTQ ESI Positive Ion Calibration Solution.

MS was used to collect two types of information: the total mass of AL proteins and their amino acid sequence. Both analyses shared the same initial workup of the fibril samples. Fibril-containing water extracts from abdominal fat or heart tissue were lyophilized, and the amount of lyophilized protein was determined using an MSE3.6P-000-DM microbalance. The lyophilized sample was sent to the Core Unit Mass Spectrometry and Proteomics at Ulm University and stored at -20 °C until further processing.

3.2.10.1 Total mass analysis

The following sample preparation and data acquisition were conducted by the technical assistant of the Core Unit Mass Spectrometry and Proteomics. The lyophilized sample was dissolved in water to reach a concentration of 1 µg/µl of protein. 2 µl of the dissolved lyophilized sample was mixed with 4 µl of MS dissociation buffer and incubated overnight at 4 °C. The next day, the sample was diluted to 15 µl with 0.1 % TFA and applied to the U3000 RSLCnano HPLC system, which was operated with HPLC buffer A and HPLC buffer B. The used columns were equilibrated with 5 % (v/v) HPLC buffer B. Salts and other polar contaminants were first removed by injecting the sample on an Acclaim™ PepMap™ 100 (0.3 mm x 5 mm) pre-column and washing with HPLC buffer A for 5 min at a flow rate of 30 µl/min into waste. Proteins were eluted directly onto the analytical column by changing the flow path from waste to the Acclaim™ PepMap™ RSLC nanoViper (0.075 mm x 500 mm)

analytical column and applying a linear gradient of 5 % to 40 % (v/v) HPLC buffer B at a flow rate of 250 nl/min over 30 min. That gradient was also applied to the analytical column, and the eluting fractions were injected directly into the electrospray module of the mass spectrometer. Mass spectra were recorded and sent to Julian Baur for the following evaluations.

The recorded mass spectra were summed using Thermo Xcalibur 3.0.63. The summed spectrum was deconvoluted using the "Quick deconvolution" function of MASH Explorer (default settings) [111]. Monoisotopic masses of deconvolutions with a confidence score of 90 % or higher corresponding to proteins with 5 or more charges were used for the subsequent sequence assignment.

3.2.10.2 Determination of the amino acid sequence of AL proteins by MS-based protein *de novo* sequencing

The following sample preparation and data acquisition were conducted by the technical assistant of the Core Unit Mass Spectrometry and Proteomics. The lyophilized sample was dissolved in water to reach a concentration of 1 µg/µl of protein. After removing 2 µl for total mass analysis, the remaining dissolved protein was divided into five wells of a polyacrylamide gel. After LDS-PAGE, the desired protein bands were cut out and washed with 0.1 ml of the respective protease buffer and a 1:1 mixture of protease buffer and acetonitrile for 10 min each. Washing was repeated two more times. The gel pieces were vacuum dried using a VZ 14 vacuum chamber, then reduced in DTT buffer for 20 min at room temperature and alkylated in alkylation buffer for 20 min at 37 °C. Proteins in the gel were digested by incubating each gel slice in 0.01 ml of a protease buffer containing 0.33 ng of the respective protease overnight at 37 °C (chymotrypsin was incubated at 25 °C). Subsequently, the gel slices were transferred into 0.01 ml of acetonitrile and 0.01 ml of 0.1 % TFA (v/v). The peptides were extracted by incubating them in an ultrasonic bath for 10 min. The gel slices were removed,

acetonitrile was evaporated, and the peptide solution was filled to 15 µl with 0.1 % TFA (v/v). The peptides were purified by HPLC and eluted directly into the ESI module as described for the total mass analysis (see 3.2.10.1 for details). Peptide fragmentation was performed either by CID or HCD. For CID, the 20 most intense ions were selected from the survey scan, and the automated gain control was set to 10,000 ions with a maximum fill time of 10 ms. FT preview scan was enabled, the collision energy was set to 35 %, and the activation time was set to 10 s with an activation Q of 0.25. No fragmentation was performed with singly charged ions, and m/z of fragmented ions was excluded for 60 s. The peptide fragments were detected in the quadrupole at normal scan speed. For HCD fragmentation, ten ions were selected per survey scan, rejecting singly charged ions and excluding m/z values of fragmented ions for 60 s for additional selection. In addition, the automatic gain control was set to 50,000 ions and a maximum fill time of 100 ms. Peptide fragments resulting from 100 ms HCD activation were detected in the Orbitrap at a resolution of 15,000. Mass spectra were recorded and sent to Julian Baur for the following evaluations. *De novo* sequencing of the selected peptides was performed using PEAKS AB 2.0 software [112], which operated with a mass accuracy of 15 ppm on intact peptide masses and either 0.05 Da or 0.5 Da for fragmented ions to obtain a complete peptide coverage. Sequence determination settings remained the default. In order to determine the entire LC sequence, Peaks AB uses an internal algorithm that identifies the correct order of *de novo* sequenced peptides based on conserved regions of LCs. Amino acids with a confidence level of 85 % and above were considered correct. Carbamidomethylation of Cys resulting from alkylation was set as a fixed PTM. Asparagine (Asn) deamidation, Gln deamidation, and methionine (Met) oxidation were set as variable PTMs. Non-distinguishable amino acids, such as isoleucine (Ile) and Leu, as well as Gln and lysine (Lys), were assigned according to the GL sequence or the respective cDNA sequence, if available.

3.2.11 Primary structure determination of the AL proteins

The AL protein's primary structure was identified using MS and the cDNA-derived sequences provided by the Medical Department V of the Heidelberg University Hospital. Assigning the monoisotopic mass of the AL proteins to a part of the corresponding provided cDNA-derived sequence identified the primary structure of the AL proteins. In cases without or with incomplete cDNA sequence, the AL protein sequence was determined by MS-based protein sequencing using Peaks AB. The parts of the precursor LC not covered by the AL proteins were considered GL. The primary structure was determined by assigning the monoisotopic mass of the AL proteins to the MS-based protein sequence. Deconvoluted monoisotopic masses of the AL proteins were assigned using the mMass software [113]. The assignment was considered correct if the sequence matched the monoisotopic mass with an error of less than 0.1 Da. Disulfide bonds, Met oxidation, and a pyroglutamate at the N-terminus were considered if necessary.

3.2.12 Computational analysis

3.2.12.1 GL assignment

For the GL assignment, the cDNA was translated into the corresponding amino acid sequence using the Expasy Translation tool [114]. V and C gene segments from cDNA and MS-derived sequences were identified using BLAST/BLAT search on UniProtKB/Swiss-Prot database [115]. The GL gene segment with the highest similarity to the cDNA or MS-derived sequence was considered correct. J gene segments were identified by directly comparing MS and cDNA-derived sequences with sequences from the IMGT/GENE database [116]. The amino acid sequence of the identified GL segments was taken from the IMGT/GENE database [116]. According to the V-J-C combination of the patient-derived LCs, GL gene segments were

combined into gapless GL sequences, and differences between GL and LC sequences were considered mutations. If the amino acid sequence did not provide enough information to assign a GL Segment, the cDNA sequence of the LCs was compared to the DNA sequence of the GL segments provided by the IMGT/GENE database [116]. The CDR position was assigned according to the IMGT/GENE database [116].

3.2.12.2 Computational analysis of the protein properties

The pI, solubility, Grand Average of Hydropathy (GRAVY), and aggregation score were calculated for biochemical analyses. The GRAVY is defined as the sum of hydropathy values of all amino acids divided by the total number of amino acids[117]. Hydropathy values of Kyte and Doolittle were used [118]. The pI and GRAVY of the proteins were calculated using the ProtParam tool from Expasy [117]. The total aggregation score reflects the aggregation propensity of a respective protein. It was calculated using TANGO [119] by dividing its AGG value by the sum of amino acids of the corresponding sequence. The solubility of proteins, represented by the CamSol solubility score, was calculated using the CamSol web server [120]. Subtracting a value of the GL sequence from the corresponding value of the LC sequence resulted in ΔGRAVY, ΔpI, ΔAGG, and Δsolubility values. N-glycosylation was predicted using the NetNGlyc 1.0 software [121] with default settings. Only glycosylation sites with the motif Asn-X-Serine (Ser) or Asn-X-Threonine (Thr) (X ≠ Proline (Pro)) were considered. Searching for proteases that cleave within the reported truncation region of the investigated AL proteins was conducted using the PeptideCutter tool from Expasy with default settings [117].

3.2.13 Three-dimensional visualization

Three-dimensional visualization of LC structures was performed with Chimera 1.15 (https://www.rbvi.ucsf.edu/chimera/) [122].

3.2.14 Statistics

Statistical significance was calculated with Excel using a two-tailed t-test. The effect size of the t-test was determined as Cohen's d (d_{Cohen}) using sample sizes (n_{LC}, n_{GL}), mean values (M_{LC}, M_{GL}), and standard deviations (SD_{LC}, SD_{GL}) from LC and GL.

$$d_{Cohen} = \frac{|M_{LC} - M_{GL}|}{\sqrt{\frac{(n_{LC} - 1) \times {SD_{LC}}^2 + (n_{GL} - 1) \times {SD_{GL}}^2}{n_{LC} + n_{GL} - 2}}}$$

4. Results

4.1 Characterization of LCs and their derivatives from serum and urine

One of the most important questions in the field of AL amyloidosis is what makes a LC amyloidogenic. It is suggested that truncation of precursor LCs could lead to destabilization, aggregation, and amyloid fibril formation [43,51,123,124]. So far, however, it is unclear whether LCs are fragmented before or after fibril formation. If fragmentation occurs afterwards, it excludes any influence on fibril formation. Therefore, this work investigated the possible existence of AL proteins in body fluids and searched for other evidence on the timing of LC proteolysis.

4.1.1 Purification of FLCs from urine and investigating their oligomeric state

In AL amyloidosis, patients excrete FLCs in the urine (called BJPs), which are used in this work to investigate the timing of LC proteolysis. However, separating the BJPs from albumin is difficult, especially in patients with albuminuria [125]. Therefore, this study presents a purification method using ammonium sulfate precipitation and IEC to separate BJPs from albumin. The purified BJPs were analyzed by SEC to determine their oligomerization state, as previous studies suggested that a disrupted dimer interface of amyloidogenic LCs reduces their stability and makes them susceptible to endoproteolysis [45,51]. Thus, finding BJPs in the urine of AL amyloidosis patients that do not form dimers would provide evidence of a disrupted dimer interface and suggest that LC truncation precedes fibril formation.

Samples of patients with AL amyloidosis are very rare. Consequently, the following work relies only on two cases, one containing FLCs of the λ isotype (P139/17; referred to as U1) and one containing FLCs of the κ isotype (P2691501; referred to as U2). Urine samples were provided

by Ute Hegenbart (University Hospital Heidelberg) in the framework of the research unit FOR2969. U1 had little albumin in the urine (383 mg/24 h), whereas the urine of U2 contained much albumin (4809 mg/24 h). The first step to remove the major portion of albumin from the urine was ammonium sulfate precipitation (Fig. 10).

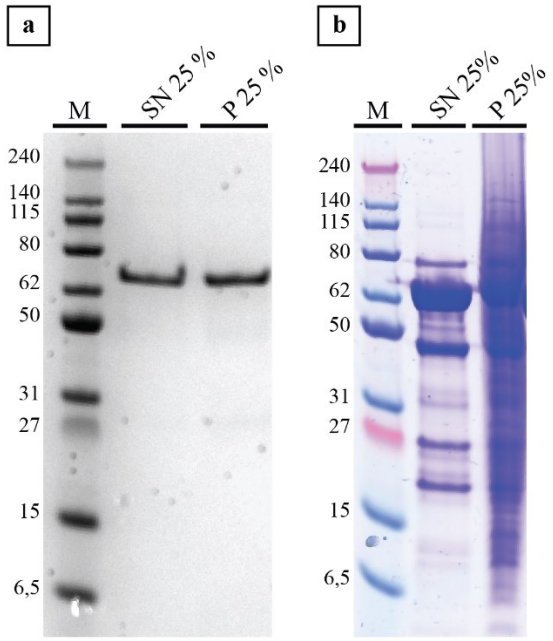


Figure 10: Ammonium sulfate precipitation of U1 and U2. a) Ammonium sulfate precipitation of U1. M: marker in kDa. SN 25%: supernatant after the precipitation with 25 % ammonium sulfate. P 25%: pellet of the precipitated proteins from 25 % ammonium sulfate precipitation. b) Ammonium sulfate precipitation of U2. Some gel columns were removed for convenience only without changing the protein bands.

Urine proteins were first precipitated with 25 % ammonium sulfate and then pelleted by centrifugation. According to LDS-PAGE, the protein pellet (P 25%) of U1 contained mainly albumin, visible as a prominent band at approximately 65 kDa (Fig. 10 a) [126]. The supernatant (SN 25 %) of U1 also contained albumin (Fig.10 a). However, no LCs were visible at an expected molecular weight of approximately 27 kDa (Fig. 10 a), indicating a low concentration of BJPs in the urine of U1. The protein pellet (P 25 %) of U2 contained an abundance of unidentifiable proteins (Fig. 10 b). The supernatant (SN 25 %) of U2 contained LCs visible as a distinct band at 27 kDa (Fig. 10 b) but also albumin and other contaminants with molecular weights of 20 to 80 kDa (Fig. 10 b). In comparison to the supernatant of U1, the supernatant of

U2 contained more albumin, reflecting the higher albumin load in the urine of patient U2. In both cases, not all albumin and other contaminants were removed by precipitation with 25 % ammonium sulfate (Fig. 10). Consequently, the remaining urinary proteins from both supernatants were precipitated by increasing the ammonium sulfate concentration to 80 %. The proteins were pelleted and supplied to IEC for further purification of the LCs (Fig. 11).

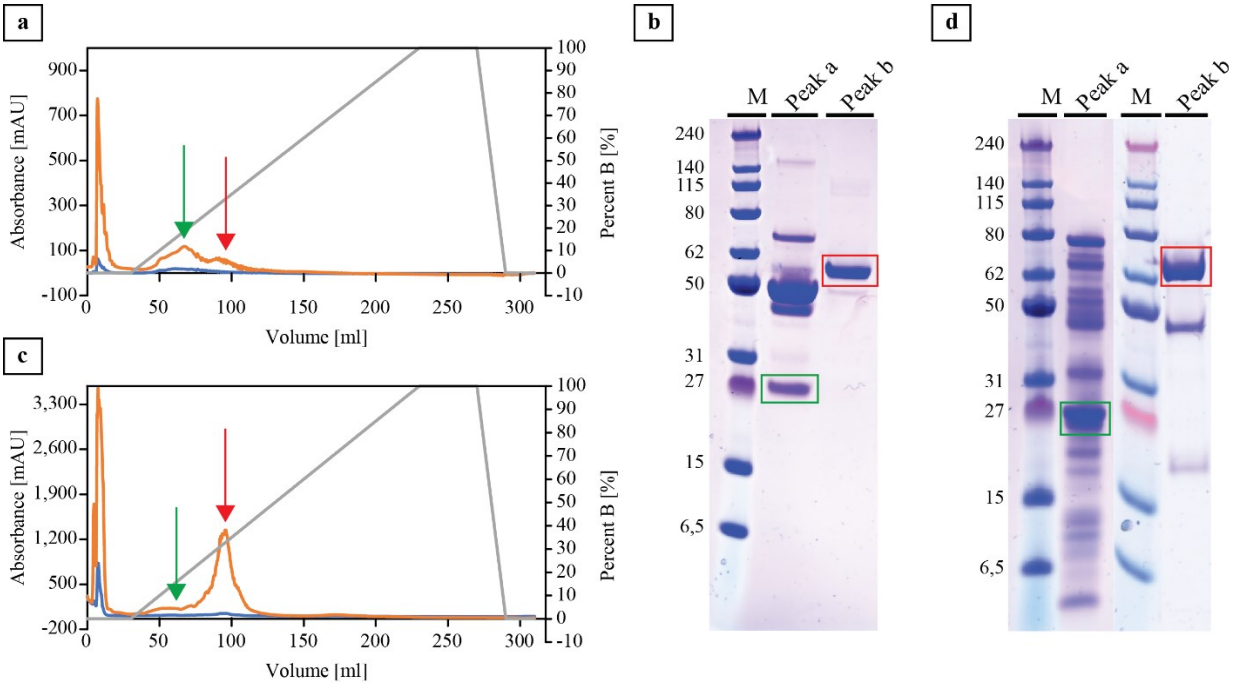


Figure 11: IEC purification of precipitated proteins from 80 % ammonium sulfate precipitation of U1 and U2. a) IEC chromatogram of U1. Primary Y-axis: milli absorbance units (mAU) (blue curve: 280 nm, orange curve: 215 nm). Secondary Y-axis: percentage (%) of elution buffer (gray curve). X-axis: elution volume. Green arrow: Peak a. Red arrow: Peak b. b) LDS-PAGE of the marked peaks from IEC of U1 (a). Green rectangle: protein band of LCs. Red rectangle: Protein band of albumin. M: marker in kDa. c) IEC chromatogram of U2. d) LDS-PAGE of the marked peaks from IEC of U2 (c). The image consists of two combined gels. Some gel columns were removed for convenience only without changing the protein bands.

The proteins from U1 eluted in two peaks (Fig. 11 a, green and red arrows). The first protein fraction eluted between 7.4 % (v/v) and 22.4 % (v/v) of IEC elution buffer (peak a), and the second protein fraction eluted between 25 % (v/v) and 37.5 % (v/v) of IEC elution buffer (peak b) (Fig. 11 a). According to LDS-PAGE, “peak a” contained LCs (Fig. 11 b, green rectangle) and four contaminants, which do not have the molecular weight of albumin. “Peak b” contained mainly albumin (Fig. 11 b, red rectangle). Proteins from U2 eluted in the same two fractions as

described for sample U1 (peak a: 7.4 % (v/v) – 22.4 % (v/v) elution buffer, peak b: 25 % (v/v) – 37.5 % (v/v) elution buffer) (Fig. 11 c). LDS-PAGE confirmed LCs in "peak a" along with several other proteins not corresponding to albumin. (Fig. 11 d, green rectangle). Albumin and two other proteins were present in the second peak (Fig. 11 d, red rectangle). Consequently, LCs were successfully separated from albumin by ammonium sulfate precipitation and IEC.

SEC was performed to analyze the oligomerization state of the purified LCs. For this purpose, the LCs were concentrated and injected onto a Superdex™ 75 Increase 5/150 GL column (Fig. 12).

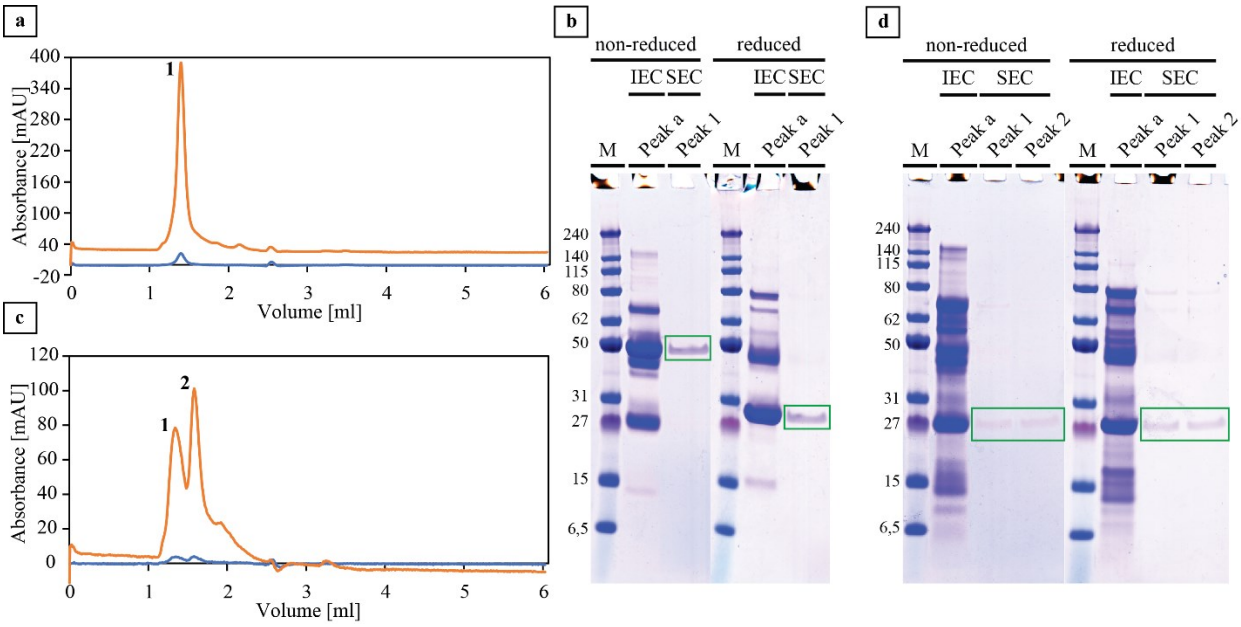


Figure 12: SEC of purified LCs from cases U1 and U2. a) SEC chromatogram of U1. Y-axis: milli absorbance units (mAU) (blue curve: 280 nm, orange curve: 215 nm). X-axis: elution volume in ml. Number: labeling the peak. b) LDS-PAGE of IEC "Peak a" and the SEC peak of U1 under reduced and non-reduced conditions. Green rectangles: LC monomers and dimers. M: marker in kDa. c) SEC chromatogram of U2. d) LDS-PAGE of IEC "Peak a" and the SEC peaks of U2 under reduced and non-reduced conditions.

The purified LCs from case U1 eluted in one peak at 1.339 ml corresponding to a molecular weight of 61 kDa (Fig. 12 a, App. Fig. 68), which indicates the presence of LC dimers. Non-reduced LDS-PAGE of "peak a" from IEC and peak 1 from SEC showed a LC band at 50 kDa (Fig. 12 b, green rectangles). The band shifted to approximately 27 kDa in reduced LDS-PAGE,

indicating the presence of covalently linked dimers with a disulfide bond (Fig. 12 b, green rectangles). The purified LCs from case U2 eluted in two peaks at molecular weights of 61 kDa and 31 kDa (Fig. 12 c, App. Fig. 68), indicating the presence of LC dimers and monomers. However, when peaks 1 and 2 were analyzed by non-reduced LDS-PAGE, only LC monomers with a molecular weight of approximately 27 kDa were detected, even in the dimeric fraction from SEC (peak 1) (Fig. 12 d, green rectangle). Performing LDS-PAGE under reduced conditions did not influence the migration behavior of the LCs (Fig. 12 d, green rectangles), suggesting the presence of non-covalent LC dimers, which are disrupted by heating and LDS.

In summary, this work presents a purification method to separate FLCs from albumin, even when using urine with a high albumin load. SEC and LDS-PAGE indicated different LC dimers in cases with low albumin load (covalent) and high albumin load (non-covalent). SEC also confirmed the presence of monomeric LCs in urine with a high albumin load, suggesting a disrupted dimer interface that destabilizes the LCs and fosters proteolysis.

4.1.2 Analysis of LC truncations in serum

Investigating the oligomeric state of FLCs from urine provided initial suggestions on the timing of LC proteolysis in AL amyloidosis. For further evidence, serum from AL amyloidosis patients was investigated next, considering the following aspects. AL amyloid fibrils consist of LC fragments called AL proteins [14,59,60,63,79,127]. Thus, if the AL proteins from tissue deposits also occur in serum, that would indicate proteolysis before fibril formation. However, investigating the existence of AL protein in the serum required detecting a protein for which no antibody is available. Thus, the only option for identifying AL proteins was to use MS, which required the removal of contaminants before the measurement. For this reason, a purification method for AL proteins from fibril deposits of the heart was established, which was then blindly

applied to serum samples from the same patient. The purified proteins were investigated by MS. Comparing the spectra of the purified fraction from the heart and serum should show whether AL protein from the deposits is also present in the serum.

This work includes four cases where heart tissue and serum were available (FOR005, FOR010, FOR012, FOR140). The patient samples were provided by the University Hospital Heidelberg in the framework of the research unit FOR2969. Results are shown as an example for FOR012. At first, fibrils were extracted from heart tissue according to the standard water extraction protocol [63] (Fig. 13 a). For this purpose, the heart tissue was cut into small pieces and washed with TCB to remove blood components. The material was then incubated with collagenase to destroy the connective tissue and expose the deposited fibrils. Subsequent washing with TEB buffer removed serum amyloid P (SAP), which is tightly bound to the amyloid fibrils [128]. Washing with distilled water brought the fibrils to the solution phase and thus separated them from the cardiac tissue.

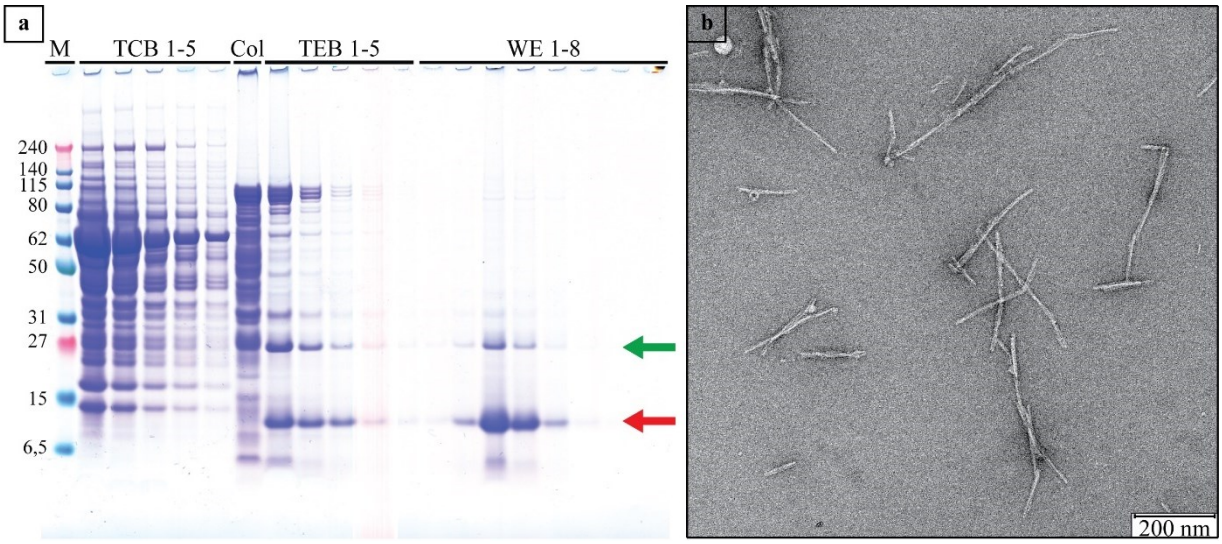


Figure 13: Fibril extraction from heart tissue of FOR012 using the standard water extraction protocol* [63].** ***a) LDS-PAGE of the water extraction procedure. M: marker in kDa. Col: supernatant after collagenase incubation. WE: water extracts. Red arrow: AL protein band. Green arrow: SAP band b) Negative stain TEM image of the extracted amyloid fibrils of WE 4 (a).

According to LDS-PAGE (Fig. 13 a), TCB fractions contained many proteins, which became less by multiple washing. After collagenase incubation and multiple washes with TEB, the protein amount decreased further. The protein band with decreasing intensity at approximately 27 kDa corresponded to human SAP [128] (Fig. 13 a, green arrow) and disappeared with increasing TEB washing steps. During water extraction, a protein band enriched at approximately 12 kDa, corresponding to the fibril-forming AL proteins (Fig. 13 a, red arrow). Investigation of WE4 by negative stain TEM confirmed amyloid fibrils in the sample (Fig. 13 b). Fibril-containing water extracts were pooled, and the fibrils were disrupted by incubation with 6 M Gnd-HCl. The sample was desalted using a PD10 column, and AL proteins were purified by IEC (Fig. 14 a + b).

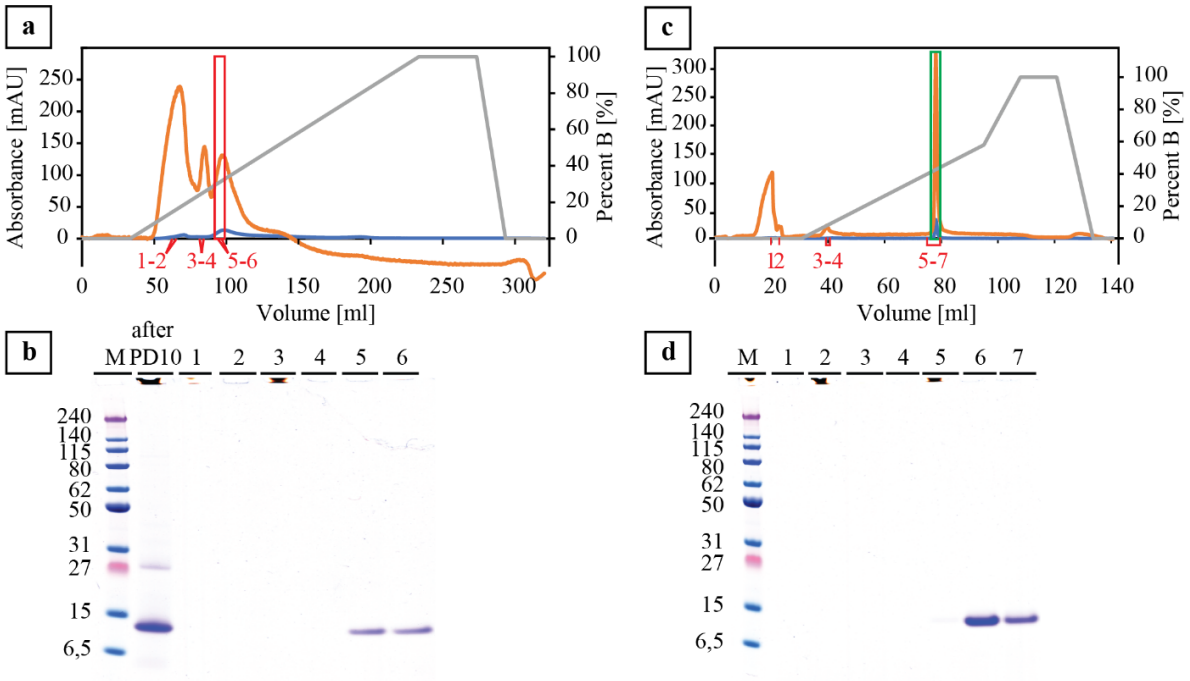


Figure 14: Chromatographic purification of AL protein extracted from heart tissue of case FOR012. a) IEC of the heart extracted AL protein. Primary Y-axis: milli absorbance units (mAU) (orange curve: 215 nm, blue curve: 280 nm). Secondary Y-axis: percentage of elution buffer (gray curve). X-axis: volume in ml. Red numbers: fractions investigated by LDS-PAGE. Red rectangle: fractions further purified by RPC. b) LDS-PAGE analysis of the IEC fractions. M: marker in kDa. After PD10: desalted sample before IEC. Numbers: fractions marked in the chromatogram. Gel columns from LDS-PAGE were removed for convenience only without changing protein bands. c) RPC of the pooled fractions from IEC. Green rectangle: pooled fractions investigated by MS. d) LDS-PAGE analysis of the RPC fractions.

AL protein of case FOR012 eluted in fractions 5 and 6, as indicated by a protein band at approximately 12 kDa in LDS-PAGE (Fig. 14 b). These fractions corresponded to the elution volume at 30.0 % to 35.0 % of the IEC elution buffer (Fig. 14 a, red rectangle). The fractions were pooled and applied to RPC (Fig. 14 c + d). In RPC, the AL proteins eluted in the fractions 5 to 7, visible as protein bands at 12 kDa in LDS-PAGE (Fig. 14 d). These fractions corresponded to a narrow peak in RPC at 42.3 % to 44.2 % of the RPC elution buffer (Fig. 14 c). AL protein-containing fractions were pooled, lyophilized, and analyzed by MS (Fig. 15).

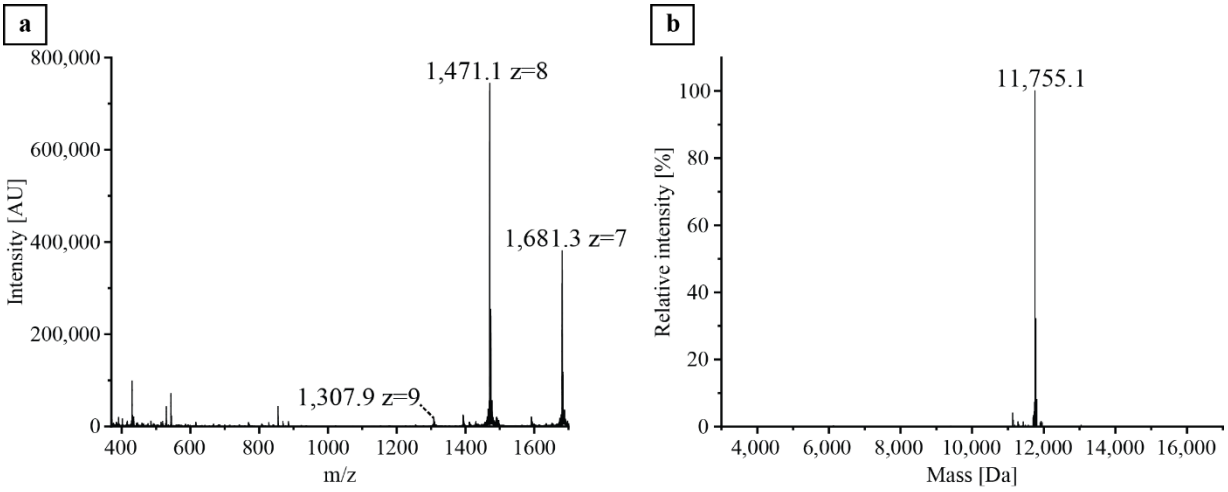


Figure 15: MS analysis of purified AL protein from heart tissue of case FOR012. a) Mass spectrum of the purified AL protein. Y-axis: intensity in arbitrary units (AU). X-axis: mass to charge ratio (m/z). Numbers: most intense isotopes of isotopic distributions corresponding to the AL protein (charge = z). b) Deconvoluted mass spectrum of a). Y axis: relative intensity in percent. X-axis: monoisotopic mass in Da. Number: monoisotopic mass of the AL protein in Da.

The mass spectrum of the purified AL protein displays isotopic distributions of proteins with seven, eight, and nine charges (Fig. 15 a). Deconvolution of these isotopes resulted in a monoisotopic mass of 11,755.1 Da (Fig. 15 b), confirming the protein band at 12 kDa in LDS-PAGE as AL-protein.

In the next step, the IEC and RPC columns were cleaned. Then it was checked if there was still AL protein on the columns that could be transferred to the purification from serum. For this

purpose, a buffer sample was prepared as follows. 50 mM ethanolamine was incubated with 6 M Gnd-HCl overnight. The mixture was desalted using a PD10 column equilibrated with IEC running buffer, and the resulting eluate was applied to IEC (Fig. 16 a + b).

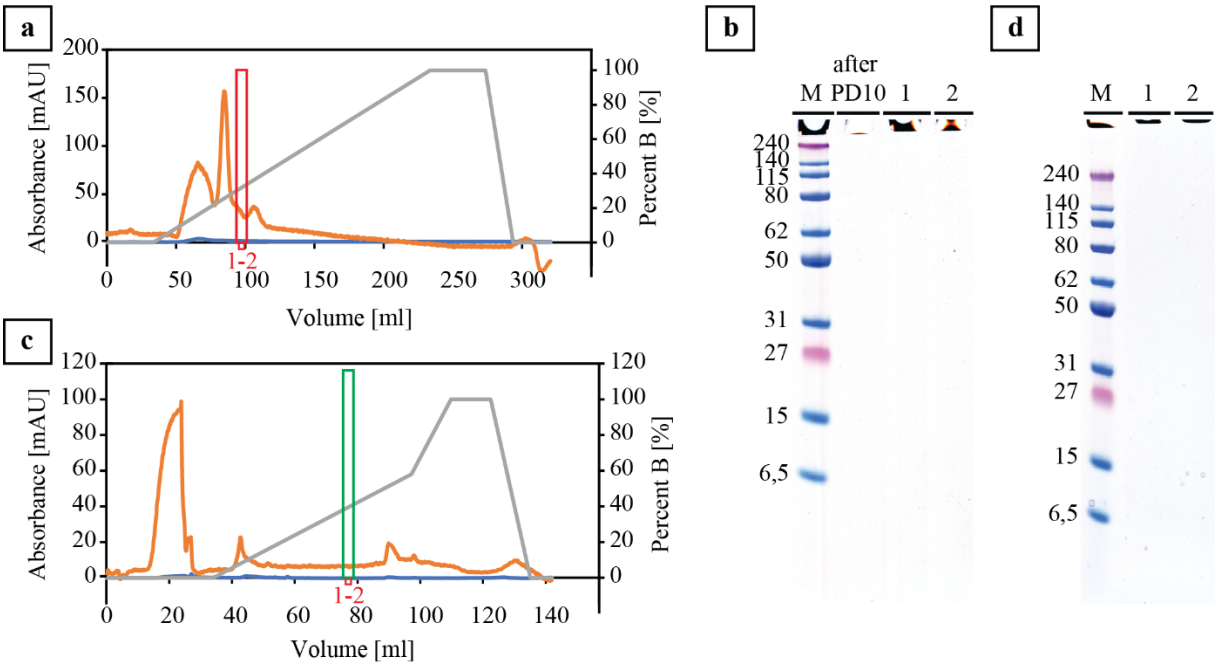


Figure 16: Chromatographic purification of the buffer. a) IEC of the buffer sample. Primary Y-axis: milli absorbance units (mAU) (orange curve: 215 nm, blue curve: 280 nm). Secondary Y-axis: percentage of elution buffer (gray curve). X-axis: volume in ml. Red numbers: fractions investigated by LDS-PAGE. Red rectangle: fractions further purified by RPC. b) LDS-PAGE analysis of the IEC fractions (a). M: marker in kDa. After PD10: desalted sample before IEC. Numbers: fractions marked in the chromatogram. Gel columns from LDS-PAGE were removed for convenience only without changing protein bands. c) RPC of the pooled fractions from IEC. Green rectangle: Pooled fractions investigated by MS. d) LDS-PAGE analysis of the RPC fractions (c).

The absorbance at 215 nm increased during the IEC purification of the buffer sample (Fig. 16 a). However, no protein bands were visible in LDS-PAGE when investigating the fractions that previously contained AL protein from cardiac tissue (30 % to 35 % of IEC elution buffer, red rectangle) (Fig. 16 a + b). When these fractions were applied to RPC, no increase in absorbance was observed in the RPC fractions that previously contained AL protein from cardiac tissue (42.3 % to 44.2 % of RPC elution buffer, green rectangle) (Fig. 16 c). Moreover, no protein bands were visible in the LDS-PAGE analysis of these fractions (Fig. 16 d). To

ensure that these fractions do not contain even small amounts of AL protein, they were analyzed by MS (Fig. 17).

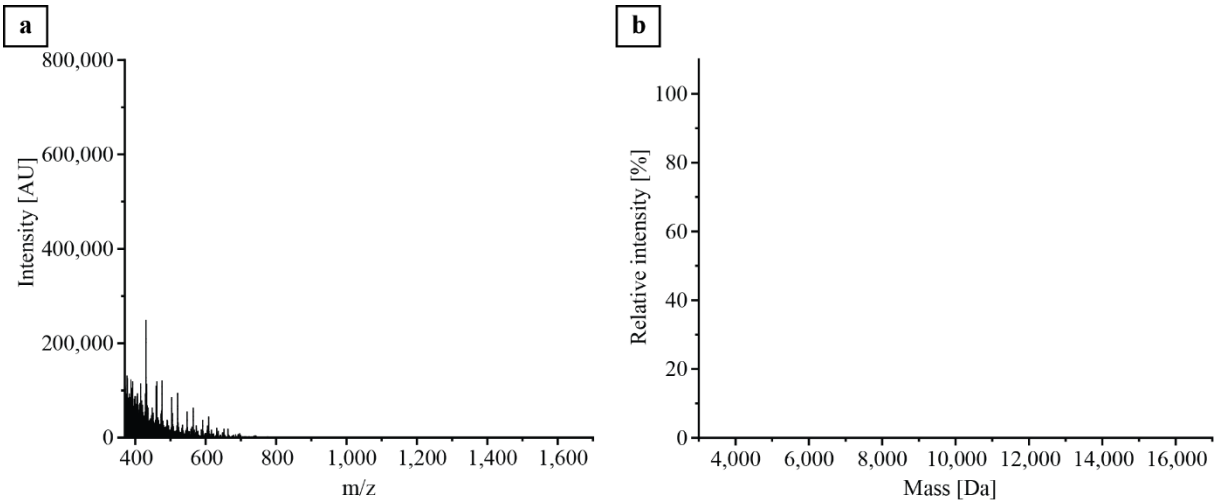


Figure 17: MS analysis of the buffer sample. a) Mass spectrum of the buffer sample. Y-axis: intensity in arbitrary units (AU). X-axis: mass to charge ratio (m/z). b) Deconvoluted mass spectrum of a). Y-axis: relative intensity in percent. X-axis: monoisotopic mass in Da.

MS revealed only molecules of low molecular weight in the RPC fractions of the buffer sample (Fig. 17 a). Deconvolution of the mass spectrum did not provide evidence of the presence of any molecule with a monoisotopic mass of 3000 Da or more (Fig. 17 b). Consequently, the columns were considered successfully cleaned.

As a next step, the purification procedure of AL protein from heart tissue was applied to serum from case FOR012. For this purpose, the serum was incubated overnight in 6 M Gnd-HCl to disrupt multimeric molecules. The mixture was desalted using a PD10 column and applied to IEC (Fig. 18 a + b).

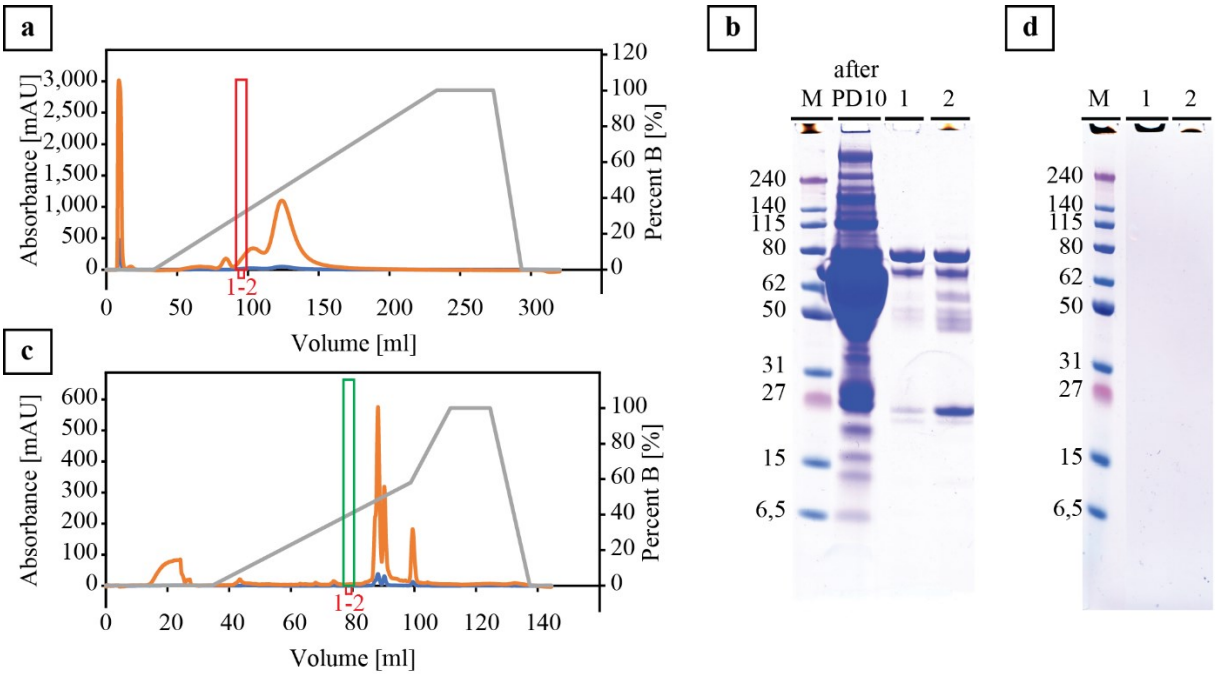


Figure 18: Chromatographic purification of serum from case FOR012. a) IEC of the serum sample. Primary Y-axis: milli absorbance units (mAU) (orange curve: 215 nm, blue curve: 280 nm). Secondary Y-axis: percentage of elution buffer (gray curve). X-axis: volume in ml. Red numbers: fractions investigated by LDS-PAGE. Red rectangle: fractions further purified by RPC. b) LDS-PAGE analysis of the IEC fractions (a). M: marker in kDa. After PD10: desalted sample before IEC. Numbers: fractions marked in the chromatogram. Gel columns from LDS-PAGE were removed for convenience only without changing protein bands. c) RPC of the pooled fractions from IEC. Green rectangle: pooled fractions investigated by MS. d) LDS-PAGE analysis of the RPC fractions (c).

After desalting with the PD10 column, the serum sample contained a lot of proteins confirmed by LDS-PAGE (Fig. 18 b). In IEC, most of these proteins were present in the flow-through or eluted at approximately 50 % of the IEC elution buffer (Fig. 18 a). Based on LDS-PAGE, few serum proteins were present in the IEC fractions that previously contained AL protein from heart tissue (30 % to 35 % of IEC elution buffer, red rectangle) (Fig. 18 b). However, no AL protein bands were detected, and the fractions were applied blindly to RPC. No protein was present in the RPC fractions that previously contained AL protein from heart tissue (42.3 % to 44.2 % of RPC elution buffer, green rectangle), as illustrated by LDS-PAGE (Fig. 18 c + d). MS was used to identify even the lowest amounts of protein in these RPC fractions (Fig. 19).

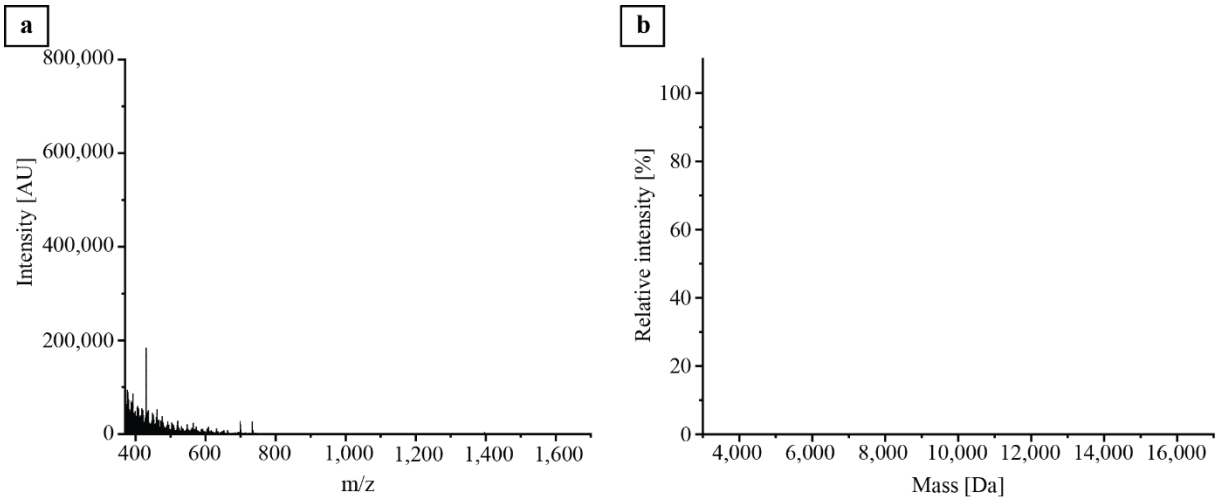


Figure 19: MS analysis of the serum sample of case FOR012. a) Mass spectrum of the serum sample. Y-axis: intensity in arbitrary units (AU). X-axis: mass to charge ratio (m/z). b) Deconvoluted mass spectrum of a). Y-axis: relative intensity in percent. X-axis: monoisotopic mass in Da.

MS revealed only molecules of low molecular weight in the RPC fractions of the serum sample (Fig. 19 a). Deconvolution of the mass spectrum did not provide evidence of the presence of any molecule with a monoisotopic mass of 3,000 Da or more (Fig. 19 b), suggesting that no AL protein is present in the serum of FOR012.

In addition to FOR012, the same strategy was used in three other cases (FOR005, FOR010, FOR140) to investigate the presence of AL protein in serum. In all four cases (FOR005, FOR010, FOR012, FOR140), AL protein from heart tissue eluted at different concentrations of elution buffer (v/v) in IEC (FOR005: 22.5 % – 27.5 %, FOR010: 22.7 % – 27.7 %, FOR012: 30.0 % – 35.0 %, FOR140: 32.5 % – 42.5 %) and RPC (FOR005: 37.9 % – 41.5 %, FOR010: 41.5 % – 43.3 %, FOR012: 42.3 % – 44.2 %, FO140: 40.5 % – 43.3 %) (App. Fig. 69 – 72). After purification, the RPC fractions of AL protein from heart tissue and serum were analyzed by MS (Fig. 20).

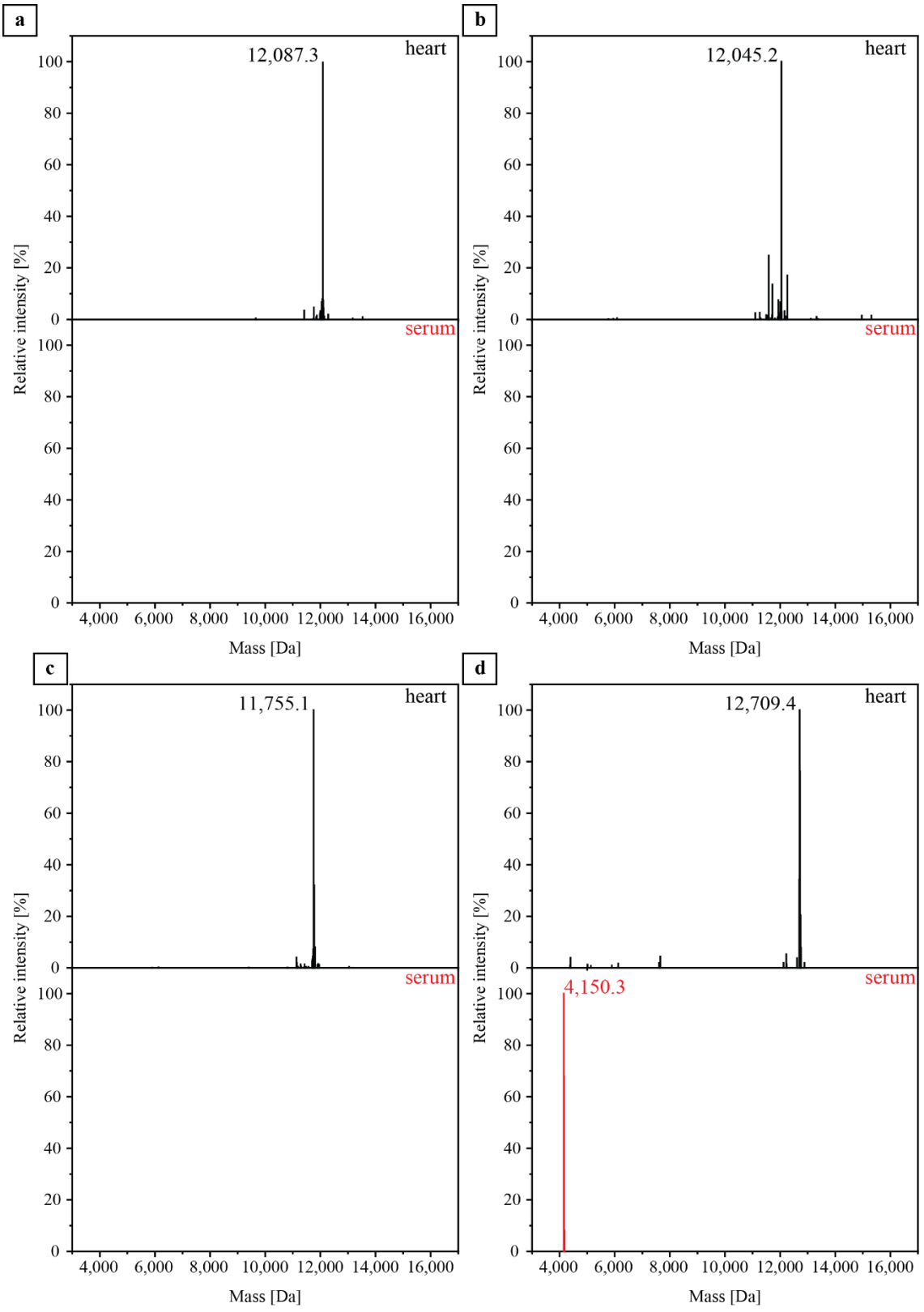


Figure 20: Deconvoluted mass spectra of serum (red) and AL protein from heart tissue (black). a) Deconvoluted mass spectra of FOR005. Top: RPC fraction of heart extracted AL protein (black). Bottom: RPC fraction of Serum (red). Numbers: Monoisotopic mass of the most prominent peak. Y-axis: relative intensity in percent. X-axis: monoisotopic mass in Da. b) Deconvoluted mass spectra of FOR010. c) Deconvoluted mass spectra of FOR012. d) Deconvoluted mass spectra of FOR140.

The deconvoluted mass spectra of AL protein from heart tissue showed a major peak between 11.7 and 12.8 kDa, indicating that the AL proteins have a mass of approximately 12 kDa (Fig. 20 a – d). In contrast, the deconvoluted mass spectra of the serum fractions contained no peaks at approximately 12 kDa. In general, only one peak was found in the deconvoluted mass spectra of the serum fractions. That peak had a molecular weight of 4 kDa and occurred exclusively in FOR140 (Fig. 20 d). However, the corresponding protein was too small to be an AL protein and thus may have represented other serum components that have been copurified.

In summary, these data demonstrated the successful purification of AL protein from heart tissue by IEC and RPC. Applying the same purification protocol to serum and investigating the purified fractions by MS provided no evidence that the AL proteins found in amyloid deposits are also present in the blood. This observation suggests that LC fragmentation occurs at the deposition site after fibril formation.

4.2 Establishment of a universal protocol to determine the primary structure of AL proteins extracted from abdominal fat and heart tissue

A central aim of this thesis was to establish a method to resolve the primary structure of AL proteins in as many cases as possible. The determination of the primary structure of AL proteins has been successfully demonstrated in two cases [14,63]. However, in these studies, AL protein was extracted from heart tissue, which is rarely available. Therefore, AL proteins should also be extracted from abdominal fat in this work to increase the number of identifiable primary structures. The advantage of abdominal fat is that it is a readily accessible patient tissue, which contains amyloid fibrils representative of amyloid deposits throughout the body [14,129]. In addition, methods for extracting amyloid fibrils from abdominal fat already exist and only had to be optimized in this work. Optimization focused on two things, increasing the yield of amyloid fibrils and extracting amyloid fibrils without using collagenase. Collagenase is commonly used in the field of amyloidosis to break up connective tissue before fibril extraction [59,60,130–132]. However, it is discussed that collagenase leads to artefactual degradation of amyloid fibrils, which is why extraction methods without collagenase were tested.

4.2.1 Establishment of an extraction method for amyloid fibrils from abdominal fat tissue

At first, a previously established protocol from Annamalai et al. [14] to extract amyloid fibrils from abdominal fat was optimized. For this purpose, the volumes used in the washing and extraction steps were reduced relative to the initial amount of tissue to increase the fibril concentration. Furthermore, a fat phase forming in the washing steps was retained until after incubation with collagenase because it could contain amyloid fibrils. These adaptations resulted in the following protocol. Abdominal fat was washed with 0.2 ml TCB per 100 mg tissue to remove blood components. The number of washing steps was adjusted to the level of

contamination. After each washing step, a fat layer formed on the aqueous phase, which was retained throughout the TCB washing. The aqueous supernatant was removed with a syringe to avoid disturbing the fat layer. The remaining tissue was incubated with 0.4 ml TCB and 2 mg collagenase per 100 mg of initial tissue at 37 °C overnight. After incubation with collagenase, the fat layer was removed, and the remaining tissue was washed a minimum of three times with 0.2 ml TEB per 100 mg of initial tissue. Fibrils were extracted by resuspending the remaining tissue at least five times with 0.1 ml water per 100 mg of initial tissue. After centrifugation, the fibrils were present in the supernatant. The fibril extraction was verified by LDS-PAGE, and fibrils were confirmed by negative stain TEM. Results are representatively shown for case FOR103 (Fig. 21).

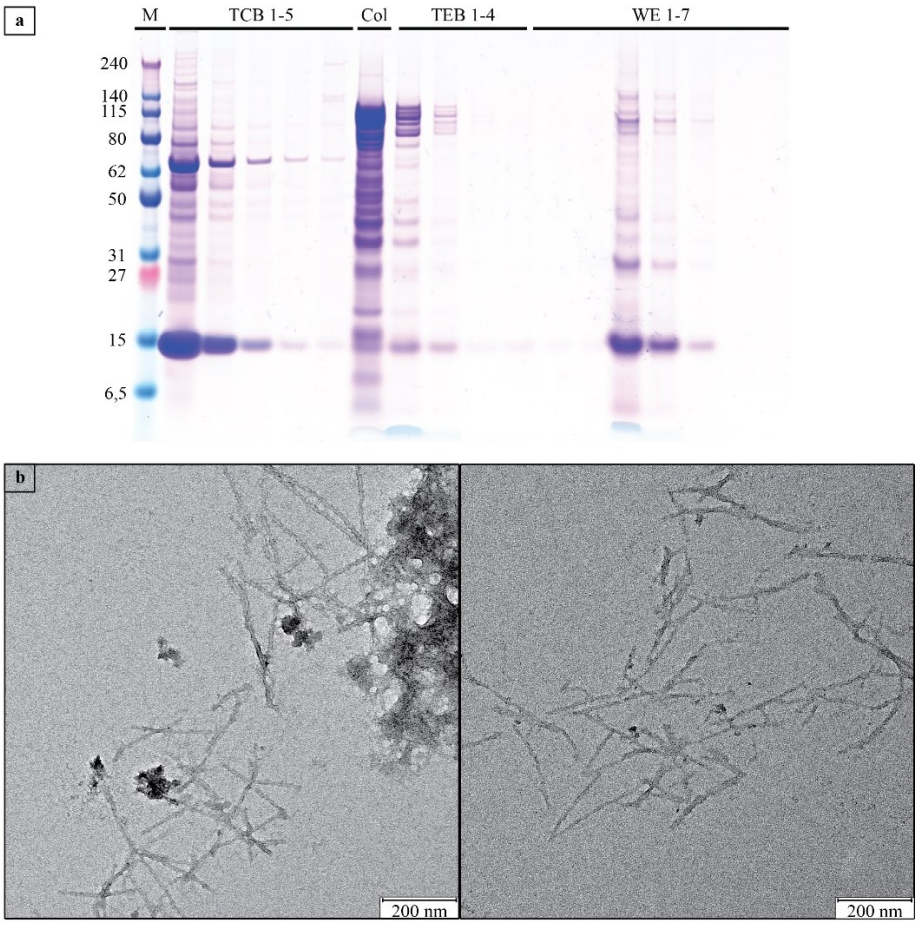


Figure 21: Fibril extraction from abdominal fat of FOR103 using an adapted version of the water extraction protocol of Annamalai et al.* [14]. *(a) LDS-PAGE analysis. M: marker in kDa. Col: supernatant after collagenase incubation. WE: water extracts. (b) Negative stain TEM images of WE5.

The protein bands in LDS-PAGE were similar to those from fibril extractions from cardiac tissue. According to LDS-PAGE, the tissue contained many proteins that were mainly removed by washing with TCB (Fig. 21 a). The most prominent protein bands were located at 15 kDa and 65 kDa, suggesting the presence of hemoglobin and albumin (Fig. 21 a) [126,133]. After incubation with collagenase, contaminants were largely removed using TEB (Fig. 21 a). In the water extracts, proteins with a molecular weight of approximately 15 kDa were enriched (Fig. 21 a). Negative stain TEM confirmed these as the AL proteins since amyloid fibrils are visible in WE5 (Fig. 21 b). LDS-PAGE and negative stain TEM showed that the adapted water extraction protocol successfully extracted amyloid fibrils (Fig. 21). The AL protein was enriched, and the downscaling worked (Fig. 21).

As a next step, chloroform was tested to extract amyloid fibrils from abdominal fat without using collagenase. The idea was to expose the amyloid fibrils by removing the surrounding lipids with chloroform, removing contaminants by TCB and TEB, and extracting the amyloid fibrils by resuspension in water. For this purpose, 40 mg of abdominal fat (from case P6/16) was resuspended in a mixture of 0.25 ml TCB and 0.25 ml chloroform. After incubating the sample for 1 min, an organic and an aqueous phase formed. The organic phase containing the lipids was removed, and the aqueous TCB phase was centrifuged at 3,100 xg for 5 min. The resulting pellet was washed with 0.5 ml TEB and extraction was performed by resuspending the remaining tissue in 0.5 ml water three times. The extraction approach was evaluated by LDS-PAGE and negative stain TEM (Fig. 22).

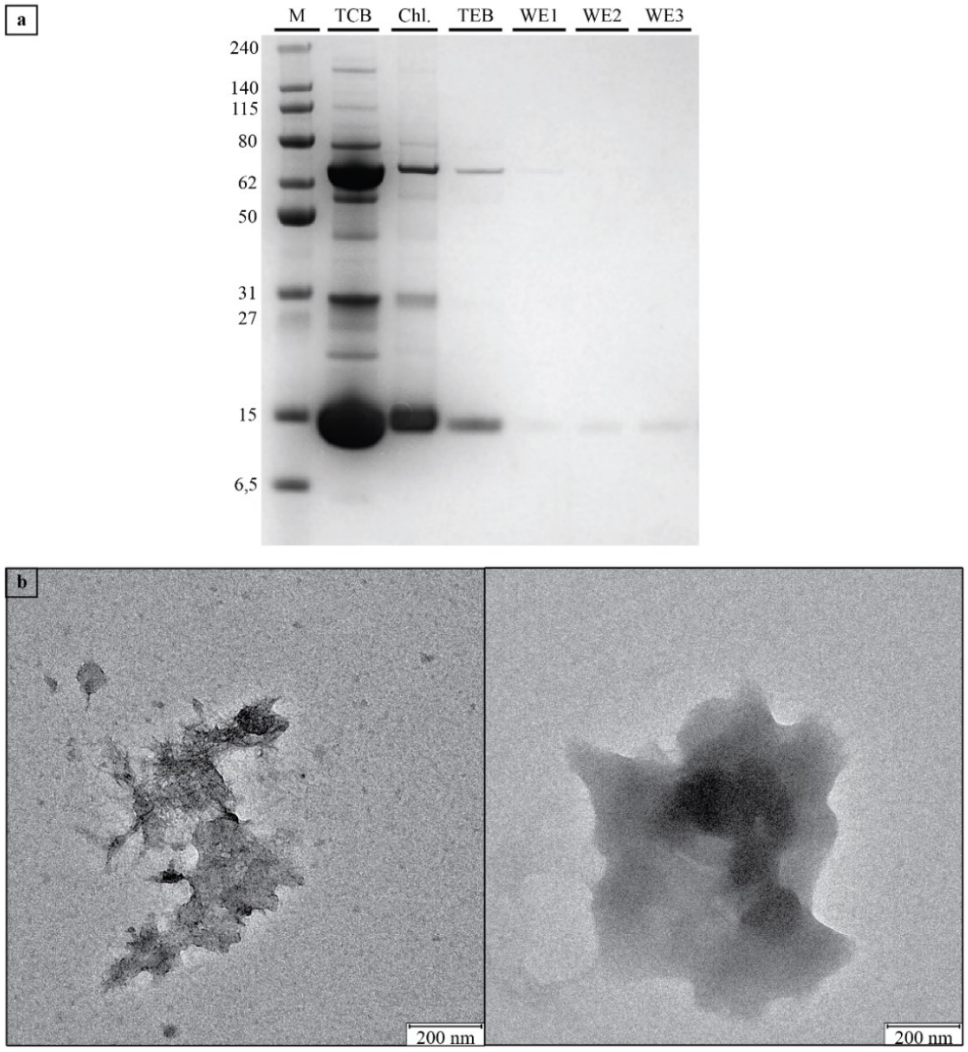


Figure 22: Approach to extract amyloid fibrils from abdominal fat of case P6/16 using chloroform. a) LDS-PAGE analysis. M: marker in kDa. Chl.: chloroform phase. WE: water extract. b) Negative stain TEM images of WE3.

According to LDS-PAGE, the TCB phase contained various proteins, especially in the range of 15 kDa and 65 kDa (Fig. 22 a), suggesting the presence of hemoglobin and albumin [126,133]. The discarded chloroform phase contained the same proteins but in a lower concentration (Fig. 22 a). Washing with TEB removed the major contaminants so that only one protein band was visible in the water extracts (Fig. 22 a). However, no amyloid fibrils were visible in negative stain TEM (Fig. 22 b), indicating no successful fibril extraction.

As a next step, acetone was used, following the same idea to extract amyloid fibrils without using collagenase, as described for chloroform. However, the procedure differed, and the extraction with acetone was conducted as follows. 40 mg of abdominal fat was washed with 0.5 ml TCB, followed by washing with 0.5 ml TEB. The remaining tissue was washed with 0.5 ml acetone to remove the lipids surrounding the fibrils. Subsequently, fibril extraction was performed by resuspending the remaining tissue in 0.5 ml of water, centrifuging, and collecting the supernatant. The extraction approach was evaluated by LDS-PAGE and negative stain TEM (Fig. 23).

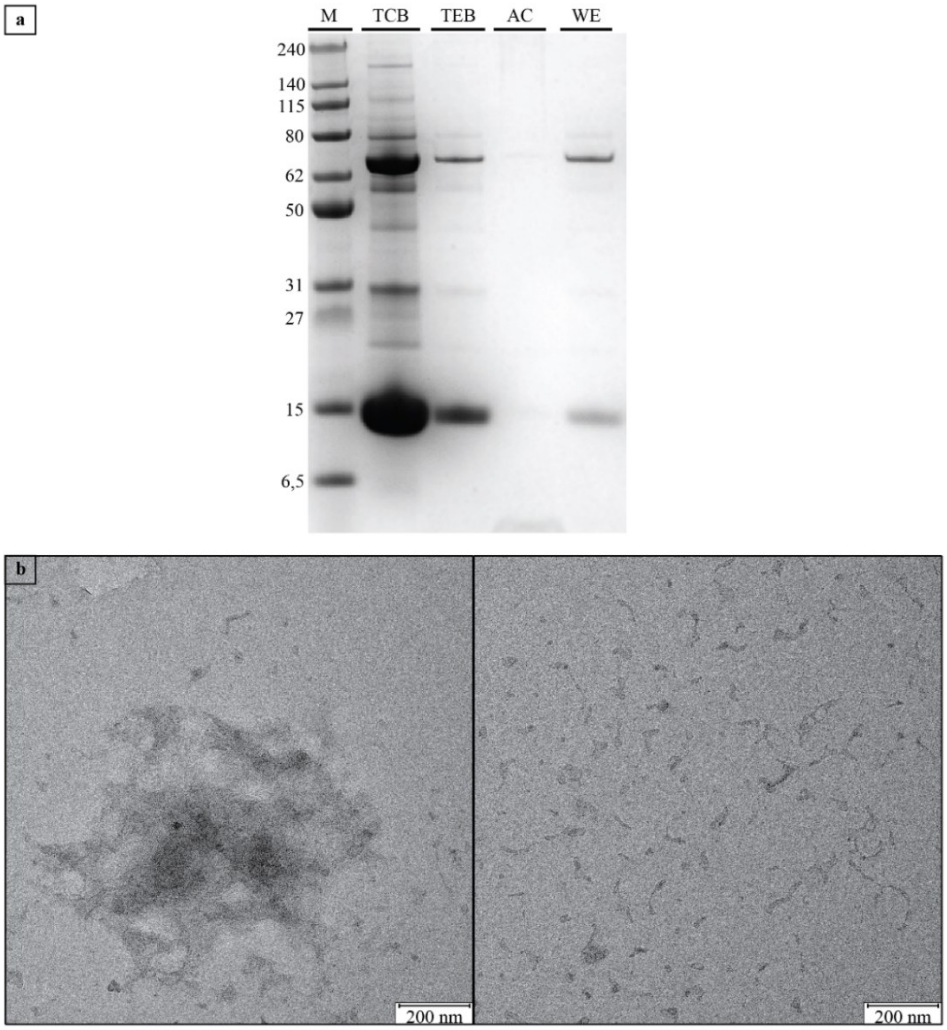


Figure 23: Approach to extract amyloid fibrils from abdominal fat of case P6/16 using acetone. a) LDS-PAGE analysis. M: marker in kDa. AC: acetone phase. WE: water extract. b) negative stain TEM analysis of WE.

According to LDS-PAGE, TCB and TEB washing removed the majority of proteins (Fig. 23 a). The discarded acetone phase only showed faint protein bands at 15 kDa and 65 kDa in LDS-PAGE (Fig. 23 a). Those two protein bands were also visible in the water extract (Fig. 23 a), but no amyloid fibrils could be detected by electron microscopy (Fig. 23 b), indicating no successful fibril extraction.

As a next step, detergents were used to remove lipids and thus extract amyloid fibrils from abdominal fat without using collagenase. For this purpose, TX-114, SDC, and CHAPS were tested, which are known to be mild and do not readily denature proteins [106]. Each detergent was dissolved in high and low salt extraction buffer, and 40 mg of abdominal fat (from case P2/16) was resuspended in 0.5 ml of the respective detergent solution. The resuspended tissue was centrifuged at 8,000 xg for 20 min, and the supernatant was removed. The pellets were resuspended in 20 µl of water and analyzed with the supernatants by LDS-PAGE (Fig. 24 a). The resuspended pellets were also examined for amyloid fibrils by negative stain TEM (Fig. 24 b – g).

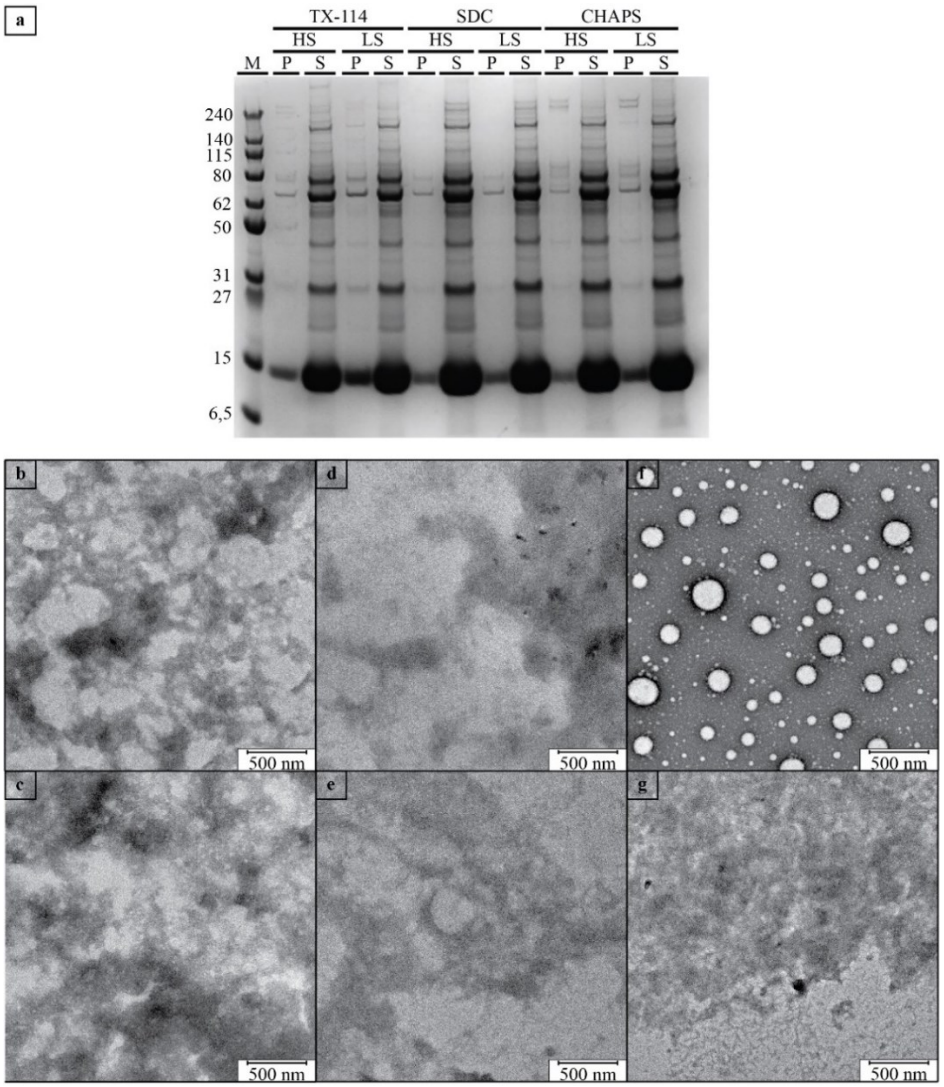


Figure 24: Extraction of amyloid fibrils from abdominal fat of case P2/16 using the detergents TX-114, SDC, and CHAPS. a) LDS-PAGE analysis. M: marker in kDa. HS: high salt. LS: low salt. P: pellet. S: supernatant. b) – g) Negative stain TEM images of the HS pellet of TX-114 (b), the LS pellet of TX-114 (c), the HS pellet of SDC (d), the LS pellet of SDC (e), the HS pellet of CHAPS (f) and the LS pellet of CHAPS (g).

LDS-PAGE analysis showed a few protein bands in the pellet of the detergent samples, with the most prominent band at approximately 15 kDa (Fig. 24 a). The supernatant of each sample contained protein bands at the same height as the pellet with increased intensity (Fig. 24 a). No difference in band intensity and height was observed comparing the different detergents and the high and low salt conditions. Negative stain TEM detected no amyloid fibrils in all resuspended pellets (Fig. 24 b – g). These results suggest that the detergents TX-114, SDC, and CHAPS do

not successfully extract amyloid fibrils, and the different salt conditions have no effect on the precipitation of amyloid fibrils.

Next, a protocol was applied, which was previously established to extract tau fibrils from the brain without using collagenase [109,134]. For this purpose, 100 mg of abdominal fat (from case P232/15) was dissolved in 1 ml of UB and centrifuged at 16,900 xg for 20 min. The supernatant was stored, and the remaining tissue was resuspended in UB and centrifuged one additional time. Both resulting supernatants were pooled and incubated with 1 % N-lauryl sarcosine for 1 h. Subsequently, the pooled supernatants were centrifuged at 100,000 xg for 1 h to pellet and extract the amyloid fibrils. However, no pellet was visible. To test whether the pellet was too small to be visible, the supernatant was removed, and potentially pelleted proteins were resuspended in LDS sample buffer (SB) and investigated by LDS-PAGE (Fig. 25 a). The supernatant was investigated by LDS-PAGE and negative stain TEM to exclude that the amyloid fibrils remained soluble during centrifugation (Fig. 25 a + b).

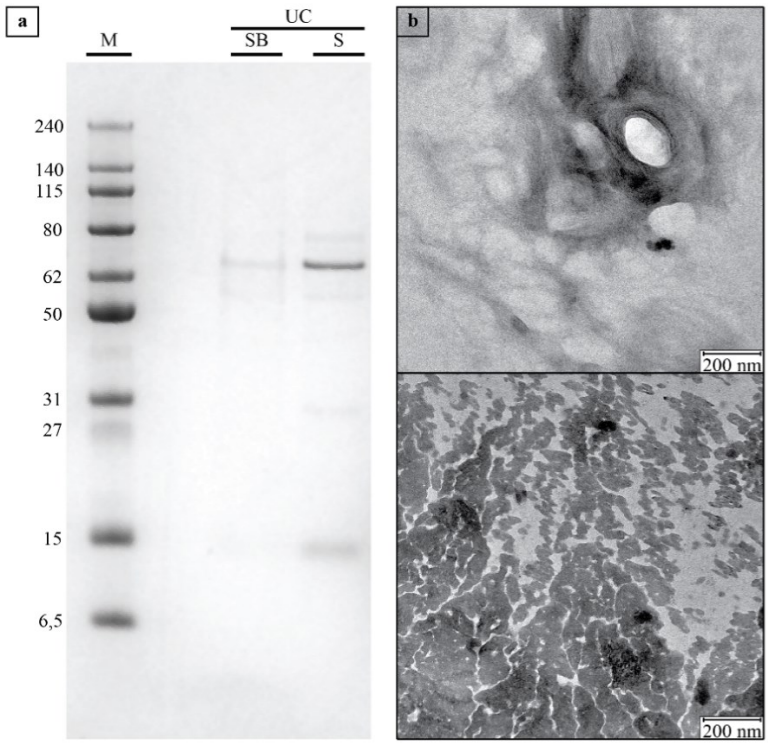


Figure 25: Fibril extraction from abdominal fat of case P232/15 with N-lauryl sarcosine. a) LDS-PAGE analysis of the UC fractions. M: marker in kDa. SB: sample buffer used to dissolve potential amyloid fibrils sticking to the bottom of the test tube. S: supernatant. b) Negative stain TEM of the supernatant (S) from UC.

LDS-PAGE only showed one protein band in the SB fraction at approximately 65 kDa (Fig. 25 a). The same and a few other protein bands were visible in the supernatant (Fig. 25 a), but no amyloid fibrils were detected by negative stain TEM (Fig. 25 b). In conclusion, no amyloid fibrils could be extracted by this method.

Another method to mechanically extract amyloid fibrils from cardiac tissue without using collagenase has recently been published by Lavatelli et al. [73] and was tested in this work. 250 mg of heart tissue (from case FOR005) was diced and washed with 1 ml of PBS. Then, the tissue was homogenized with a pestle and mortar and washed with 0.5 ml TEB. After repeating homogenization and washing nine times, the extraction protocol of Lavatelli et al. was finished. In this work, the resulting tissue homogenate was resuspended three more times in water to drive the amyloid fibrils to the solution phase. The extraction was validated by LDS-PAGE, and WE3 was analyzed for amyloid fibrils by negative stain TEM (Fig. 26).

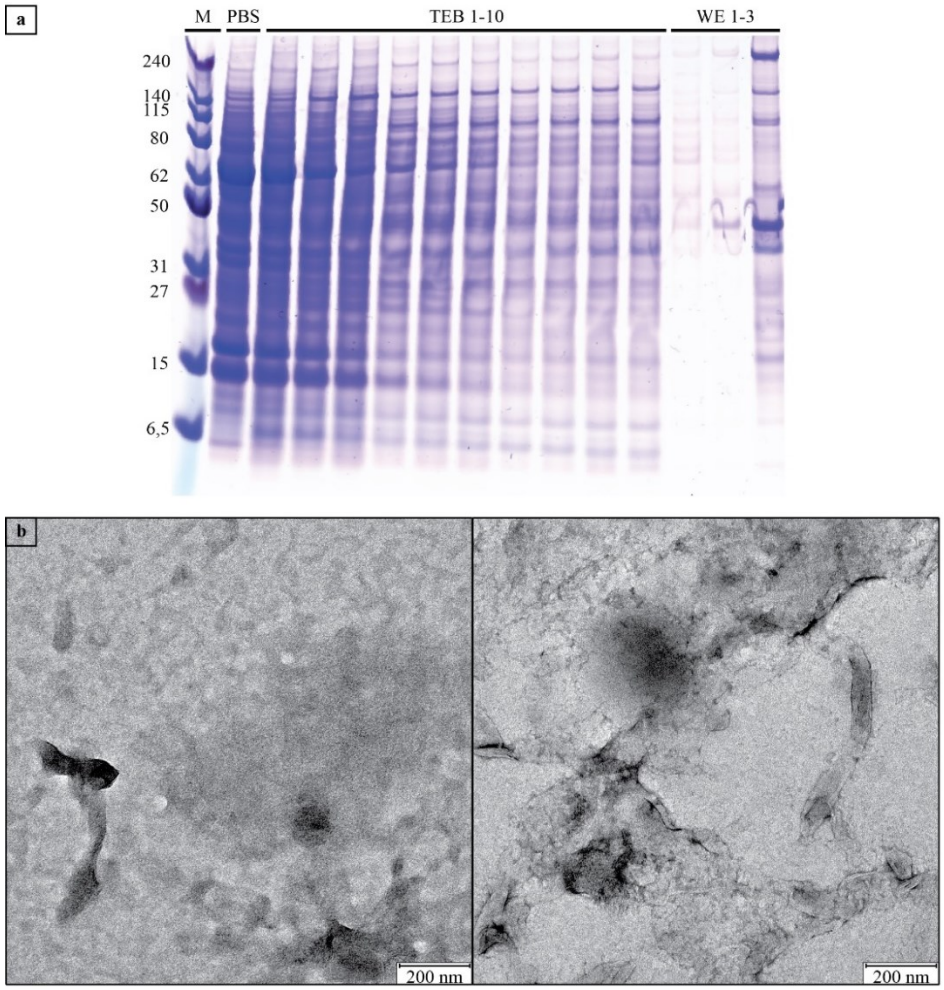


Figure 26: Mechanical extraction of amyloid fibrils from heart tissue of case FOR005 using the protocol of Lavatelli et al.* [73]. *a) LDS-PAGE analysis. M: marker in kDa. PBS: supernatant after washing with PBS. WE: water extracts. b) negative stain TEM of WE3.

According to LDS-PAGE, the tissue homogenate contained a lot of contaminants, which decreased slightly after several washing steps (Fig. 26 a). WE1 to WE3 comprised fewer proteins than the tissue homogenates, but none corresponded to the molecular weight of the AL proteins. Negative stain TEM of water extract 3 showed heavy contamination of the sample and could not detect amyloid fibrils (Fig. 26 b). Consequently, no intact amyloid fibrils could be extracted by this method.

For this reason, another method was tested to extract amyloid fibrils mechanically. It is based on a previous publication using a fractionator to dissociate lipoaspirates [110]. The idea was to build a fractionator from two connected syringes disrupting the tissue and exposing the fibrils. First, 400 mg abdominal fat aspirate (from case P6/16) was washed with 0.8 ml TCB five times. Subsequently, the tissue was resuspended in 1.6 ml TCB containing EDTA-free protease inhibitor and transferred into the fractionator. The resuspended tissue was disrupted by pushing it from one side of the fractionator to the other for 1 min and 50 s. Disrupting the tissue was repeated three times, and the resulting homogenate was centrifuged at 3,100 xg for 30 min. The fat layer and the supernatant were removed, and the pellet was washed with 0.15 ml TEB thrice. The fibrils were extracted by resuspending the remaining tissue in 0.15 ml water, centrifuging, and collecting the supernatant six times. Subsequently, the extraction was evaluated by LDS-PAGE and negative stain TEM (Fig. 27).

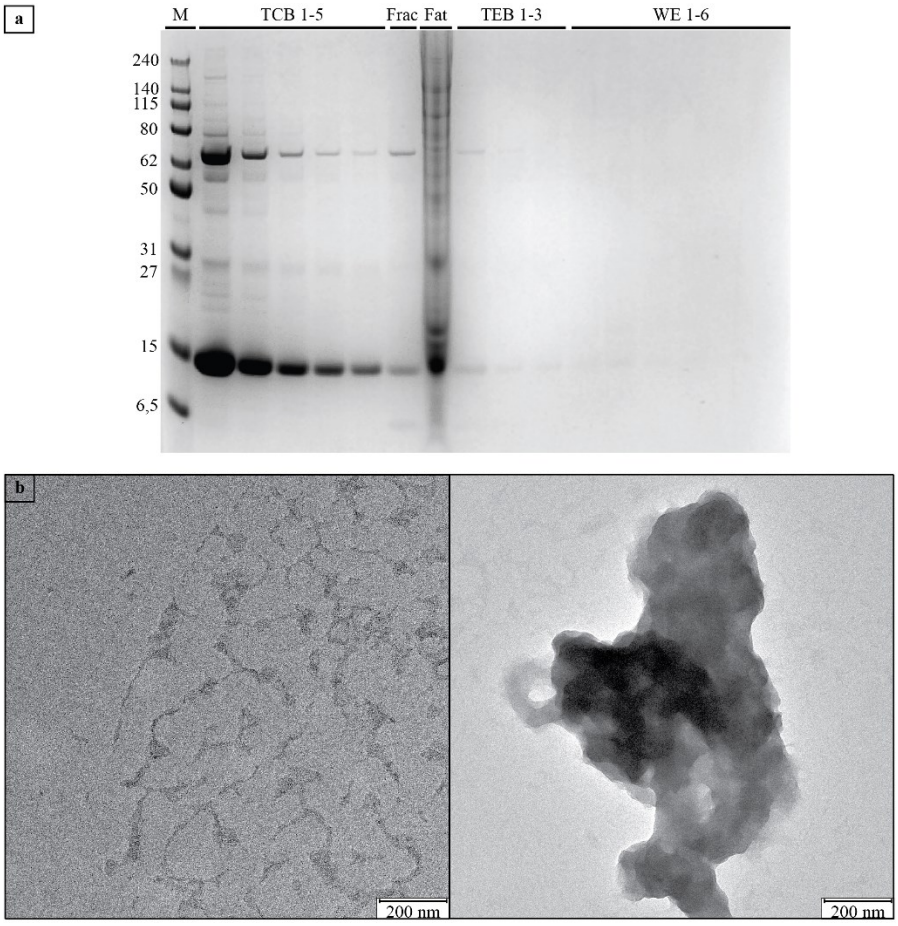


Figure 27: Mechanical extraction of amyloid fibrils from case P6/16 using a fractionator. a) LDS-PAGE analysis. M: marker in kDa. Frac: supernatant after fractionation. Fat: fat layer. WE: water extract. b) Negative stain TEM analysis of water extract 5.

LDS-PAGE analysis revealed mainly two proteins with a molecular weight of approximately 15 kDa and 65 kDa in the TCB fractions (Fig. 27 a). After homogenization, most of the proteins were present in the fat (Fig. 27 a). Removing the fat layer removed most of the proteins from the homogenate, resulting in minute amounts of protein with a molecular weight of about 15 kDa in the water extracts (Fig. 27 a). Negative stain TEM could not detect amyloid fibrils in WE5 (Fig. 27 b), indicating that this method failed to extract pure amyloid fibrils. Furthermore, this approach required too much initial abdominal fat aspirate and was rejected.

In summary, neither organic solvents, detergents, nor mechanical extraction were an alternative to collagenase and resulted in the successful extraction of intact and pure amyloid fibrils. The

adapted water extraction protocol still provided the best results and was used for all further extractions from abdominal fat.

4.2.2 Establishment of a workflow to determine the primary structure of AL proteins

At the beginning of this project, the aim was to use cDNA and the total mass of the AL proteins to determine their primary structure. For this purpose, collaboration partners from Heidelberg University Hospital should determine the LC sequence of each case by cDNA sequencing of bone marrow samples and send corresponding patient tissue to Ulm University. In Ulm, fibrils should be extracted from the tissue, and the total mass of fibril-forming AL proteins should be determined by MS. The comparison of the cDNA sequence to protein mass should then provide the fragment length and PTMs of the AL protein. Within the framework of the research unit FOR2969, bone marrow samples from 192 cases of systemic AL amyloidosis were collected at Heidelberg University Hospital. The bone marrow from 162 cases was subjected to cDNA sequencing by Stefanie Huhn. The sequences differed in quality and were graded from 1 to 6. Grade 1 comprises monoclonal sequences and grade 2 polyclonal sequences with less than five unidentifiable amino acids, referred to as sequence uncertainties. Grade 3 defines polyclonal sequences with 5 to 10 uncertainties, grade 4 polyclonal sequences with 11 to 15 uncertainties, grade 5 polyclonal sequences with 16 to 20 uncertainties, and grade 6 polyclonal sequences with more than 20 uncertainties. Heidelberg University Hospital collected abdominal fat aspirates from all sequenced 162 cases. The amyloid load of each sample was determined in Heidelberg based on a semiquantitative Congo red staining method [101] that defines an amyloid score. An amyloid score of 0 means that there was no detectable Congo red signal, 1+ indicates minute amounts of amyloid, 2+ indicates small amyloid deposits, 3+ indicates substantial amyloid quantities, and 4+ indicates large amounts of amyloid in the fat [101]. The

plan of the current PhD project was to extract amyloid fibrils from all 162 cases, measure the total mass of the corresponding AL proteins, and assign it to the corresponding cDNA-derived amino acid sequence from Heidelberg. However, after attempting to extract amyloid fibrils from cases with amyloid scores of 1+ and 2+, it became clear that these cases did not contain sufficient amyloid fibrils for further biochemical investigations. Consequently, only 55 abdominal fat aspirates with an amyloid load of +3 and +4 and 5 heart biopsies with cDNA information were sent to Ulm University and subjected to further analyses (Tab. 9).

Table 9: Information on the cases sent to Ulm University for biochemical analyses (n=60) regarding sex, age at the time of tissue sampling, organ involvement, tissue samples, amyloid score, and cDNA sequence grade. GI: Gastrointestinal tract. PN: peripheral nerves. No amyloid score was determined for cases with cardiac tissue. The information was provided by Ute Hegenbart (Heidelberg University Hospital).

Case	Sex	Age	Organ involvement	Tissue	Amyloid load	Sequence grade
FOR101	Male	72	Heart	Abdominal fat	4+	2
FOR102	Female	34	Heart	Abdominal fat	3+	1
FOR103	Female	80	Heart	Abdominal fat	4+	1
FOR104	Female	76	Heart	Abdominal fat	4+	1
FOR105	Female	69	Heart	Abdominal fat	3+	3
FOR106	Male	68	Heart	Abdominal fat	3+	2
FOR107	Male	55	Kidney	Abdominal fat	3+	2
FOR108	Female	50	Kidney	Abdominal fat	3+	3
FOR109	Female	41	Kidney	Abdominal fat	3+	1
FOR112	Male	62	Heart	Abdominal fat	3+	4
FOR113	Male	61	Heart	Abdominal fat	3+	3
FOR116	Male	60	Heart	Abdominal fat	3+	1
FOR117	Female	52	Heart	Heart	(no score)	5
FOR120	Male	71	Heart	Abdominal fat	3+	1
FOR122	Female	75	Heart	Abdominal fat	4+	2
FOR124	Female	72	Heart	Abdominal fat	3+	2
FOR127	Female	68	Heart	Abdominal fat	3+	6
FOR128	Male	58	Heart	Abdominal fat	4+	1
FOR130	Male	57	Heart	Heart	(no score)	2
FOR131	Female	80	Heart	Abdominal fat	3+	4
FOR132	Male	57	Heart	Abdominal fat	4+	1
FOR133	Female	67	Heart	Abdominal fat	4+	4
FOR135	Female	38	Kidney	Abdominal fat	3+	1
FOR136	Female	59	Heart	Abdominal fat	3+	1
FOR137	Female	49	Kidney	Abdominal fat	4+	1
FOR138	Male	70	Heart	Abdominal fat	3+	1
FOR139	Male	63	Kidney	Abdominal fat	3+	2
FOR140	Female	57	Heart	Heart	(no score)	4
FOR142	Male	50	Kidney	Abdominal fat	3+	2
FOR143	Female	68	Kidney	Abdominal fat	3+	1

FOR144	Male	64	Heart	Abdominal fat	3+	1
FOR146	Female	59	Heart/Liver	Abdominal fat	3+	6
FOR147	Male	54	Heart	Abdominal fat	4+	1
FOR148	Male	57	Heart	Abdominal fat	4+	3
FOR149	Female	60	Heart/Kidney	Abdominal fat	3+	2
FOR150	Male	51	Heart/Kidney	Abdominal fat	3+	1
FOR153	Male	64	Heart/Kidney	Abdominal fat	3+	1
FOR154	Female	54	Heart/Kidney	Abdominal fat	4+	4
FOR156	Female	60	Heart/Kidney/GI	Abdominal fat	4+	2
FOR157	Male	71	Heart	Abdominal fat	4+	2
FOR159	Female	77	Heart	Abdominal fat	4+	2
FOR161	Female	62	Heart	Heart	(no score)	2
FOR162	Male	60	Heart	Abdominal fat	3+	5
FOR164	Female	63	Heart/Kidney	Abdominal fat	3+	3
FOR165	Male	69	Kidney	Abdominal fat	3+	2
FOR169	Female	54	Heart	Abdominal fat	4+	1
FOR171	Male	70	Heart	Abdominal fat	3+	5
FOR176	Male	64	Heart	Abdominal fat	3+	1
FOR179	Male	74	Heart	Abdominal fat	3+	2
FOR182	Female	66	Heart	Abdominal fat	4+	1
FOR183	Male	76	Heart/Kidney/PN	Abdominal fat	3+	1
FOR185	Female	68	Heart/Kidney	Abdominal fat	3+	1
FOR186	Male	53	Heart/Kidney	Abdominal fat	4+	1
FOR189	Male	83	Heart/Kidney	Abdominal fat	3+	6
FOR1001	Female	53	Heart/Kidney	Abdominal fat	3+	1
FOR1006	Male	80	Heart	Abdominal fat	3+	1
FOR1007	Female	53	Heart	Heart	(no score)	1
FOR1010	Female	52	Heart	Abdominal fat	3+	1
FOR1015	Female	78	Heart	Abdominal fat	3+	1
FOR1021	Female	65	Heart	Abdominal fat	3+	1

4.2.2.1 Determination of the primary structure of the AL proteins using the monoisotopic mass and the provided cDNA-derived sequence

Fibrils were extracted from abdominal fat using the herein-established water extraction protocol (see section 4.2.1 for details) or from heart tissue using the standard water extraction protocol [63]. The resulting amyloid fibrils were lyophilized and weighted to estimate the amount of the extracted AL protein. The lyophilized sample was sent to the Core Unit Mass Spectrometry and Proteomics at Ulm University to process the samples and acquire MS data. At the core unit, the fibrils were disrupted by incubating in MS dissociation buffer overnight. An aliquot corresponding to 1 µg of AL protein was removed, diluted with 0.1 % (v/v) TFA to 15 µl, and subjected to LC-MS. For each case, an HPLC chromatogram was acquired (Fig. 28 a), and

12,000 to 15,000 mass spectra were recorded over the period of the sample elution (65 min) (Fig. 28 b). The spectra were forwarded to the Institute of Protein Biochemistry and analyzed as part of this work. The analysis is exemplified in figure 28 for case FOR142. At first, all 12,000 to 15,000 mass spectra were combined into one. The summed mass spectrum (Fig. 28 c), which is based on an m/z scale, was deconvoluted using MASH Explorer [111] to plot all molecules according to their monoisotopic mass. In addition, the MASH Explorer determined the charge of each molecule and calculated a confidence score for each deconvolution by comparing the theoretical isotopic distribution with the experimental isotopic distribution of a molecule (Fig. 28 d + e).

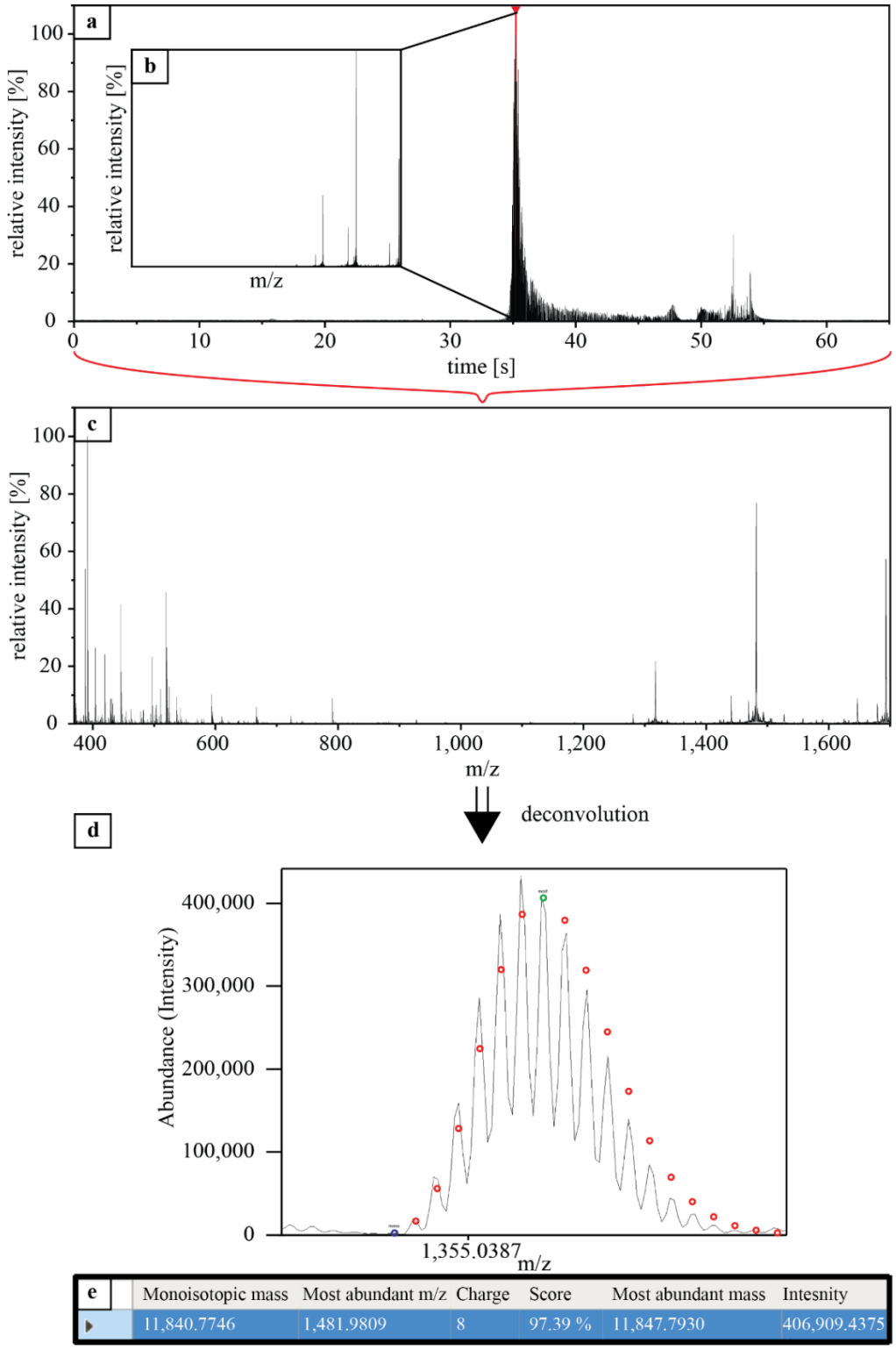
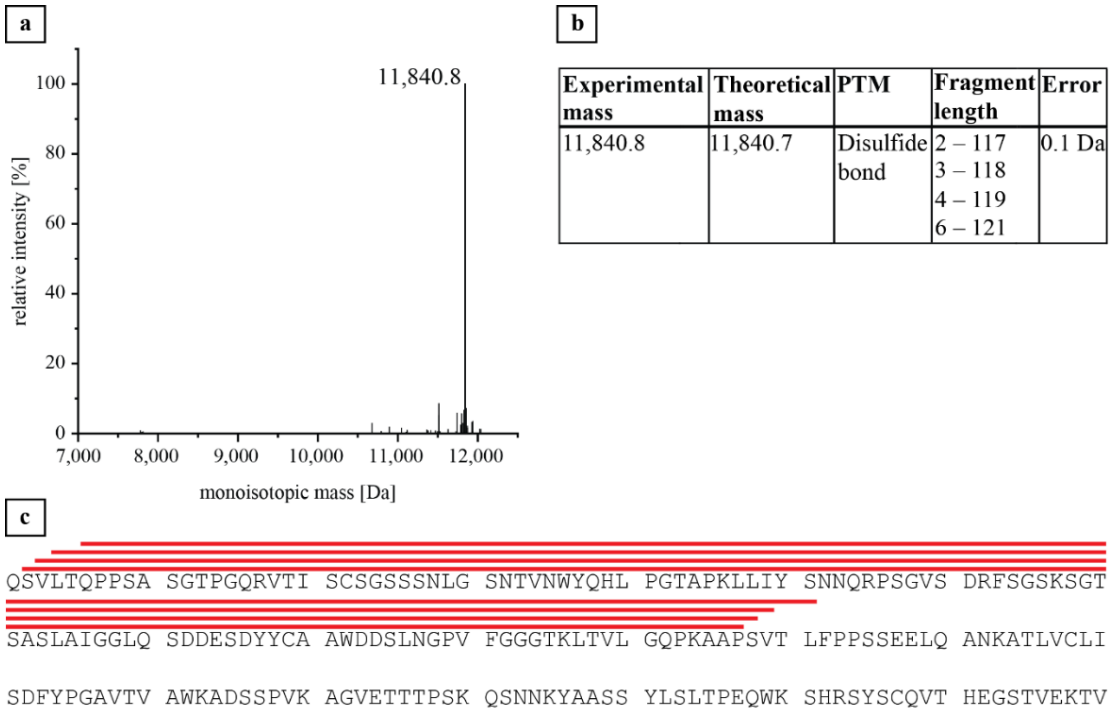


Figure 28: Analysis of the MS data illustrated for FOR142. a) HPLC chromatogram of all eluted proteins. b) Mass spectrum of molecules eluted at a given retention time (red line in a). c) Summed mass spectrum. d) Experimental isotopic distribution aligned with the theoretical isotopic distribution calculated by the MASH explorer* [111] *(circles). Blue circle: monoisotopic peak. Red circles: peaks with different carbon isotopes. Green circles: most intense peak. e) Different parameters of the deconvolution in d).

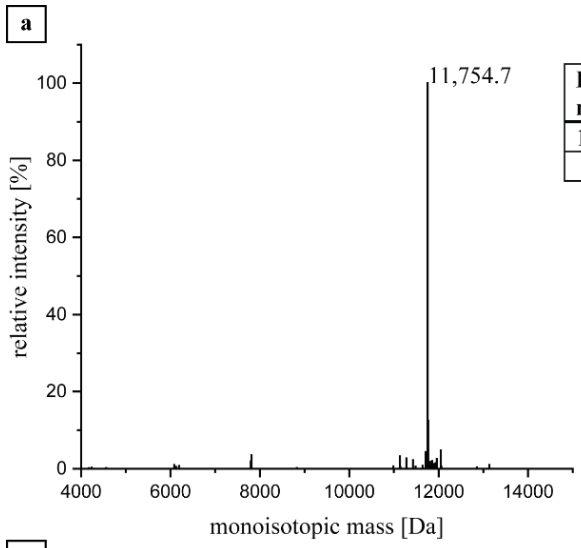
Monoisotopic masses from deconvolutions with a confidence score of 90 % or more that correspond to molecules with at least five charges were selected and plotted against their relative intensity (Fig. 29 a). Mass peaks with the highest intensities were considered to belong to the AL proteins because the AL proteins represented the main fraction of the analyzed water extracts, as indicated by the LDS-PAGE analysis from fibril extraction (App. Fig. 73 – 91). The software mMass [113] was used to assign the experimental mass of the AL proteins to the theoretical mass of the cDNA-derived sequences from Heidelberg (Fig. 29 b + c).



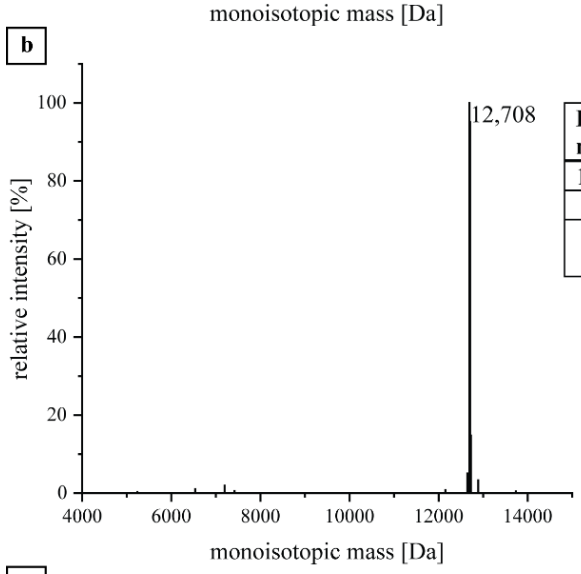
Experimental mass	Theoretical mass	PTM	Fragment length	Error
11,840.8	11,840.7	Disulfide bond	2 – 117 3 – 118 4 – 119 6 – 121	0.1 Da

Figure 29: Sequence assignment of case FOR142 using the total monoisotopic mass and the cDNA-derived amino acids sequence from Heidelberg. a) Deconvoluted mass spectrum. Number: Assigned monoisotopic mass. b) Experimental mass assignment to a theoretical mass of a LC fragment. PTM: PTM needed to assign the sequence. Error: Deviation of experimental and theoretical mass. c) cDNA-derived precursor LC sequence. Red line: assigned fragments fitting to the monoisotopic mass from a).

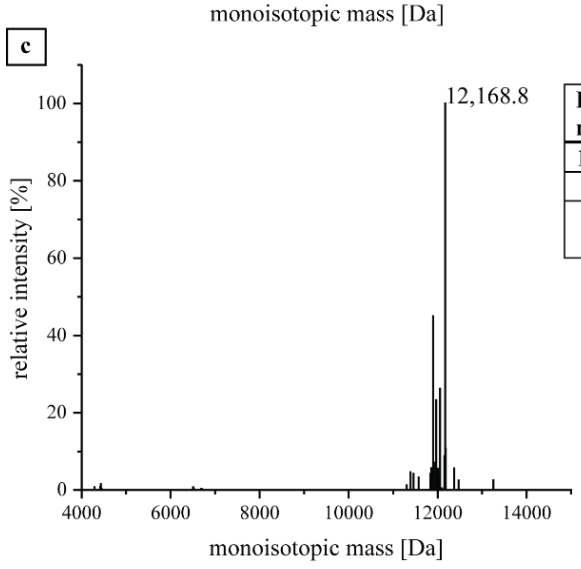
In these assignments, common PTMs such as disulfide bonds, pyroglutamate formation, and oxidation were considered. The number of considered PTMs was kept as low as possible to prevent overfitting the mass peaks. If the experimental and theoretical mass differed by 0.1 Da or less, the primary structure of the AL protein was considered correctly assigned. (Fig. 30).



Experimental mass	Theoretical mass	PTM	Error
11,754.7	11,756.7	-	2.0
	11,754.7	Disulfide bond	0.0



Experimental mass	Theoretical mass	PTM	Error
12,708	12,694	-	-14.0
	12,692	Disulfide bond	-16.0
	12,708	Disulfide bond Oxidation	0.0



Experimental mass	Theoretical mass	PTM	Error
12,168.8	12,187.9	-	19.1
	12,185.8	Disulfide bond	17.0
	12,168.8	Disulfide bond Pyroglutamate	0.0

Figure 30: Mass assignment of the cases FOR012, FOR140, and FOR159. a) Mass assignment for FOR012 revealing a disulfide bond. Experimental mass: assigned monoisotopic mass. Theoretical mass: calculated mass of the assigned amino acid sequence with and without PTMs. Error: discrepancy between experimental and theoretical mass. b) Mass assignment for FOR140 revealing a disulfide bond and oxidation. c) Mass assignment for FOR159 revealing a disulfide bond and pyroglutamate formation.

Resolving the primary structure of AL proteins by assigning a monoisotopic mass to a provided cDNA-derived sequence was only possible for sequences with a maximum of two sequence uncertainties. For this reason, 16 out of 60 cases with sequence grades 3 to 6 had to be excluded from the evaluation because they contained more than four sequence uncertainties (Fig. 31). Four (FOR120, FOR139, FOR149, FOR156) out of the remaining 44 cases with sequence grades 1 and 2 could also not be evaluated, as they contained more than two sequence uncertainties (Fig. 31). In 10 of the remaining 40 cases (FOR101, FOR102, FOR107, FOR124, FOR128, FOR137, FOR138, FOR142, FOR157, FOR159), the primary structure of the fibril proteins could be determined through the combination of the provided cDNA sequence (App. Tab. 17) with MS (App. Fig. 76, 77, 79, 80 – 83, 85, 87, 88) (Fig. 31).

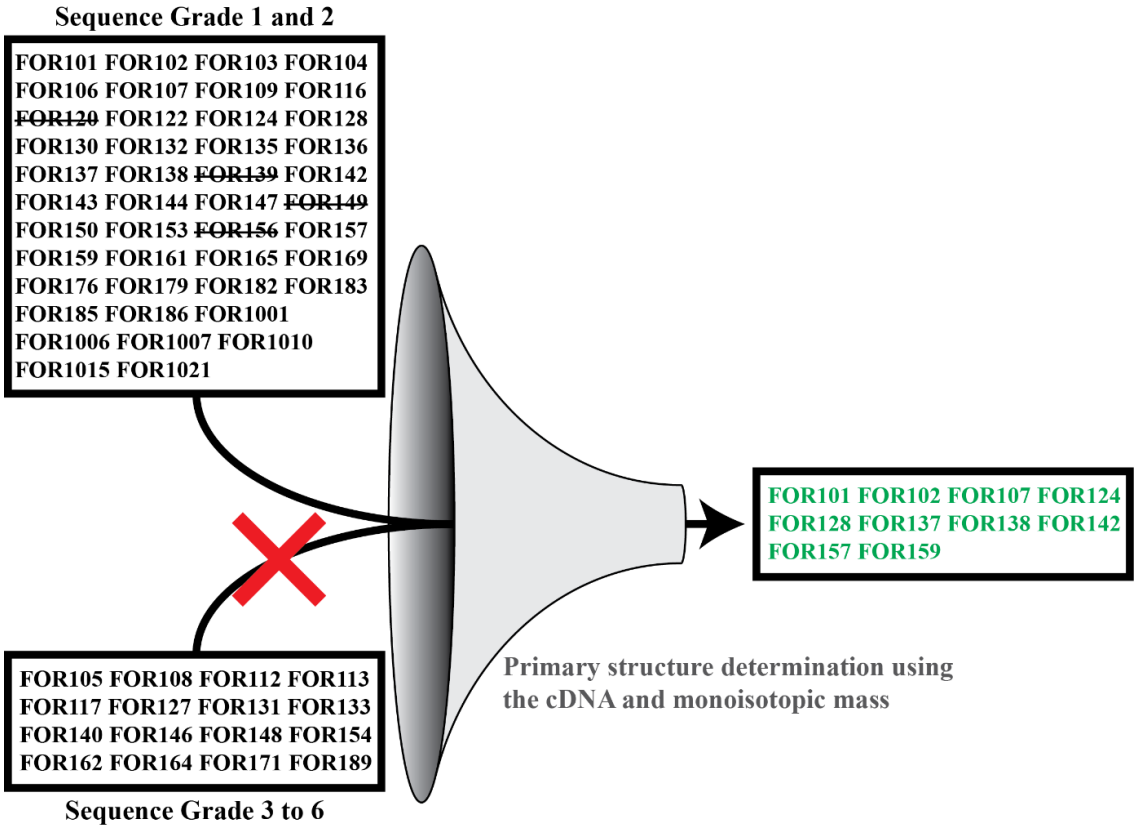


Figure 31: Schematic illustration of the primary structure determination from AL proteins with cDNA information using the provided cDNA-derived sequence and the monoisotopic mass. Red cross: Cases with sequence grades 3 to 6 could not be used for sequence assignments. Crossed-out cases: Cases with sequence grades 1 or 2 containing more than two sequence uncertainties not used for sequence assignments. Green cases: Successful sequence assignment and primary structure determination of AL proteins in cases with cDNA information using the cDNA-derived sequence from Heidelberg and the monoisotopic mass.

In four of these 10 cases (FOR101, FOR142, FOR157, FOR159), the provided cDNA-derived sequence was ambiguous and contained one or two positions where two amino acids were possible (App. Tab. 17). With the help of the monoisotopic mass it became clear which amino acid was correct. The only exception was the amino acid uncertainty at position 138 of FOR159, which is not within the sequence assigned to the monoisotopic mass. At this position, the GL amino acid alanine (Ala) was assumed. More cases were not suitable to determine the primary structure of AL proteins with this approach. The reasons were that the cDNA-derived sequences were incomplete or contained too many unidentified amino acids (App. Tab. 17), or the collected abdominal fat aspirate was too small or contained too little amyloid to extract amyloid fibrils.

4.2.2.2 Determination of the primary structure using the total mass and MS-based protein sequencing

Therefore, the next step was to become independent from the cDNA sequences and provide clear sequence information through MS-based protein sequencing of the extracted AL proteins. To that end, the lyophilized amyloid fibrils from water extraction were handed over to the Core Unit Mass Spectrometry and Proteomics at Ulm University to process the samples and acquire MS data. At the core unit, the amyloid fibrils were loaded on five wells of a polyacrylamide gel. If possible, at least 10 µg protein was loaded per well. The AL protein bands were excised, disulfide bonds were reduced, and Cys were alkylated. Then the AL proteins were digested in the gel by five proteases (Trypsin, Elastase, LysC, Chymotrypsin, and Pepsin). The resulting peptides were extracted from the gel and subjected to LC-MS. Peptides from each protease digest were analyzed separately with MS. For each protease digest, an HPLC chromatogram was acquired (Fig. 32 a), and approximately 14,000 mass spectra were recorded over the period of the sample elution (65 min) (Fig. 32 b). Analyzed peptides from the mass spectra were

fragmented by collision with N_2, and a fragment spectrum was recorded for the fragments of each peptide, resulting in one or more fragment spectra per mass spectrum (Fig. 32 c).

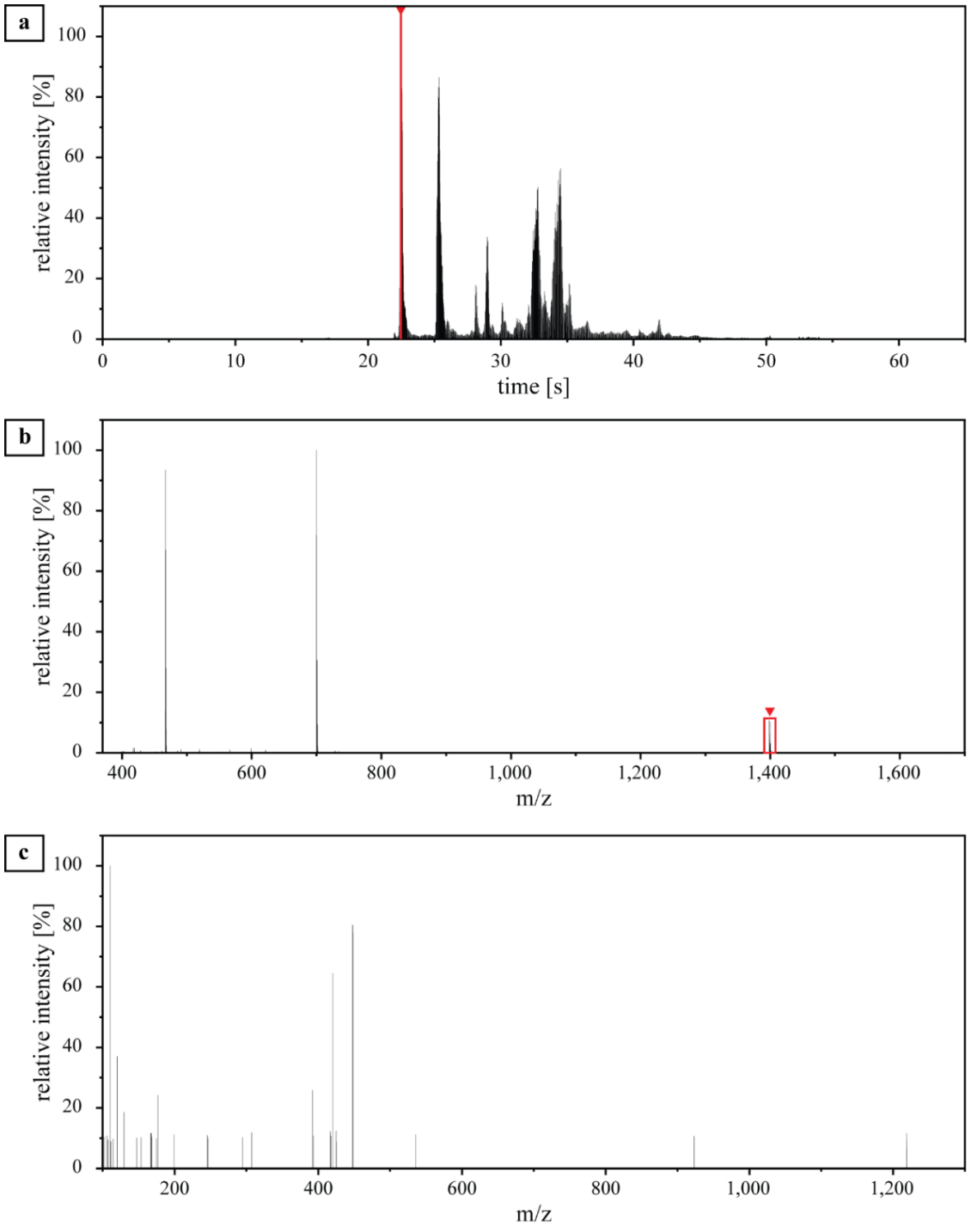


Figure 32: Illustration of the MS data used for MS-based protein sequencing exemplified for the trypsin-digested sample of FOR103. a) HPLC chromatogram of all eluted peptides analyzed by MS. b) Mass spectrum of peptides eluted at a specific retention time (red line in a). c) Fragment spectrum of one peptide of b) (red rectangle).

Spectra were forwarded to the Institute of Protein Biochemistry and evaluated using the Peaks AB 2.0 software [112] as part of this work. Peaks AB performs *de novo* sequencing, in other words, sequences proteins based on MS data. For this purpose, it analyzes the fragment spectra and compares the fragments of the individual peptides, considering the following aspects (Fig. 33 b). Peptide fragmentation involves the cleavage of a peptide bond, resulting in an N-terminal and C-terminal half of the peptide (Fig. 33 b). It is random which peptide bond is cleaved, resulting in multiple fragments that differ in their number of amino acids (Fig. 33 b). If two fragments of the same peptide differ by the mass of a particular amino acid, the larger fragment is defined as the smaller fragment plus the amino acid. By comparing all peptide fragments, Peaks AB identifies the amino acids that are missing from one fragment to another, revealing the amino acid sequence of a peptide (Fig. 33 b). When assigning a peptide sequence, Peaks AB also considers PTMs like carbamidomethylation as a consequence of Cys alkylation, oxidation, and pyroglutamate formation (Fig. 33 a). Peaks AB assembles the entire protein sequence by overlapping all sequenced peptides from all five protease digests using an internal algorithm (Fig. 33 a). The algorithm is specialized for antibodies and identifies the correct order of the *de novo* sequenced peptides based on the conserved regions of LC sequences. Depending on the number and quality of sequenced peptides that identified an amino acid at a given position, Peaks AB assigns a confidence score to each amino acid (Fig. 33 a).



Figure 33: MS-based protein sequencing results from Peaks AB exemplified for FOR103. a) Overlap of the sequenced peptides from trypsin (light blue), chymotrypsin (light green), elastase (yellow), pepsin (violet), and lysC (cyan) digest. Yellow C: carbamidomethylation. Above the peptide coverage (top to bottom): amino acid position, amino acid sequence, CDRs and FRs, confidence score per amino acid. Red letters: confidence score >95 %. Dark blue letters: confidence score >85 %. Bold black letters: confidence score <85 %. Black letters: unverified amino acid suggestion. b) Fragment spectrum of one peptide (red rectangle) with assigned N-terminal (blue b2 – b8) and C-terminal (red y1 – y10) fragments. Y-axis: relative intensity in percent (%). X-axis: m/z ratio. The peptide sequence is illustrated above the spectrum.

In most cases, however, the first analysis of Peaks AB did not provide the complete amino acid sequence of an AL protein. One reason was that Peaks AB sometimes identified only the amino acids that make up a peptide but not their order because too few peptide fragments were present. Another problem arose from regions without overlapping peptides, resulting in gaps within the reconstructed protein sequence. In these cases, Peaks AB provided a sequence from an internal database, which was not confirmed by protein sequencing and served as a gap filler (Fig. 33 a). The correct amino acids and their order were manually identified as follows. As a first step, gaps in the protein sequence of Peaks AB were filled with the corresponding amino acids of the cDNA-derived sequence from Heidelberg. Peaks AB compared the new parts of the sequence with the sequenced peptides. If peptides could be assigned to the new sequence so that the protein sequence was completely covered by peptides, the gaps were considered to be filled correctly. In cases where the protein sequence still contained gaps, its GL sequence was determined. For this purpose, the incomplete protein sequence of Peaks AB was blasted against the UniProtKB/Swiss-Prot database [115]. The GL segments with the lowest difference from the protein sequence were considered correct. Subsequently, the corresponding GL V-, J-, and C segments were combined into a gapless GL sequence, which was used for “sequence validation” in Peaks AB. This tool compares the sequenced peptides with a predefined sequence. Peaks AB uses more sequenced peptides for sequence validation than for determining the protein sequence. That is because assigning peptides to a predefined sequence requires only the identification of the amino acids that make up a peptide, but not their order. Thus, Peaks AB also uses peptides where too few fragments are available to identify the order of amino acids. Sequence validation identified differences between the GL sequence and the sequenced peptides. Each GL amino acid that differed was replaced with the corresponding amino acid from the sequenced peptides, and the resulting sequence was compared with the incomplete protein sequence of Peaks AB. Gaps within the protein sequence were filled with the corresponding amino acids of sequence validation. The gaps were considered to be filled

correctly if the protein sequence was completely covered with sequenced peptides of Peaks AB. If the protein sequence still contained gaps, amyloidogenic LCs of previous publications [22] or the Amyloid Light Chain Database were compared to the incomplete protein sequence from Peaks AB. Replacing amino acids in the protein sequence with amino acids from the previously published LC sequences could also improve the peptide coverage and fill gaps in the protein sequence. When all gaps were filled, the corresponding AL protein's total mass was assigned to the gapless protein sequence of Peaks AB using mMass [113]. The assignment was performed as described in section 4.2.2.1, replacing the cDNA-derived sequence from Heidelberg with the protein sequence from Peaks AB. The assignment was considered correct if the total mass matched the theoretical mass of the protein sequence with an error of 0.1 Da or less. MS-based protein sequencing could only identify the LC sequence covered by the AL protein. Consequently, protein sequencing did not provide the complete LC sequence because the AL proteins were shorter. For all analyses where the complete precursor LC was necessary, the missing part of the LCs was considered GL.

MS-based protein sequencing was performed in the 50 cases in which the total mass of the AL proteins could not be assigned to the cDNA-derived sequence from Heidelberg, provided that a substantial amount of AL protein was extracted. The 50 cases included the remaining 34 cases with sequence grades 1 and 2 and the 16 cases with sequence grades 3 to 6 (Fig. 34). In six of the 50 cases (FOR103, FOR140, FOR147, FOR161, FOR1001, FOR1007), the MS-based protein sequence (App. Fig. 95 – 100) combined with the total mass (App. Fig. 78, 84, 86, 89 – 91) could identify the primary structure of the AL proteins. (Fig. 34). Together with the cases in which the primary structure of the AL proteins was determined using the cDNA-derived sequence from Heidelberg, this results in a total of 16 out of 60 cases with resolved primary structure (Fig. 34).

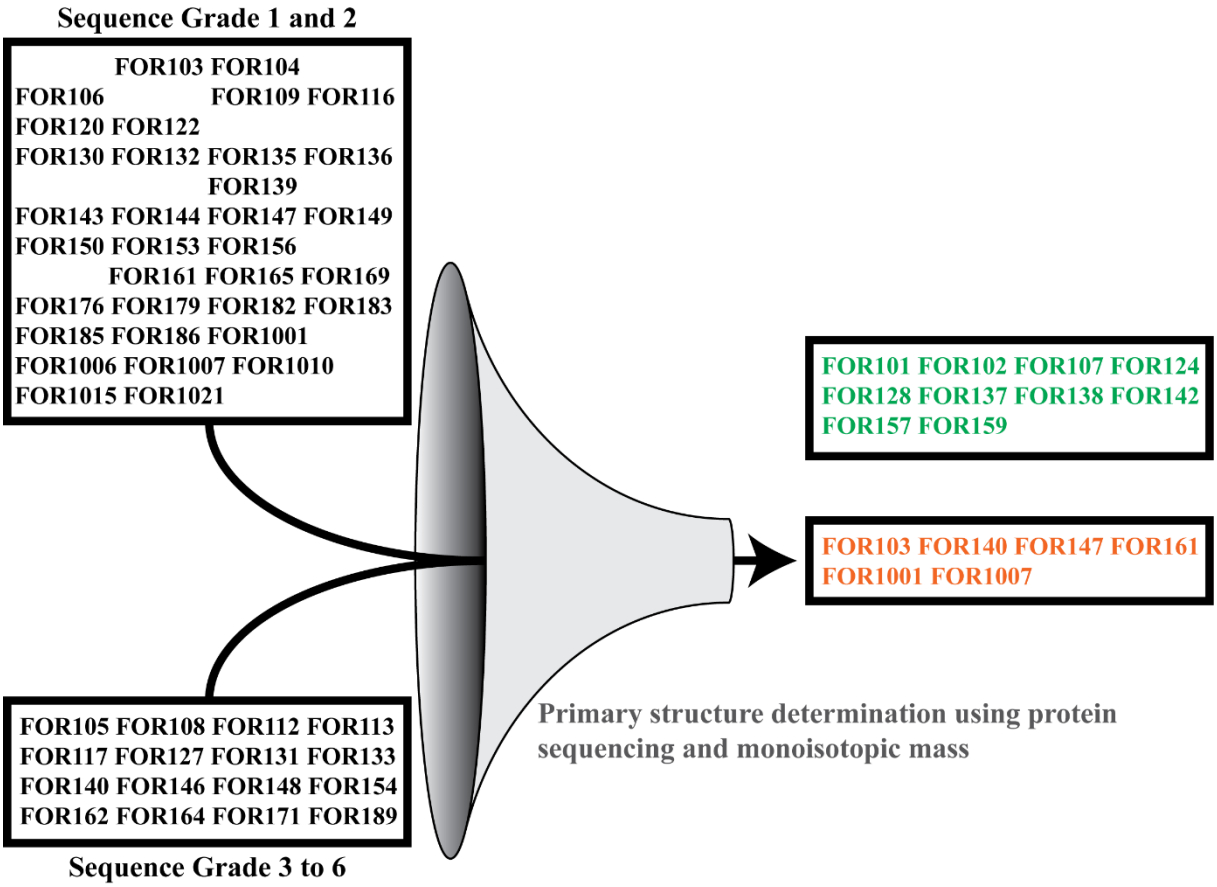


Figure 34: Schematic illustration of the primary structure determination from AL proteins with cDNA information using the MS-based protein sequence and the monoisotopic mass. Green cases: Successful sequence assignment and primary structure determination of AL proteins from cases with cDNA information using the cDNA-derived sequence from Heidelberg and the monoisotopic mass. Orange: Successful sequence assignment and primary structure determination of AL proteins in cases with cDNA information using the MS-based protein sequence and the monoisotopic mass.

A limiting factor of MS-based protein sequencing was the availability of AL protein because it required a fifty-fold amount compared to the total mass analysis. Consequently, many cases had to be excluded from this work because the abdominal fat aspirate was too small or contained too little AL protein to assemble a gap-free protein sequence.

4.2.2.3 Application of MS-based protein sequencing to cases without cDNA sequence

Until now, primary structure analysis was limited to cases where cDNA sequences were available. However, with the advent of MS-based protein sequencing, cDNA information was no longer necessary. Therefore, heart biopsies from four cases (FOR007, FOR009, FOR010, FOR012) without cDNA sequence were used to determine the primary structure of the AL proteins by MS-based protein sequencing and total mass analysis. These four cases expanded the total number of analyzed cases from 60 to 64.

Fibril extraction was performed according to the standard water extraction protocol [63]. MS data for total mass analysis and protein sequencing was acquired in the Core Unit Mass Spectrometry and Proteomics, as described in sections 4.2.2.1 and 4.2.2.2. Total mass analysis and protein sequencing were part of this work (App. Fig. 73 – 75, 92 – 94), were conducted as described in sections 4.2.2.1 and 4.2.2.2, and could resolve the primary structure of the AL proteins in three (FOR009, FOR010, FOR012) of four cases (Fig. 35). The parts of the LC sequence that could not be identified by MS-based protein sequencing because the AL proteins did not cover the entire LC were considered GL.

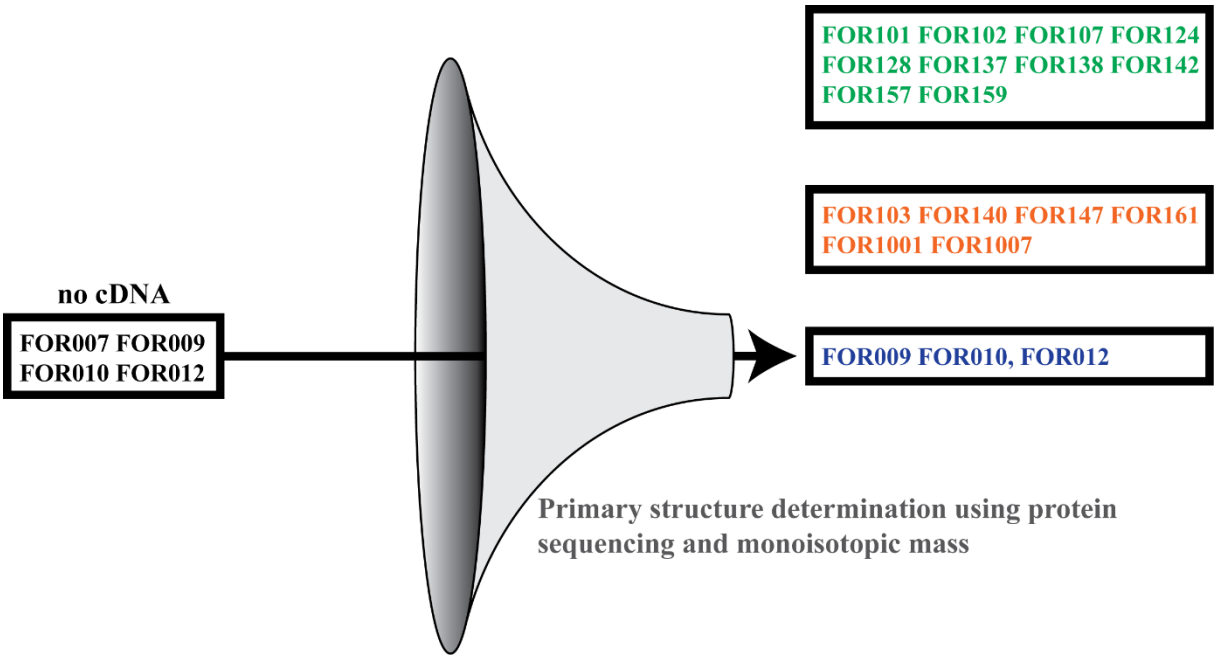


Figure 35: Schematic illustration of the primary structure determination from AL proteins without cDNA information using only the MS-based protein sequence and the monoisotopic mass. Green cases: Successful sequence assignment and primary structure determination of AL proteins in cases with cDNA information using the cDNA-derived sequence from Heidelberg and the monoisotopic mass. Orange: Successful sequence assignment and primary structure determination of AL proteins in cases with cDNA information using the MS-based protein sequence and the monoisotopic mass. Blue: Successful sequence assignment and primary structure determination of AL proteins in cases without cDNA information using the MS-based protein sequence and the monoisotopic mass.

In summary, MS data could resolve the primary structure of AL proteins in 19 of 64 cases. In 10 of the 19 cases, the cDNA and total mass were sufficient to determine the primary structure of the AL proteins (Fig. 36, black cases). In the remaining nine of 19 cases (Fig. 36, white cases), MS-based protein sequencing was necessary because the cDNA information was incomplete (Fig. 36, dashed ellipse) or none was available (Fig. 36, gray background). For 16 of the 19 cases, cDNA sequencing was performed in Heidelberg. 15 of the 16 cases had a sequence grade of 1 or 2, and one of the 16 cases had a sequence grade of 3 to 6 (Fig. 36, green and blue background). In 13 of the 16 cases, the amyloid fibrils were extracted from abdominal fat, whereas in three of the 16 cases, fibrils were extracted from heart tissue (Fig. 36, red and black circle). The remaining three of the 19 cases had no cDNA sequence from Heidelberg, and the fibrils were extracted from heart tissue (Fig. 36, gray background and black circle). Fibril

extraction from heart tissue yielded more AL protein than extraction from abdominal fat, resulting in a better success rate of protein sequencing and determining the primary structure of AL proteins.

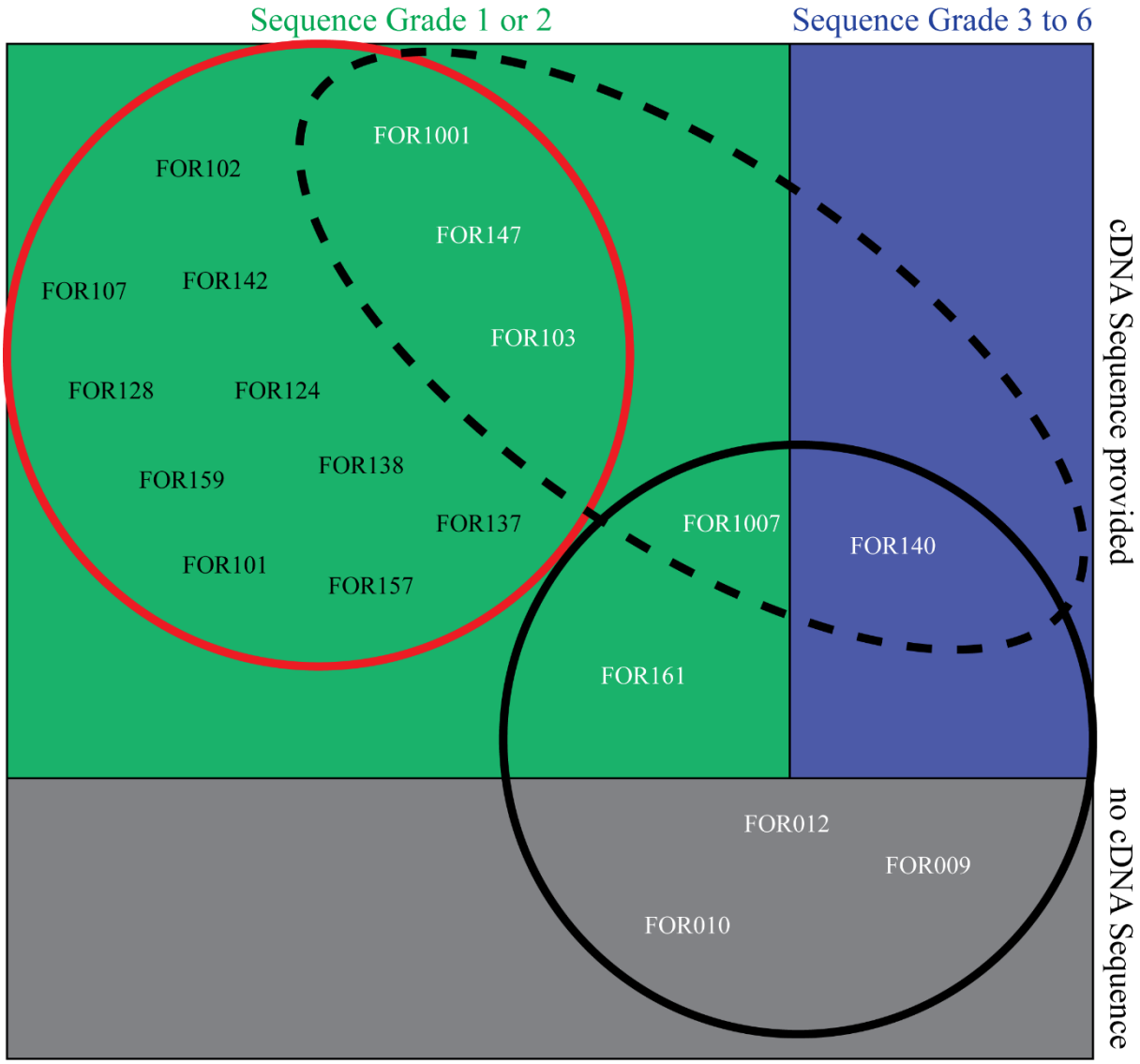


Figure 36: Classification of the 19 cases with the identified primary structure of the AL proteins. White cases: Primary structure determination using MS-based protein sequencing. Black cases: Primary structure determination using the cDNA-derived sequence. Top: cases with cDNA sequence from Heidelberg University Hospital. Bottom: Cases without cDNA Sequence. Green: Cases with sequence grade 1 or 2. Blue: Cases with Sequence grades 3 to 6. Gray: No cDNA sequence. Dashed ellipse: Cases with missing N-terminal amino acids in the cDNA-derived sequence. Red circle: Cases with AL protein from abdominal fat. Black circle: Cases with AL protein from the heart.

4.3 Determination and analysis of AL proteins' primary structures to identify parameters determining amyloidogenicity

4.3.1 Characterization of the cases whose primary structure could be determined

The 19 patients whose AL protein primary structures could be determined included nine women and ten men between 34 and 80 years of age at the time of tissue sampling (Tab. 10). 15 patients suffered from a prominent cardiac involvement, three from a strong renal involvement, and one showed a combination of cardiac and renal involvement (Tab. 10).

Table 10: Clinical information on the patients whose primary structure of the AL proteins could be determined. Sex, age at the time of tissue sampling, and organ involvement are displayed. The information was provided by Ute Hegenbart (Heidelberg University Hospital).

Case	Sex	Age	Organ involvement
FOR009	51	male	Heart
FOR010	64	male	Heart
FOR012	56	male	Heart
FOR101	72	male	Heart
FOR102	34	female	Heart
FOR103	80	female	Heart
FOR107	55	male	Kidney
FOR124	72	female	Heart
FOR128	58	male	Heart
FOR137	49	female	Kidney
FOR138	70	male	Heart
FOR140	57	female	Heart
FOR142	50	male	Kidney
FOR147	54	male	Heart
FOR157	71	male	Heart
FOR159	77	female	Heart
FOR161	62	female	Heart
FOR1001	53	female	Heart/kidney
FOR1007	56	female	Heart

4.3.1.1 Amyloid fibril extraction in the cases with resolved primary structures

In six cases, amyloid fibrils were extracted from 250 mg cardiac tissue using the standard water extraction procedure [63]. In the remaining 13 cases, extraction was performed from abdominal fat aspirates using the herein-established water extraction (see section 4.2.1 for details). The

total amount of provided abdominal fat was used for fibril extraction and varied from case to case between 20 mg and 280 mg (Tab. 11).

Table 11: Tissue information of the 19 patients whose primary structure of the AL proteins could be determined. Sample tissue, amount of tissue used for extraction, and amyloid load are displayed. No amyloid score was determined for cases with cardiac tissue. The information was provided by Ute Hegenbart (Heidelberg University Hospital).

Case	Sample tissue	Amount of tissue	Amyloid load
FOR009	Heart	250 mg	(no score)
FOR010	Heart	250 mg	(no score)
FOR012	Heart	250 mg	(no score)
FOR101	Fat	60 mg	4+
FOR102	Fat	35 mg	3+
FOR103	Fat	110 mg	4+
FOR107	Fat	250 mg	3+
FOR124	Fat	100 mg	3+
FOR128	Fat	35 mg	4+
FOR137	Fat	100 mg	4+
FOR138	Fat	20 mg	3+
FOR140	Heart	250 mg	(no score)
FOR142	Fat	60 mg	3+
FOR147	Fat	50 mg	3+
FOR157	Fat	40 mg	4+
FOR159	Fat	190 mg	4+
FOR161	Heart	250 mg	(no score)
FOR1001	Fat	280 mg	3+
FOR1007	Heart	250 mg	(no score)

Each extraction was analyzed by LDS-PAGE and negative stain TEM (Fig 37, App. Fig. 73 – 91). LDS-PAGE of all 19 fibril extractions showed an enriched protein in the water extracts with a molecular weight of approximately 15 kDa (Fig 37 a – c, App. Fig. 73 – 91). However, the intensity of the corresponding protein band differed between the 19 cases, with two cases (FOR102, FOR124) showing only a weak signal (Fig 37 b, App. Fig. 77, 80). In addition to the proteins with a molecular weight of 15 kDa, the water extracts of five cases (FOR137, FOR140, FOR159, FOR161, FOR1007) contained other enriched proteins below the 15 kDa marker band resulting in a total of two or three protein bands (Fig. 37 c, App. Fig. 82, 84, 88, 89, 91). LDS-PAGE provided no evidence on whether the smaller protein bands corresponded to fragments of the protein with a molecular weight of 15 kDa. Negative stain

TEM showed amyloid fibrils in the water extracts of all 19 cases, confirming the successful extraction of intact amyloid fibrils and that the protein band at 15 kDa corresponded to the AL proteins (Fig 37, App. Fig. 73 – 91). Due to contamination, the amyloid fibrils could be characterized in only 17 out of 19 cases (Fig. 37 d – f, App. Fig. 73 – 80, 82, 84 – 91). 14 cases (FOR010, FOR012, FOR101, FOR102, FOR103, FOR107, FOR137, FOR140, FOR142, FOR147, FOR157, FOR161, FOR1001, FOR1007) showed short amyloid fibrils (Fig. 37 d), and three cases (FOR009, FOR124, FOR159) long amyloid fibrils without visible crossovers (Fig. 37 e). In two cases with long (FOR009, FOR124) and five cases with short amyloid fibrils (FOR010, FOR012, FOR142, FOR161, FOR1001), multiple fibrils were wrapped around each other (Fig. 37 f).

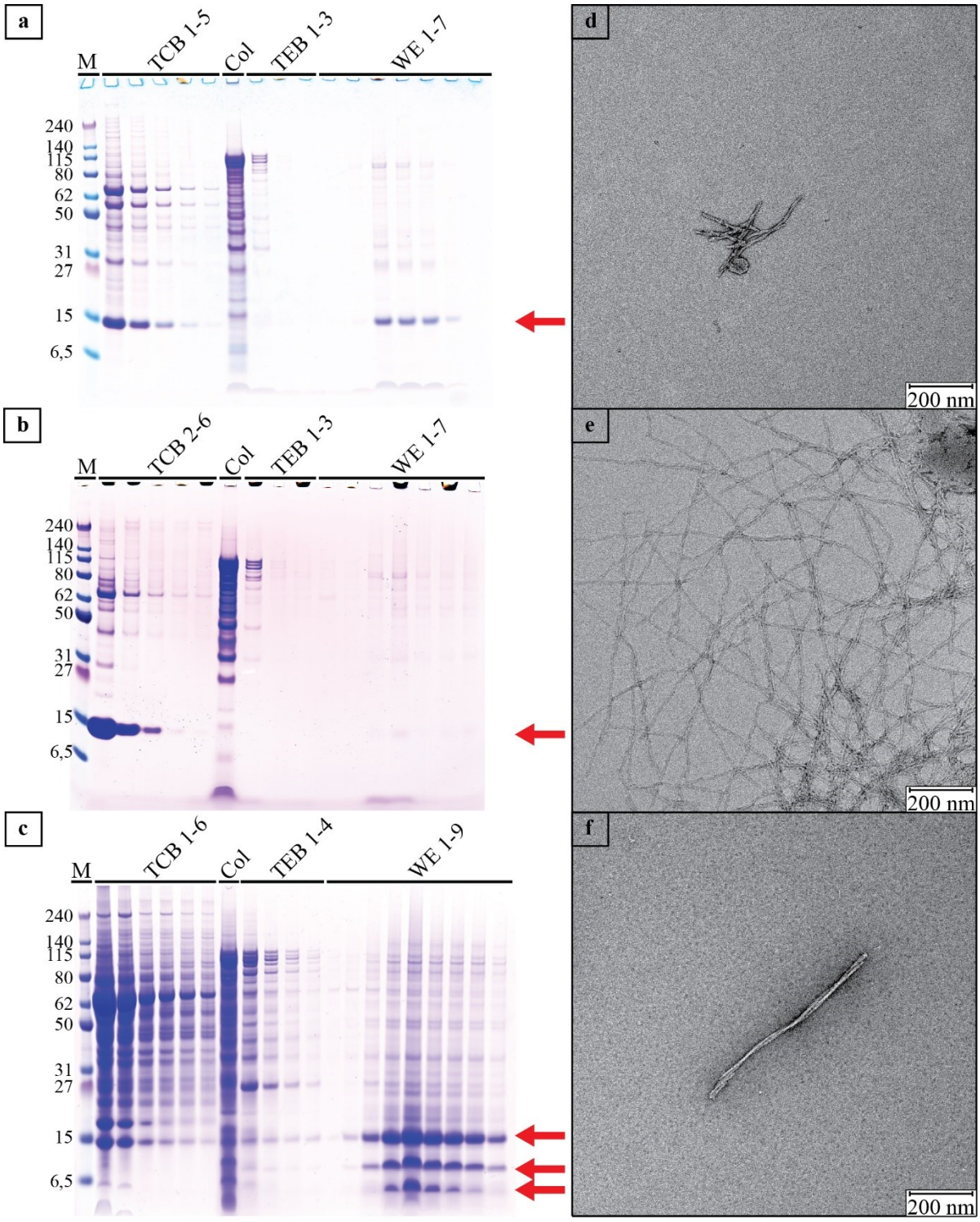


Figure 37: Examples of different AL protein bands and fibril morphologies in the water extracts of the 19 cases whose AL protein primary structures could be determined. a) LDS-PAGE of water extraction with one AL protein band at 15 kDa in the water extracts (FOR101). M: marker in kDa. Col: supernatant after collagenase incubation. WE: water extract b) LDS-PAGE of water extraction with one weak AL protein band at 15 kDa in the water extracts (FOR124). c) LDS-PAGE of water extraction with multiple protein bands in the water extracts (FOR140). d) Negative stain TEM image of short amyloid fibrils (FOR140). e) Negative stain TEM image of long amyloid fibrils (FOR159). f) Negative stain TEM image of two amyloid fibrils wrapped around each other (FOR124).

In summary, the amyloid fibril extraction differed between the cases regarding the band intensity of the AL proteins and the number of enriched proteins in the water extracts. Negative stain TEM confirmed the success of fibril extraction and provided information on the fibril length. No correlation was observed between the patient parameters and differences in the number of protein bands in LDS-PAGE or fibril morphology.

4.3.1.2 The primary structure of the AL proteins from the 19 cases

The corresponding AL proteins of the extracted amyloid fibrils were analyzed by MS (App. Fig. 73 – 91). Consistent with previous studies [14,127], all identified AL proteins had a molecular weight of 10 to 15 kDa (App. Fig. 73 – 91). Assigning a monoisotopic mass to a specific LC fragment under consideration of common PTMs identified the primary structure of the fibril-forming AL proteins. In 18 of 19 cases, the molecular mass with the highest intensity could be assigned to a precursor LC fragment (Fig. 38).

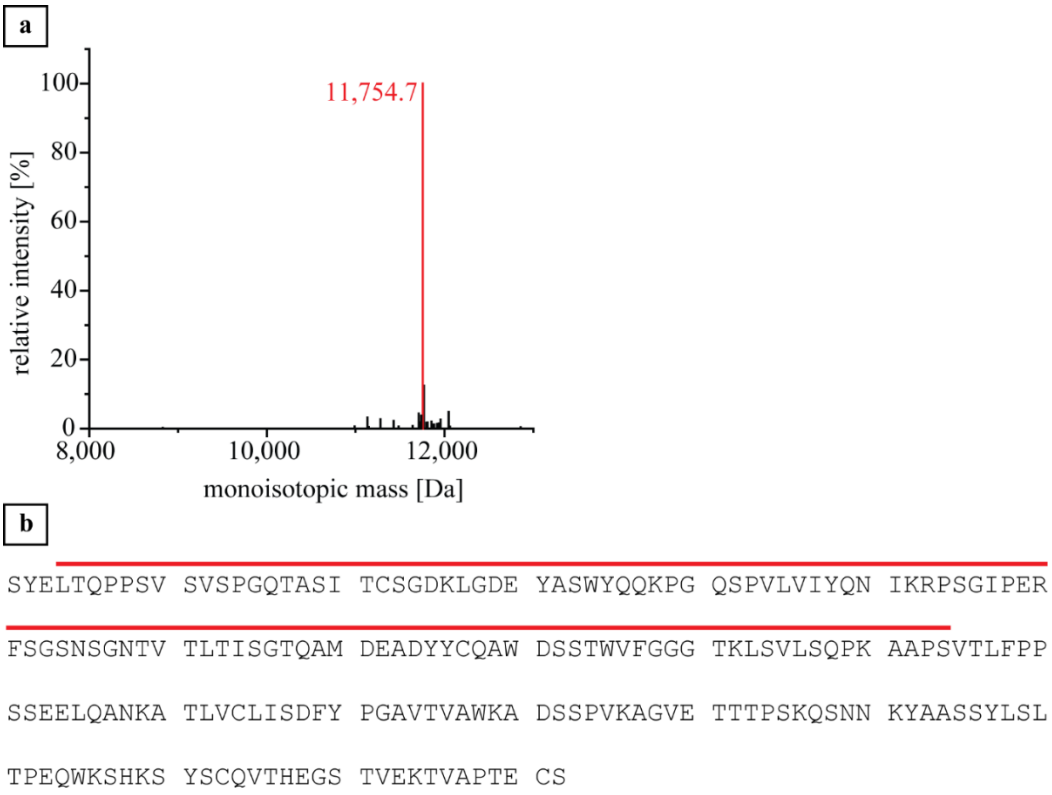


Figure 38: Assignment of the most intense monoisotopic mass to a precursor LC fragment at the example of FOR012. a) Deconvoluted mass spectrum. Red number: assigned monoisotopic mass. Red peak: peak of the assigned mass. b) Precursor LC sequence, determined by MS-based protein sequencing. Red line: Precursor LC fragment that could be assigned to the monoisotopic mass.

Only in one case (FOR161) (Fig. 39, App. Fig. 89) the most intense peak did not match a LC fragment. Instead, the second and third most intense peaks could be assigned (Fig. 39). These peaks had a side peak at a distance of 18 Da, one being the most intense peak (Fig. 39), which indicates a PTM that shifts the molecular weight of the AL proteins by 18 Da. However, no such PTM was found, and the only conjecture was that some AL proteins were hydrolyzed and the resulting fragments remained connected by a disulfide bond.

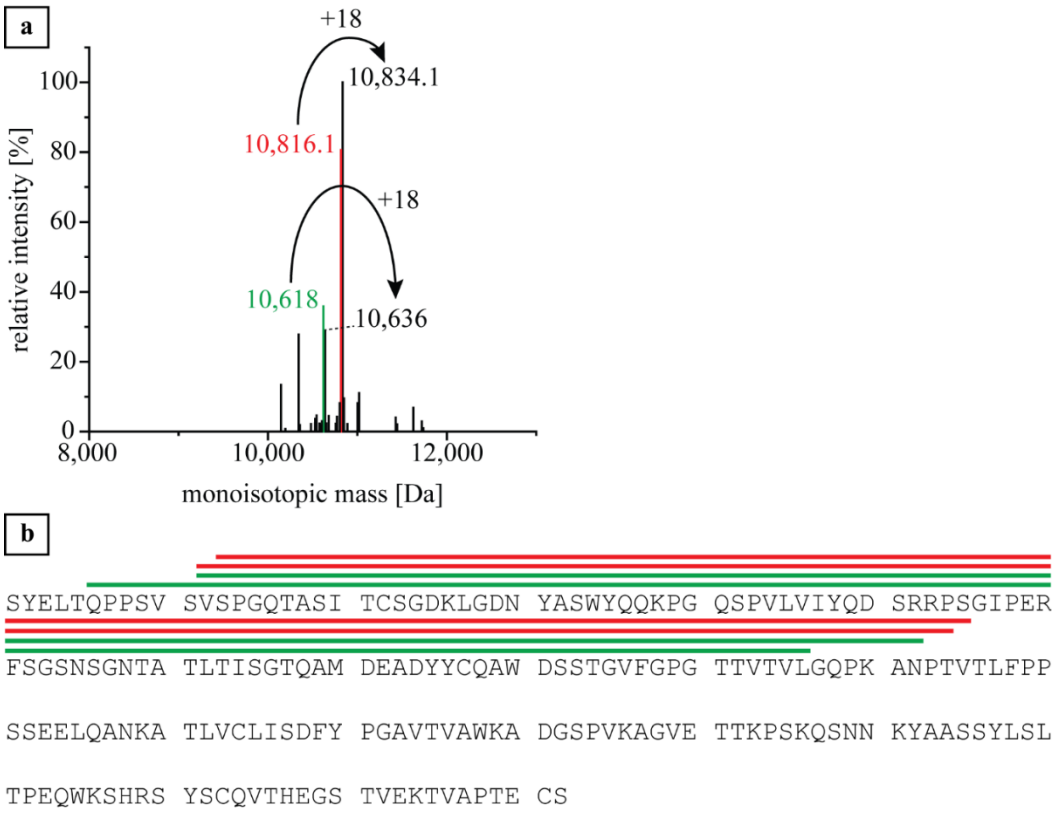


Figure 39: Mass assignment of case FOR161, whose most intense monoisotopic mass could not be assigned. a) Deconvoluted mass spectrum. Red number: assigned monoisotopic mass. Red peak: peak of the assigned red mass. Green number: assigned monoisotopic mass. Green peak: peak of the assigned green mass. Black numbers: non-assignable masses. Black peaks: non-assignable peaks. +18: mass difference in Da. b) Precursor LC sequence, determined by MS-based protein sequencing. Red line: Precursor LC fragment that could be assigned to the red monoisotopic mass. Green line: Precursor LC fragment that could be assigned to the green monoisotopic mass.

To compare the AL proteins of the different cases concerning their length and N- and C-terminal truncations, the LC fragments were aligned, and the amino acid positions were labeled according to consecutive numbering (Fig. 40).

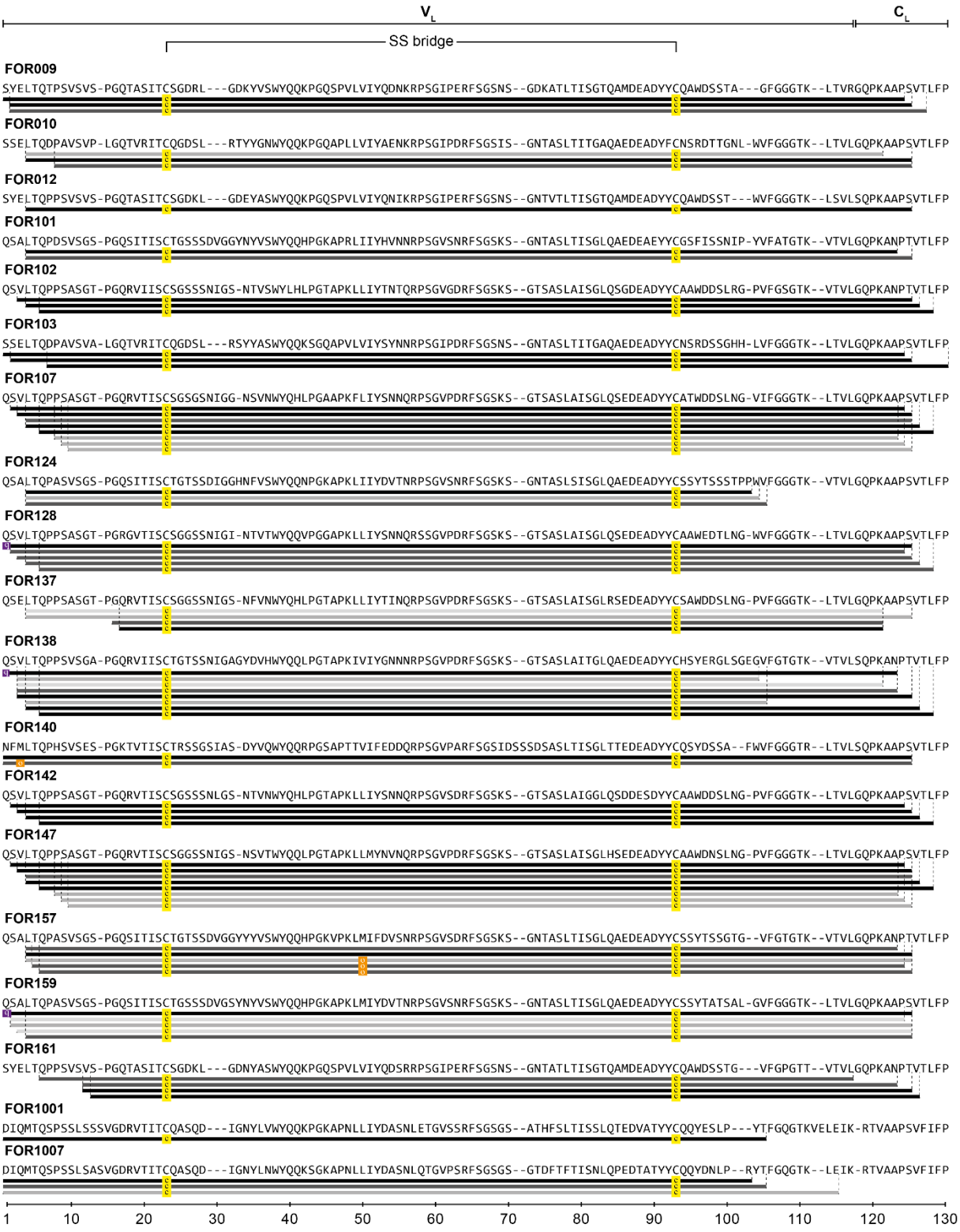


Figure 40: Alignment of the AL proteins' primary structures from the 19 different patients. Yellow c: disulfide bond. Violet q: pyroglutamate formation. Orange o: Met oxidation. Grayscale: Indicates LC fragments from the same mass peak; Dark gray: high intensity of the peak; light gray: low intensity of the peak.

In five out of 19 cases (FOR012, FOR102, FOR103, FOR142, FOR1001), only one monoisotopic mass could be assigned to the corresponding LC sequence (Fig. 40, Tab 12). However, in three of these five cases (FOR102, FOR103, FOR142), the monoisotopic mass matched multiple LC fragments (Fig. 40, Tab. 12), as amino acids with the same molecular mass were present at the N-terminus and C-terminus of the fragments. In the alignment of the AL proteins' primary structures, AL proteins identified from one monoisotopic mass are labeled in the same shade of gray (Fig. 40). In five of 19 cases (FOR009, FOR101, FOR128, FOR140, FOR161), two monoisotopic masses could be assigned to LC fragments (Fig. 40, Tab. 12). In six out of 19 cases (FOR010, FOR107, FOR124, FOR147, FOR157, FOR1007), three monoisotopic masses could be assigned, and in three out of 19 cases (FOR137, FOR138, FOR159), four monoisotopic masses could be assigned (Fig. 40, Tab. 12). Finding multiple AL proteins in 17 out of 19 cases suggests that the amyloid fibrils of one patient are composed of AL proteins with heterogeneous N- and C-termini and different PTMs. Adding up all assigned LC fragments from the 19 cases resulted in 75 AL proteins whose primary structure could be determined. In summary, the alignment revealed commonalities between the AL proteins concerning N- and C-terminal truncations and PTMs, which were investigated in more detail.

Table 12: Information on the assignment of the monoisotopic masses to the precursor LCs in the 19 cases. The table displays the experimental mass of the AL protein, the theoretical mass of the assigned LC fragment, the start and end of the assigned LC fragments according to the consecutive numbering of figure 40, and the PTMs of the AL proteins. MW: molecular weight in Da. SS: disulfide bond. pQ: pyroglutamate formation. ox: oxidation. AL proteins identified from one monoisotopic mass are displayed in the same row.

Case	MW [Da]		assignable LC fragment(s)			PTMs		
	experimental	theoretical				SS	pQ	ox
FOR009	11,977.8	11977.8	1–124	2–125		X		
	12,177.9	12177.9	2–127			X		
FOR010	11,587.8	11587.8	8–125			X		
	11,718.9	11718.8	4–121			X		
	12,045.0	12045.0	4–125			X		
FOR012	11,754.7	11754.7	4–125			X		
FOR101	12,002.9	12,002.8	4–123			X		
	12,200.9	12,200.9	4–125			X		
FOR102	11,852.9	11,852.9	3–125	4–126	6–128	X		
FOR103	12,183.0	12,182.9	1–124	2–125		X		
		12,182.9	7–130			X		

FOR107	11,132.4	11,132.4	8–123	9–124	10–125		X		
	11,755.7	11,755.7	4–125				X		
	11,854.8	11,854.8	2–124	3–125	4–126	6–128	X		
FOR124	10,047.8	10047.7	4–103				X		
	10,233.9	10233.8	4–104				X		
	10,332.9	10332.9	4–105				X		
FOR128	11,830.8	11,830.8	2–124	3–125	4–126	6–128	X		
	12,028.9	12,028.9	1–125				X	X	
FOR137	10,433.1	10433.1	17–121				X		
	10,490.2	10490.1	16–121				X		
	11,526.7	11526.7	4–121				X		
	11,852.9	11852.9	4–125				X		
FOR138	10,312.0	10,312.0	3–104	4–105			X		
	11,854.9	11,854.8	3–121				X		
	12,039.9	12,039.9	3–123				X		
	12,238.0	12,238.0	1–123				X	X	
		12,238.0	3–125	4–126	6–128		X		
FOR140	12,692.0	12692.0	1–125				X		
	12,708.0	12708.0	1–125				X		X
FOR142	11,840.8	11,840.7	2–124	3–125	4–126	6–128	X		
FOR147	11,128.5	11128.4	8–123	9–124	10–125		X		
	11,751.8	11751.7	4–125				X		
	11,850.9	11850.8	2–124	3–125	4–126	6–128	X		
FOR157	11,709.6	11709.6	4–123				X		
		11709.5	5–124	6–125			X		X
	11,907.7	11907.7	4–125				X		
	11,923.7	11923.7	4–125				X		X
FOR159	11,899.8	11,899.7	4–125				X		
	11,970.8	11,970.8	2–124	3–125			X		
	12,057.8	12,057.8	2–125				X		
	12,168.8	12,168.8	1–125				X	X	
FOR161	10,618.0	10618.0	6–117	12–123			X		
	10,816.1	10816.1	12–125	13–126			X		
FOR1001	10,577.1	10577.0	1–105				X		
FOR1007	10,517.0	10516.9	1–103				X		
	10,781.1	10781.1	1–105				X		
	11,641.5	11641.5	1–115				X		

4.3.1.2.1 N-terminal variability of the AL proteins

64 of 75 AL proteins showed N-terminal truncations. 53 AL proteins started within the first four amino acids of the precursor LC, whereas 22 AL proteins started between amino acid positions 5 to 17 (Fig. 41, Tab. 12). Most AL proteins started at amino acid position four (Fig. 41, Tab. 12). In contrast to the N-terminally truncated AL proteins, 11 LC fragments from

eight cases (FOR009, FOR103, FOR128, FOR138, FOR140, FOR159, FOR1001, FOR1007) had an intact N-terminus (Fig. 41, Tab. 12). Pyroglutamate formation was observed in all AL proteins with intact N-terminus when the precursor LC started with Gln. FOR140, FOR1001, and FOR1007 showed only AL proteins with intact N-terminus (Fig. 40, Tab. 12).

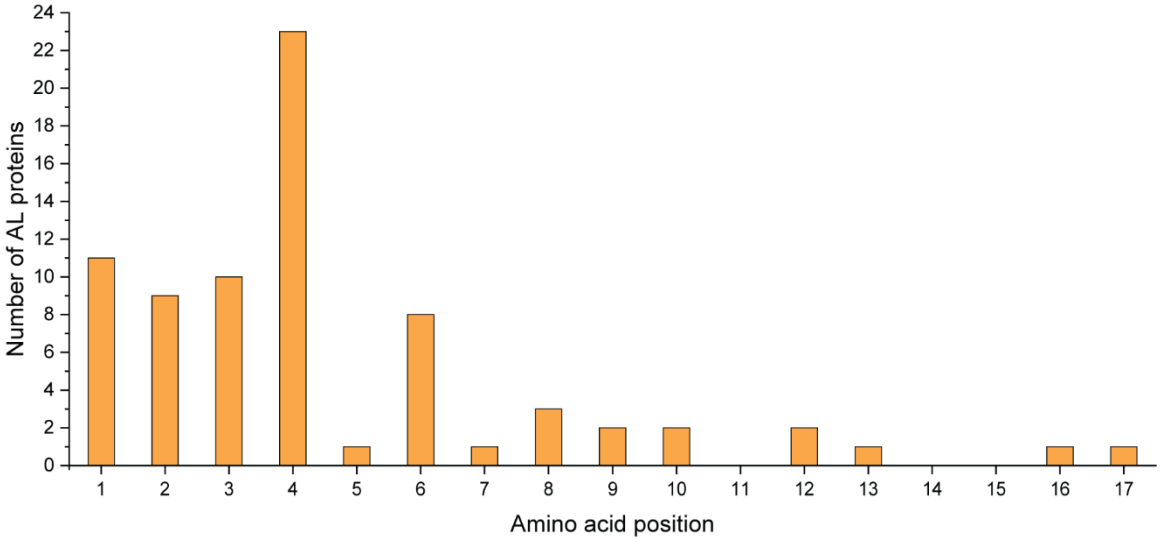


Figure 41: Plot of the N-terminal start point of the identified AL proteins (n=75). Y-axis: number of AL proteins starting at a given position. X-axis: N-terminal amino acid position of the AL proteins according to the consecutive amino acid numbering in figure 40.

4.3.1.2.2 C-terminal variability of the AL proteins

Consistent with previous studies of AL amyloidosis [14,65,66,135], 66 of the 75 identified LC fragments could be assigned to the V_L domain and a small part of the C_L domain (Fig. 42, Tab. 12). 67 of 75 AL proteins terminated between amino acids 115 and 130 (Fig. 42, Tab 12) and 8 of 75 terminated earlier between amino acids 103 and 105 (Fig. 42, Tab 12). 37 % of the AL proteins terminated at position 125 (Fig. 42, Tab. 12), suggesting a ubiquitous truncation motif in different patients.

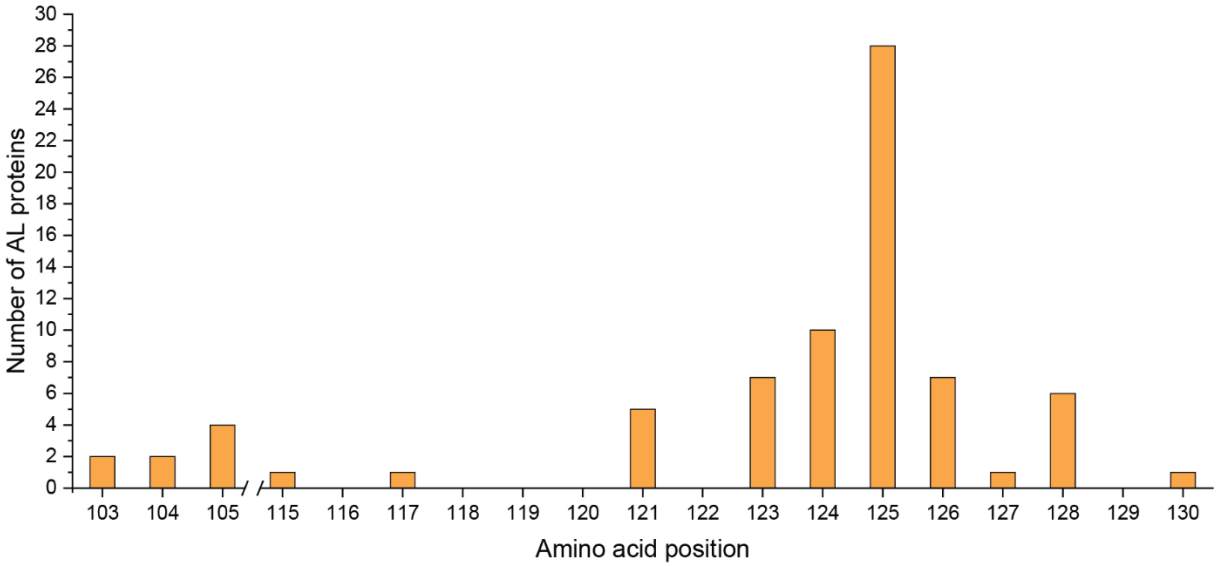


Figure 42: Plot of the C-terminal end of the identified AL proteins (n=75). Y-axis: number of AL proteins terminating at a given position. X-axis: terminal amino acid position of the AL protein according to the consecutive amino acid numbering in figure 40.

Considering the three-dimensional structure of a natively folded precursor LC, most AL proteins terminated in the N-terminal β-strand of the C_L domain, including amino acids 125 to 130 (Fig. 43). Other AL proteins terminated in the junction between the V_L domain and C_L domain at positions 115 to 124 (Fig. 43). That confined truncation region suggests that endoproteases can bind the compactly folded precursor LCs only at the exposed N-terminus of the C_L domain or that a specific protease binds a distinct motif in the C_L domain. However, searching for proteases in the PeptideCutter of Expasy [117] that cut exclusively in the region between amino acid positions 115 to 130 yielded no hits. The fragments that terminated at positions 103 to 105 did not fit the pattern of the other AL proteins. These shorter fragments ended in a C-terminal β-strand of the natively folded V_L domain (Fig. 43), opposing the usual truncation site. The fragments were found in only four cases (FOR124, FOR138, FOR1001, FOR1007) (Fig. 40). Interestingly, all AL proteins of FOR124, FOR1001, and FOR1007 terminated before the C_L domain, suggesting that certain properties of the LCs or amyloid fibrils affect the truncation site of AL proteins.

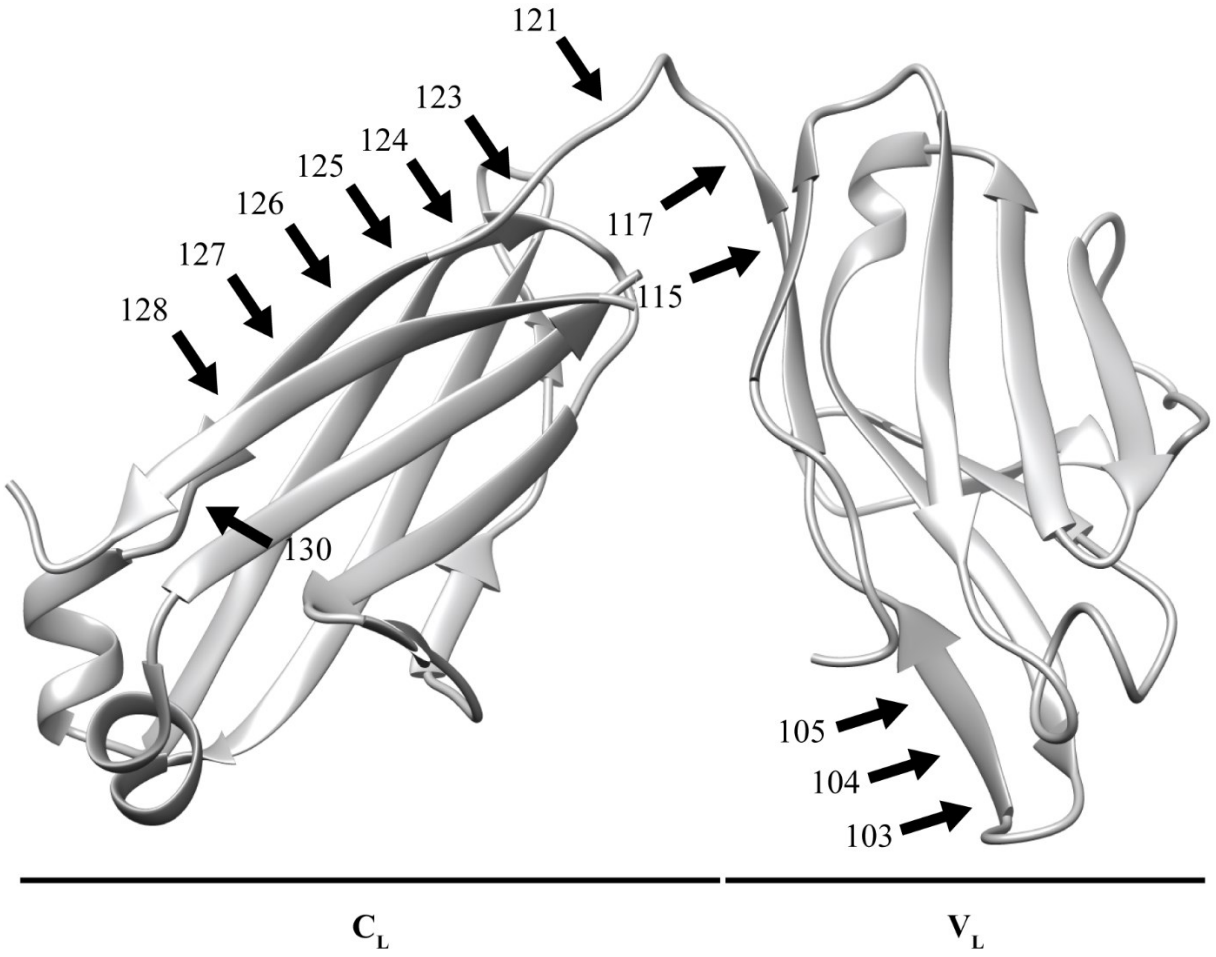


Figure 43: C-terminal truncation sites in the native 3D structure of an IGLV1-44-derived LC (PDB: 6QB6-L) [136]. Black arrows: truncation site with amino acid position according to the consecutive numbering in figure 40. Gray line: unstructured loop. Gray arrow: β-strand. Gray twist without arrowhead: α-helix. Position 117: C-terminus of the V_L domain.

4.3.1.2.3 Disulfide bonds and other PTMs

The assignment of monoisotopic masses to LC fragments revealed that all AL proteins in this work contained the disulfide bond of a natively folded V_L domain. The conservation of the disulfide bond suggests a universal mechanism where the unfolding of the LCs and fibril formation occurs around the disulfide bond. In addition to the disulfide bond, two other PTMs were found in the AL proteins. Pyroglutamate formation of the N-terminal Gln of AL proteins was found in three cases (FOR128, FOR138, FOR159) (Fig. 40, Tab. 12). There is no evidence that a pyroglutamate is characteristic of amyloidogenic LCs, as it is a common PTM of naturally

occurring antibodies [137]. Oxidation of the AL proteins was observed in two cases (FOR140, FOR157) (Fig. 40, Tab. 12). Since MS did not provide information on the modified amino acid, the oxidation of Met was suggested because it is one of the most readily oxidized amino acids in proteins [138]. A previous study showed that ESI causes Met oxidation in Aβ [139], suggesting that the rare oxidation of AL proteins may be an artifact of ionization. Although previous studies showed that LC glycosylation is increased in AL amyloidosis patients [19,61], MS did not provide evidence for the glycosylation of AL proteins. The small amount of provided abdominal fat and extracted AL protein prevented further experiments on N-glycosylation.

In summary, the MS data showed similarities between the different AL proteins in terms of molecular weight, truncation sites, and other PTMs. The MS results suggest a ubiquitous C-terminal truncation region of AL proteins with slight heterogeneity at the N- and C-terminus. There is evidence that an intact N-terminus and a C-terminal truncation in the V_L domain occur only rarely and depend on certain characteristics of a LC. Furthermore, the mass assignments showed that a disulfide bond is present in all AL proteins, suggesting a fibril formation mechanism around the intact disulfide bond.

4.3.2 Computational analysis of the AL proteins to identify parameters determining amyloidogenicity

After resolving the primary structure of the AL proteins, the next step was to analyze the cases concerning properties that affect amyloidogenicity. The investigation focused on mutations and biochemical parameters such as net charge, solubility, hydrophobicity, and aggregation propensity.

4.3.2.1 Assignment of the GLs and identification of mutations in the precursor LC sequences

The identification of mutations in the amino acid sequence of the precursor LCs required the corresponding GL sequence as a reference. The GL segments of each case were initially assigned at the University Hospital Heidelberg and sent to Ulm University with the cDNA information. However, the resulting GL sequences from Heidelberg lacked amino acids at the end of the V segment and the start of the C segment (Fig. 44).

```
IGLV3-19*01_Heidelberg SSELTQDPAVSVALGQTVRITCQGDSLRSYYASWYQQKPGQAPVLVIYGKNNRPSGIPDR 60
IGLV3-19*01_Ulm        SSELTQDPAVSVALGQTVRITCQGDSLRSYYASWYQQKPGQAPVLVIYGKNNRPSGIPDR 60
IGLJ3*02_Heidelberg    ------------------------------------------------------------ 0
IGLJ3*02_Ulm           ------------------------------------------------------------ 0
IGLC2_Heidelberg       ------------------------------------------------------------ 0
IGLC2_Ulm              ------------------------------------------------------------ 0
                       ************************************************************

IGLV3-19*01_Heidelberg FSGSSSGNTASLTITGAQAEDEADYYCNSRDSSG---------------------- 94
IGLV3-19*01_Ulm        FSGSSSGNTASLTITGAQAEDEADYYCNSRDSSGNH-------------------- 96
IGLJ3*02_Heidelberg    ----------------------------------WVFGGGTKLTVL------------ 12
IGLJ3*02_Ulm           ----------------------------------WVFGGGTKLTVL------------ 12
IGLC2_ Heidelberg      ------------------------------------------------XQPKAAPSVTLF 12
IGLC2_Ulm              ------------------------------------------------GQPKAAPSVTLF 12
                       ********************************::************:************

IGLV3-19*01_Heidelberg ------------------------------------------------------------ 94
IGLV3-19*01_Ulm        ------------------------------------------------------------ 96
IGLJ3*02_Heidelberg    ------------------------------------------------------------ 12
IGLJ3*02_Ulm           ------------------------------------------------------------ 12
IGLC2_Heidelberg       PPSSEELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYL 72
IGLC2_Ulm              PPSSEELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYL 72
                       ************************************************************

IGLV3-19*01_Heidelberg ---------------------------------    94
IGLV3-19*01_Ulm        ---------------------------------    96
IGLJ3*02_Heidelberg    ---------------------------------    12
IGLJ3*02_Ulm           ---------------------------------    12
IGLC2_Heidelberg       SLTPEQWKSHRSYSCQVTHEGSTVEKTVAPTECS    106
IGLC2_Ulm              SLTPEQWKSHRSYSCQVTHEGSTVEKTVAPTECS    106
```

Figure 44: Problems of the GL assignment from Heidelberg University Hospital in comparison to the GL assignment from Ulm University. Blue letters: assigned GL sequence from Ulm University using the IMGT/GENE-DB* [116]. *Black letters: assigned GL sequence from Heidelberg University Hospital. Yellow: difference between the assignment of Heidelberg and Ulm.

In Heidelberg, V segments were assigned using the Vbase2 database, which aligned the precursor LC sequence to multiple sequences of different databases [140]. Vbase2 suggested a GL sequence based on the sequences from the alignment. However, the sequence of all V segments lacked the last two amino acids (Fig. 44) and was thus unable to provide information

about the junctional diversity of the LCs. The C segments were assigned in Heidelberg based on ENSEMBL [141], which compared the precursor LC sequence with cDNA, mRNA, and protein sequences from different databases [142]. As a result, ENSEMBL provided a reference sequence composed of the different sequences from the databases. However, in the case of the C segments, ENSEMBL did not identify an amino acid at position 1, and thus an “X” was reported (Fig. 44).

In this work, it was essential to identify all amino acids encoded in the GL segments of an undifferentiated bone marrow cell. Thus, comparing different sequences in the GL assignment using Vbase2 and ENSEMBL was redundant and resulted in incomplete information. In order to identify the complete GL sequence, all GL assignments were repeated as part of this work. V and C GL segments were identified by blasting the corresponding amino acid sequence of the precursor LCs against the UniProtKB/Swiss-Prot database [115]. The GL segments with the smallest difference from the precursor LC sequences were considered correct. The GL J segment was identified by directly comparing the corresponding part of the precursor LC sequence with the J segments deposited in the IMGT/GENE-DB [116]. As a next step, the sequences of the identified GL V, J, and C segments were taken from the IMGT/GENE-DB [116], as this database contains accurate gene data of the human genome and thus provides the complete genomic sequence of the different GL segments. The sequences of the V, J, and C segments were combined into a gapless GL sequence (Fig. 45).

```
FOR103          SSELTQDPAVSVALGQTVRITCQGDSLRSYYASWYQQKSGQAPVLVIYSYNNRPSGIPDR  60
IGLV3-19*01     SSELTQDPAVSVALGQTVRITCQGDSLRSYYASWYQQKPGQAPVLVIYGKNNRPSGIPDR  60
IGLJ2*01        ------------------------------------------------------------  0
IGLC2*03        ------------------------------------------------------------  0

FOR103          FSGSNSGNTASLTITGAQAEDEADYYCNSRDSSGHHLVFGGGTKLTVLGQPKAAPSVTLF  120
IGLV3-19*01     FSGSSSGNTASLTITGAQAEDEADYYCNSRDSSGNH------------------------  96
IGLJ2*01        ----------------------------------VVFGGGTKLTVL------------  12
IGLC2*03        ------------------------------------------------GQPKAAPSVTLF  12

FOR103          PPSSEELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYL  180
IGLV3-19*01     ------------------------------------------------------------  96
IGLJ2*01        ------------------------------------------------------------  12
IGLC2*03        PPSSEELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYL  72

FOR103          SLTPEQWKSHRSYSCQVTHEGSTVEKTVAPTECS  214
IGLV3-19*01     ----------------------------------  96
IGLJ2*01        ----------------------------------  12
IGLC2*03        SLTPEQWKSHRSYSCQVTHEGSTVEKTVAPTECS  106
```

Figure 45: Exemplary GL assignment to the precursor LC sequence of FOR103 performed as a part of this work. The GL segments IGLV3-19*01, IGLJ2*01, and IGLC2*03 were assigned to the precursor LC sequence. Mutations of the precursor LC in comparison to the GL are highlighted in red.

In cases where no cDNA was available and solely MS-based protein sequencing was performed, the resulting protein sequence was too short to provide information on the GL C segment. In those cases, the C segment was suggested from the identified J-segment, assuming that the combination of J and C segments does not require gene recombination. Since J and C segments are encoded side-by-side on the λ gene locus [29], it was considered that the LCs are composed of those IGLJ and IGLC pairs. For example, if the J segment could be identified as IGLJ3, the C segment was considered IGLC3. Amino acids of the LCs not covered by the AL protein were considered GL.

Two cases (FOR1001, FOR1007) could be assigned to the κ, and 17 cases (FOR009, FOR010, FOR012, FOR101, FOR102, FOR103, FOR107, FOR124, FOR128, FOR137, FOR138, FOR140, FOR142, FOR147, FOR157, FOR159, FOR161) to the λ isotype (Tab. 13).

Table 13: GL information on the LCs determined as a part of this thesis concerning isotype and GL subfamily. Slash: Indicates two possibilities for assigning a J and C segment. Bracket: Indicates a J or C GL segment, which could also be assigned to the corresponding LC sequence but is not encoded side-by-side to the corresponding unambiguously identified J or C segment on the λ gene locus and was thus considered not correct.

Case	Isotype	GL gene segments		
		V	J	C
FOR138	λ	1-40	1	1
FOR102	λ	1-44	1 (6)	1
FOR107	λ	1-44	2	2
FOR128	λ	1-44	3	3 (2)
FOR137	λ	1-44	2 (3)	2
FOR142	λ	1-44	2 (3)	2
FOR147	λ	1-44	2/3	2/3
FOR101	λ	2-14	1	1
FOR124	λ	2-14	3	3
FOR157	λ	2-14	1	1
FOR159	λ	2-14	3 (2)	3
FOR009	λ	3-1	2 (3)	2
FOR012	λ	3-1	3	3
FOR161	λ	3-1	1	1
FOR010	λ	3-19	3	3
FOR103	λ	3-19	2	2 (3)
FOR140	λ	6-57	3	3 (2)
FOR1001	κ	1-33	2	1
FOR1007	κ	1-33	2	1

The two κ LCs originated from the same V, J, and C segments (IGKV1-33, IGKJ2, IGKC) (Tab. 13). The C-segment was not part of the AL protein, so the corresponding GL segment was assigned using the cDNA information from Heidelberg. The 17 λ LCs originated from six different V segments, namely IGLV1-40 (FOR138), IGLV1-44 (FOR102, FOR107, FOR128, FOR137, FOR142, FOR147), IGLV2-14 (FOR101, FOR124, FOR157, FOR159), IGLV3-1 (FOR009, FOR012, FOR161), IGLV3-19 (FOR010, FOR103) and IGLV6-57 (FOR140) (Fig. 46, Tab. 13).

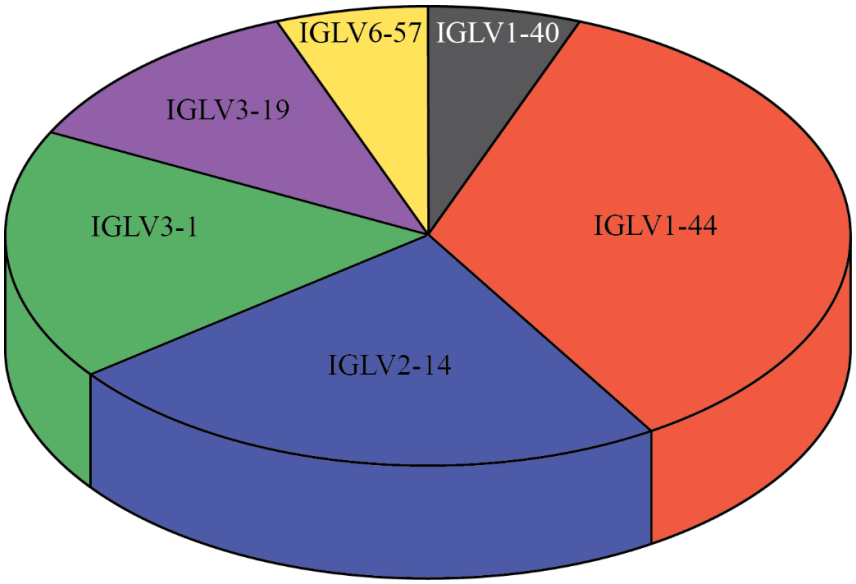


Figure 46: GL V segment distribution of the 17 λ cases. IGLV1-40: 1 case. IGLV1-44: 6 cases. IGLV2-14: 4 cases. IGLV1-3: 3 cases. IGLV3-19: 2 cases. IGLV6-57: 1 case.

Furthermore, the J and C segments of the LCs could be assigned to three different GLs. 5 LCs originated from IGLJ1/IGLC1 (FOR101, FOR102, FOR138, FOR157, FOR161), 5 LCs from IGLJ2/IGLC2 (FOR009, FOR103, FOR107, FOR137, FOR142, and 6 LCs from IGLJ3/IGLC3 (FOR010, FOR012, FOR124, FOR128, FOR140, FOR159) (Tab. 13). In six cases (FOR009, FOR102, FOR137, FOR142, FOR147, FOR159) LCs could be derived from two different J segments and in four cases (FOR103, FOR128, FOR140, FOR147) from two different C segments (Tab. 13). Assuming that a LC is composed of J and C segment pairs encoded side by side on the λ gene locus [29], the unambiguous assignment of either the J or C GL segment was sufficient to identify the correct corresponding other GL segment. For example, if the C segment was unambiguously assigned to IGLC3, IGLJ3 was considered the correct corresponding GL J segment. For FOR147, the correct IGLJ/IGLC pair could not be determined because two J and C segments were possible (Tab. 13).

Although previous studies have reported different organ involvement in AL amyloidosis depending on the use of different GL V segments [36–39,41], this study is unable to provide information on predominant deposition sites because of the small cohort size. Nevertheless, the present data are consistent with the findings of previous studies reporting predominant heart involvement in patients with IGLV2-14- and IGLV3-1-derived LCs [36,38]. All patients in this work expressing IGLV2-14- and IGLV3-1-derived LCs showed dominant heart involvement (Tab. 14). In the other GLs, however, the organ involvement did not represent the results from the literature.

Table 14: Organ involvement of each case in comparison to the corresponding GL V segment and the isotype of the expressed precursor LC.

Case	Isotype	GL V segment	Organ involvement
FOR138	λ	1-40	Heart
FOR102	λ	1-44	Heart
FOR107	λ	1-44	Kidney
FOR128	λ	1-44	Heart
FOR137	λ	1-44	Kidney
FOR142	λ	1-44	Kidney
FOR147	λ	1-44	Heart
FOR101	λ	2-14	Heart
FOR124	λ	2-14	Heart
FOR157	λ	2-14	Heart
FOR159	λ	2-14	Heart
FOR009	λ	3-1	Heart
FOR012	λ	3-1	Heart
FOR161	λ	3-1	Heart
FOR010	λ	3-19	Heart
FOR103	λ	3-19	Heart
FOR140	λ	6-57	Heart
FOR1001	κ	1-33	Heart/kidney
FOR1007	κ	1-33	Heart

After the assignment of the GL, mutations could be identified (Tab. 15). In this work, all deviations from the GL sequence were identified, such as amino acid substitutions, insertions, and deletions. Insertions and deletions in the VJ junction were interesting, as they arose most likely from the imprecise joining of the V and J segments, the so-called junctional diversity (Tab. 15).

Table 15: Information on the LCs determined as a part of this thesis concerning the number of mutations and junctional diversity. The table displays the total number of mutations, the number of mutations in the V_L and C_L domains, and the number of mutations in the CDRs and FRs of the V_L domain. Junctional diversity is indicated by a cross and divided into insertion and deletion.

Case	Number of mutations					Junctional diversity	
	Sum of mutations	V_L	CDRs	FRs	C_L	Insertion	Deletion
FOR138	12	11	7	4	1	X	
FOR102	11	11	4	7	0		
FOR107	8	8	5	3	0		
FOR128	10	10	3	7	0		
FOR137	9	9	5	4	0		
FOR142	7	7	2	5	0		
FOR147	9	9	5	4	0		
FOR101	14	14	8	6	0		
FOR124	10	10	6	4	0	X	
FOR157	7	7	4	3	0		X
FOR159	7	7	6	1	0		
FOR009	10	10	4	6	0		X
FOR012	8	7	4	3	1		X
FOR161	7	7	3	4	0		X
FOR010	13	13	6	7	0		
FOR103	6	6	4	2	0		
FOR140	10	9	4	5	1	X	
FOR1001	15	15	3	12	0		
FOR1007	7	7	2	5	0	X	

As a first step, the mutations of the cases expressing κ LCs were investigated. FOR1001 contained 15 mutations: 3 in the CDRs and 12 in the FRs (Fig. 47, Tab. 15). FOR1007 contained 7 mutations: 2 in the CDRs and 5 in the FRs (Fig. 47, Tab. 15). No mutations occurred in the C_L domain. Junctional diversity was found in one (FOR1007) of two κ cases and comprised an amino acid insertion (Tab. 15).

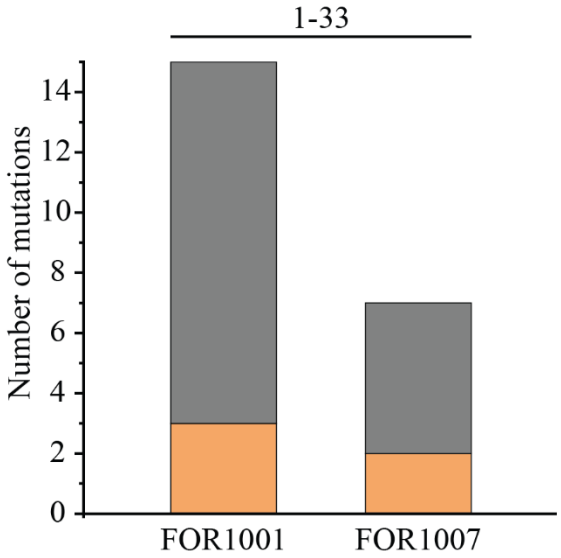


Figure 47: Mutations of the κ cases FOR1001 and FOR1007. Orange: Mutations in the CDRs. Gray: Mutations in the FRs. The GL (IGKV1-33) is illustrated above the blot.

In the next step, mutations of the λ LCs were investigated. The total number of mutations varied from 6 to 14 (Fig. 48, Tab. 15), which is a similar range as for the κ LCs. The number of mutations in the precursor LC sequences did not appear to differ significantly from GL to GL. However, no statistical test could be performed because some GL V segments were only present in the LC of one patient. The number of mutations in the CDRs varied from 2 to 8 and in the FRs from 1 to 7 (Fig. 48, Tab. 15). Three cases (FOR138, FOR140, FOR012) had a single mutation in the C_L domain (Fig. 48, Tab. 15). Junctional diversity was present in eight cases (FOR009, FOR012, FOR124, FOR138, FOR140, FOR157, FOR161, FOR1007) comprising 4 insertions (FOR124, FOR138, FOR140, FOR1007) and 4 deletions (FOR009, FOR012, FOR157, FOR161) (Tab. 15).

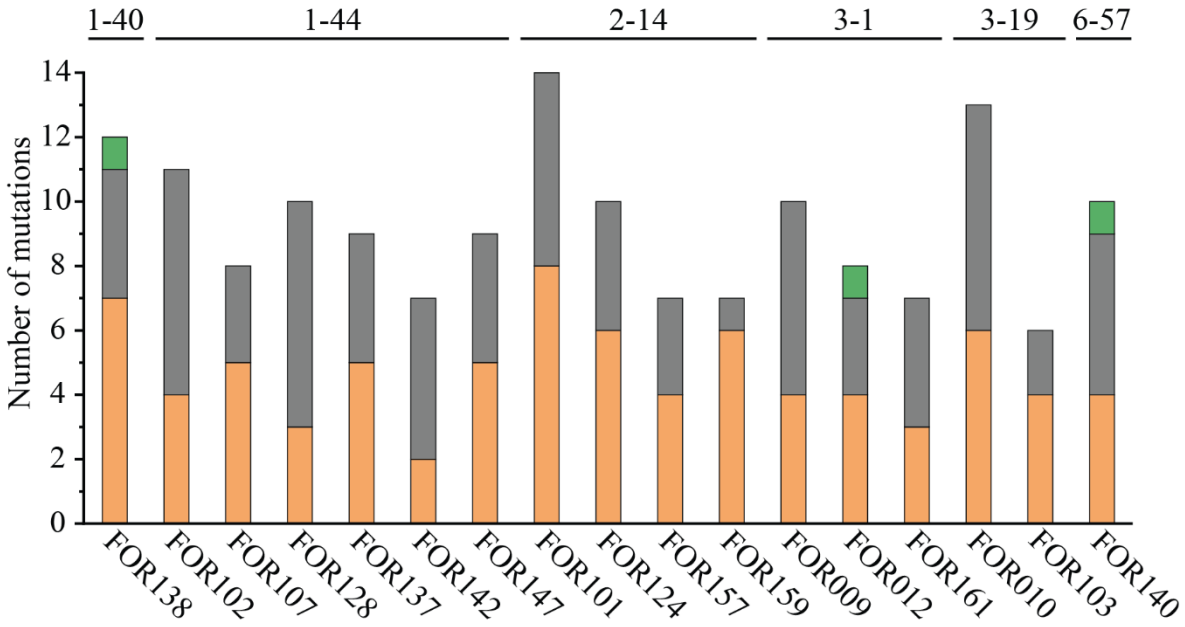


Figure 48: Mutations of the 17 λ cases. Orange: Mutations in the CDRs of the V_L domain. Gray: Mutations in the FRs of the V_L domain. Green: Mutations in the C_L domain. The GL is illustrated above the blot.

In summary, the successful GL assignment allowed the identification of mutations in the precursor LC sequence of the 19 cases. The total number of mutations and their distribution in the CDRs and FRs varied among the 19 cases, suggesting that it is not the number of mutations but their position and the biochemical properties of the amino acids that determine amyloidogenicity.

4.3.2.2 Comparison of the mutation sites of all 19 precursor LCs

Investigating mutations present in the amyloidogenic LCs of different cases could identify mutation sites, which increase the amyloidogenicity of LCs. Therefore, an alignment of all cases was made containing all mutations and the secondary structure elements of a natively folded LC (PDB: 6QB6-L [136]) (Fig. 49).

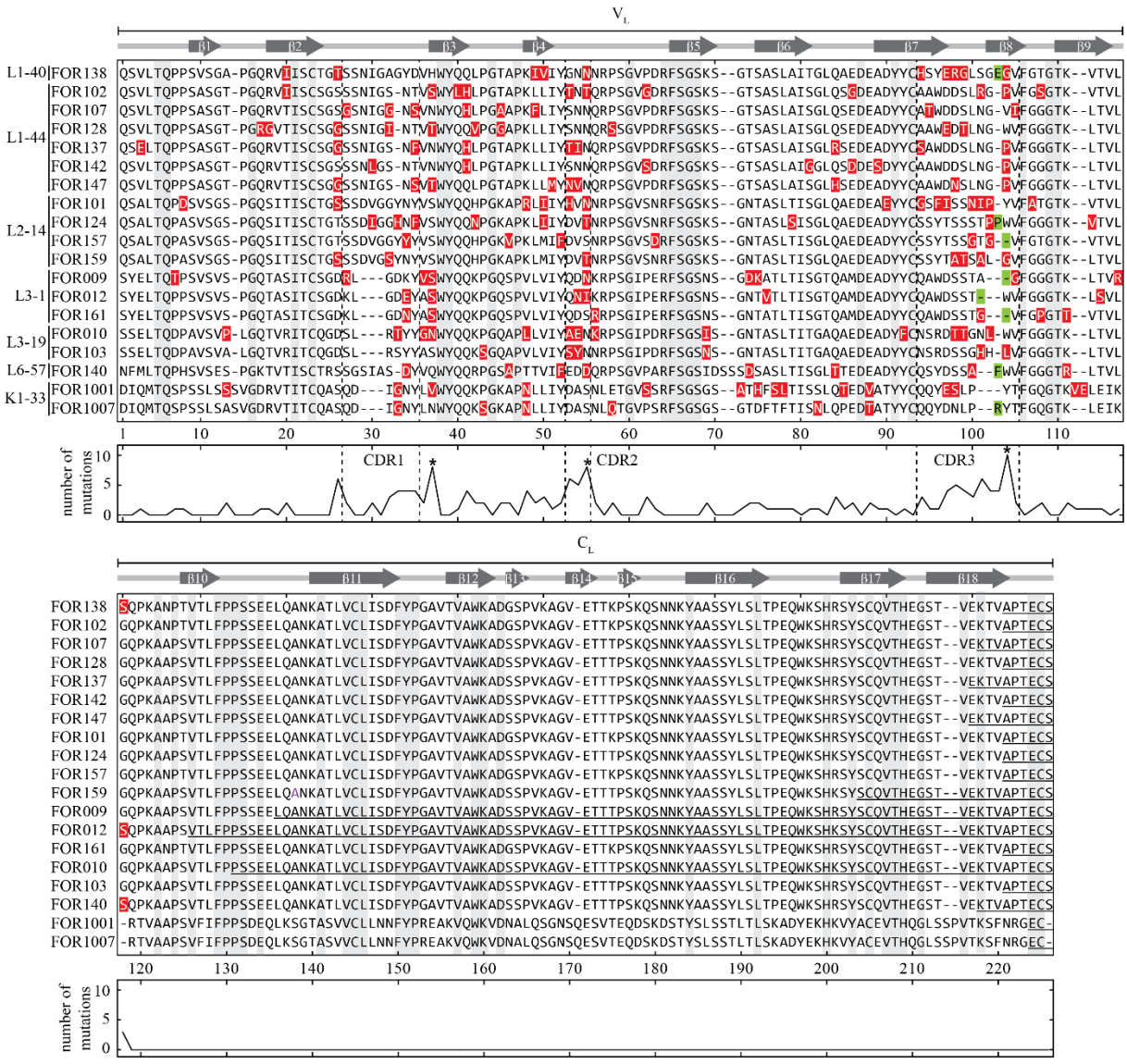


Figure 49: Alignment of all 19 precursor LCs. Amino acids are numbered consecutively. Red background: amino acid mutations compared to the GL sequence. Green background: deletions or insertions resulting from junctional diversity. Gray background: conserved amino acids. Violet letter: Uncertain amino acid. Underlined amino acids: regions neither identified by cDNA sequencing nor MS comprising the GL sequence. Dotted lines: CDRs (assigned according to IMGT/GENE-DB* [116]*). Gray Arrows: beta strands of a natively folded LC (PDB: 6QB6-L* [136]*). Gray lines: Loops of a natively folded LC (PDB: 6QB6-L* [136]*). Graph below the alignment: mutations per amino acid position. Asterisks: Mutations occur in more than 40 % of LCs.

The position of mutations always refers to the consecutive numbering of figure 49. The mutations mainly clustered in the CDRs or adjacent to them (Fig. 49). More than 40 % of all LCs had mutations at positions 37 (adjacent to CDR1), 55 (within CDR2), and 104 (within CDR3) (Fig. 49). The mutations at position 55 and 104 are consistent with previously identified mutational hotspots in λ AL amyloidosis [22]. However, other positions like 43, 46, 52, 53, 73,

89, and 92, which are known to affect the amyloidogenicity of LCs [22,33,44,49,50], were rarely mutated in this dataset (Fig. 49). V_L domain substitutions were generally LC-specific, and hardly any mutation occurred in multiple LCs from different patients. A unique but interesting mutation in the V_L domain of FOR1001 was the insertion of two amino acids at the end of the V_L domain (Fig. 49). It was located outside the CDRs eight amino acids away from the V-J junction and thus not explainable by junctional diversity. The C_L domain was mutated in only three precursor LCs (FOR012, FOR138, FOR140) (Fig. 49). In all three cases, it was an amino acid substitution of a non-polar glycine (Gly) by a polar Ser at the N-terminus of the C_L domain at position 118 (Fig. 49). In previous studies, mutations have been shown to create glycosylation sites in LC sequences that correlate with AL amyloidosis [53]. In order to investigate whether mutations also created consensus sequences for glycosylation in the precursor LCs of this study, the LC and GL sequences were analyzed bioinformatically by NetNGlyc 1.0 [121]. The analysis was limited to N-glycosylation sites with the motif Asn-X-Ser (X ≠ Pro) because no motif for O-glycosylation has so far been identified. However, neither the GL sequences nor the precursor LC sequences contained N-glycosylation sites. In summary, comparing all 19 precursor LC sequences did not identify a mutation that was present in the majority of cases. This observation suggests that it is not one specific mutation that increases the amyloidogenicity of all LCs. Instead, the data suggest that different mutations increase the amyloidogenicity of LCs, probably depending on the GL segments underlying a LC.

4.3.2.2.1 Comparison of mutation sites in LCs derived from the same GL V segment

Therefore, it was investigated whether some mutations occurred exclusively in LCs derived from the same GL V segment and thus could influence amyloidogenicity on the GL level. At first, the two κ cases, FOR1001 and FOR1007, were investigated expressing IGKV1-33-

derived LCs. Two mutations occurred in both cases: the substitution of Ser to Gly at position 33 and the substitution of Lys to Asn at position 48 (Fig. 50). Both mutations altered the biochemical properties of the amino acids, which could affect LC stability. It is also interesting that mutations occurred mainly in the FRs, although the CDRs are known for their variability [20].

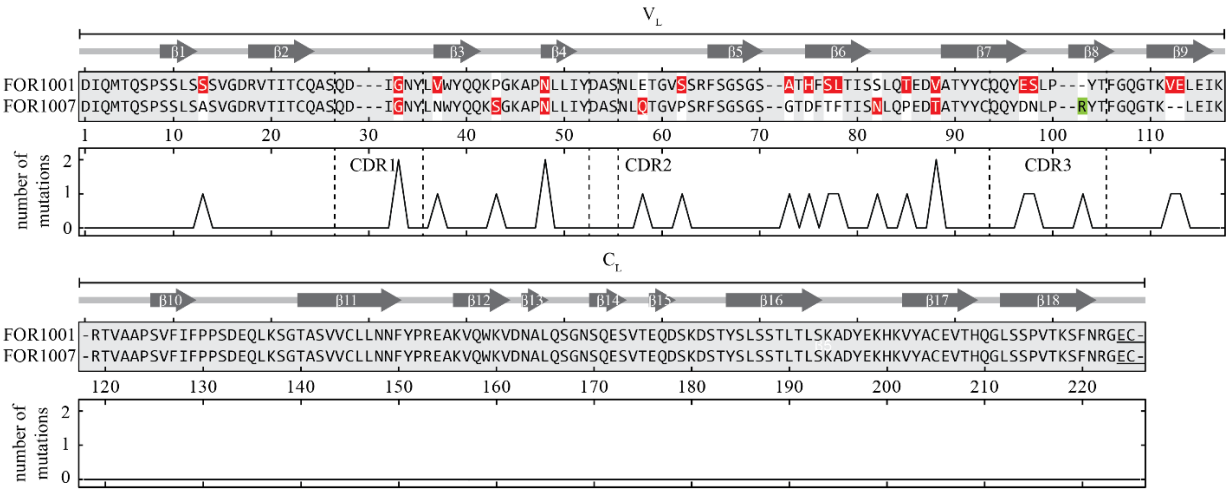


Figure 50: Sequence alignment of the κ cases FOR1001 and FOR1007. Red background: amino acid mutations compared to the GL sequence. Green background: deletions or insertions resulting from junctional diversity. Gray background: conserved amino acids. Underlined amino acids: regions neither resolved by cDNA sequencing nor MS comprising the GL sequence. Dotted lines: CDRs (assigned according to IMGT/GENE-DB* [116]*). Gray Arrows: beta strands of a natively folded LC (PDB: 6QB6-L* [136]*). Gray lines: Loops of a natively folded LC (PDB: 6QB6-L* [136]*). Graph below the alignment: mutations per amino acid position.

As a next step, the λ LCs with the same underlying GL V segment were investigated for the same mutations. 67 % of the IGLV1-44-derived LCs had an amino acid substitution from Gln to histidine (His) at position 41 (Fig. 51), and 50 % a mutation from Ser to Gly at position 26 (Fig. 51). Both substitutions altered the biochemical properties at these positions, suggesting a destabilizing effect on the IGLV1-44-derived LCs. In particular, exchanging a polar Gln for a charged His could influence the net charge or alter the LC conformation. The mutation from any amino acid to Pro at position 104 (Fig. 51) occurred in 67 % of the IGLV1-44-derived LCs. The Pro was located in β-strand 8 and could alter the structure of IGLV1-44-derived LCs due

to its ability to break β-strands [143]. Interestingly, IGLV1-44-derived LCs contained no insertions or deletions in the V-J junction.

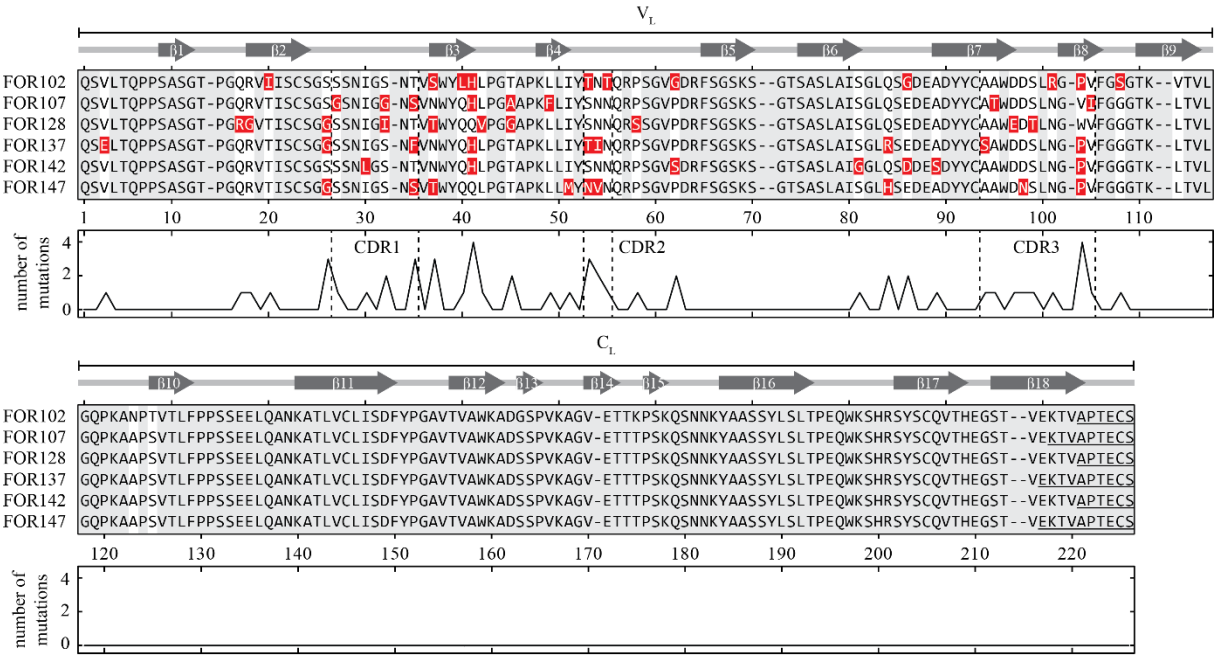


Figure 51: Sequence alignment of the IGLV1-44-derived cases. Red background: amino acid mutations compared to the GL sequence. Gray background: conserved amino acids. Underlined amino acids: regions neither resolved by cDNA sequencing nor MS comprising the GL sequence. Dotted lines: CDRs (assigned according to IMGT/GENE-DB* [116]*). Gray Arrows: beta strands of a natively folded LC (PDB: 6QB6-L* [136]*). Gray lines: Loops of a natively folded LC (PDB: 6QB6-L* [136]*). Graph below the alignment: mutations per amino acid position.

All IGLV3-1-derived LCs contained the unique amino acid exchange of Cys to Ser at position 37 (Fig. 52). It is important to note that it is not the Cys at position 37 but the Cys at position 23 that forms the canonic disulfide bond of a natively folded V_L domain. However, mutation of the Cys at position 37 could disrupt intermolecular interactions by preventing disulfide bond formation with other molecules. In contrast to the IGLV1-44-derived LCs, all IGLV3-1-derived LCs contained deletions in the V-J junction, suggesting that junctional diversity is GL-dependent.

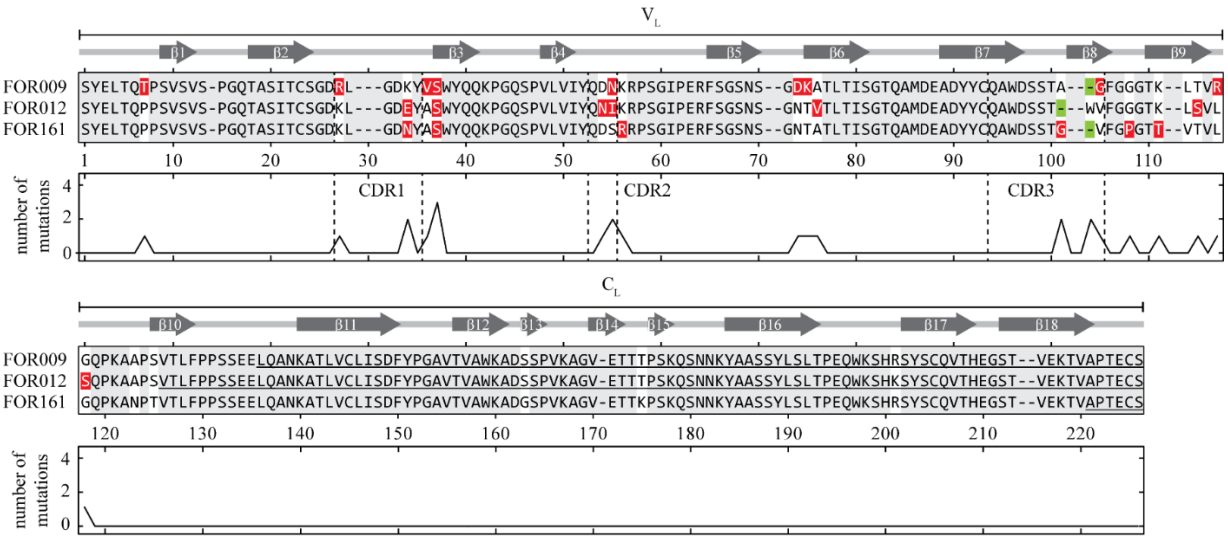


Figure 52: Sequence alignment of the IGLV3-1-derived cases. Red background: amino acid mutations compared to the GL sequence. Green background: deletions or insertions resulting from junctional diversity. Gray background: conserved amino acids. Underlined amino acids: regions neither resolved by cDNA sequencing nor MS comprising the GL sequence. Dotted lines: CDRs (assigned according to IMGT/GENE-DB* [116]*). Gray Arrows: beta strands of a natively folded LC (PDB: 6QB6-L* [136]*). Gray lines: Loops of a natively folded LC (PDB: 6QB6-L* [136]*). Mutations per amino acid position are displayed in a graph below the sequence alignment.

In summary, LCs with the same IGLV segment shared the same amino acid substitutions, suggesting that mutations increasing the amyloidogenicity of LCs are GL-specific.

4.3.2.3 Mutation sites in the three-dimensional structure of a natively folded LC

Amyloidogenic mutations disrupt either the native fold of the precursor LCs or affect fibril formation. In both cases, the biochemical properties and location of the mutated amino acid are relevant. Therefore, the next step was to investigate the spatial position of the mutations in a natively folded IGLV1-44-derived LC (Fig. 53).

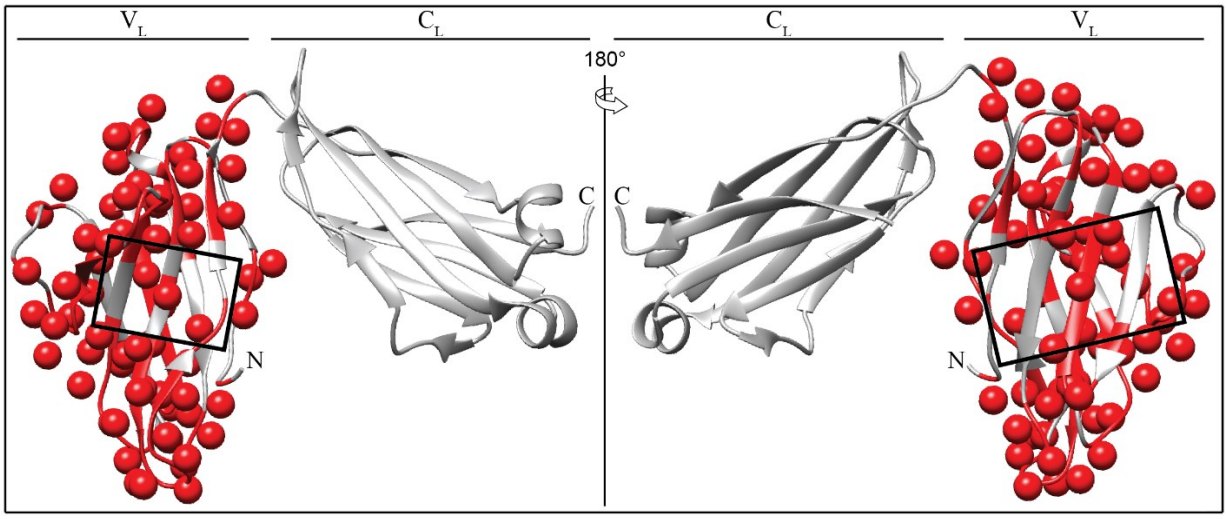


Figure 53: Illustration of the mutation sites of all 19 cases in the native structure of an IGLV1-44-derived LC (PDB: 6QB6-L) **[136].** ***Red spheres: mutation sites with spatial orientation of the amino acids. Black rectangle: beta strands in the middle of the V_L domain with few mutations. Gray line: unstructured loop. Gray arrow: β-strand. Gray twist without arrowhead: α-helix. N: N-terminus. C: C-terminus.***

The mutations were distributed throughout the V_L domain of the LC and occurred in all secondary structure elements (unstructured loops, α-helices, β-strands) (Fig. 53). However, fewer mutations occur in the beta strands in the middle of the V_L domain (Fig. 53, black rectangle), forming two mutation clusters at the top and bottom of the V_L domain (Fig. 53). The amino acid substitutions at the lower end were located in the CDRs, which build the antigen binding site and are known for their high variability [20]. The amino acids in the upper region were part of the conserved FRs. This clustering is consistent with previously identified mutation hotspots [127]. In conclusion, these observations suggest that the beta strands in the middle of the V_L domain are important to maintain the native fold of the LCs and are, therefore, rarely mutated. Mutations within this region could destabilize the LCs and increase their amyloidogenic potential. There is evidence that mutations in the upper region of the V_L domain can affect domain interactions due to their proximity to the C_L domain. Mutations in the CDRs are less likely to change the LC structure, as CDRs are located at the lower tip of the V_L domain and exhibit natural variability due to their antigen specificity. However, given the large number of different mutations, the destabilizing mutations remain unknown.

4.3.2.4 Computational analysis of biochemical parameters from the amyloidogenic LCs in comparison to the GL

In the next step, the consequences of mutations on different biochemical properties of the LCs were analyzed, such as net charge, solubility, hydrophobicity, and aggregation propensity. Corresponding values were calculated for GL and LC sequences using web-based bioinformatic tools. GL refers to the non-mutated sequence encoded in the human genome, and LC to the mutated sequence of the precursor LC. Difference values of LC and GL were expressed by delta (Δ) values, calculated by subtracting GL from corresponding LC values. All values were rounded to one decimal place.

4.3.2.4.1 Investigation of the net charge

Mutation-induced changes in net charge were investigated by calculating the pI using the ProtParam tool of ExPASy [117]. The full-length LC and the corresponding AL proteins of all 19 cases were investigated. In the 16 cases where the assignment of monoisotopic masses identified multiple AL proteins, the pI was determined for all AL proteins and averaged. AL proteins are referred to as LC fragments in the bioinformatic analysis to avoid misunderstandings. The same calculations were conducted for the full-length GLs and GL fragments (Fig. 54, App. Tab 19, 20).

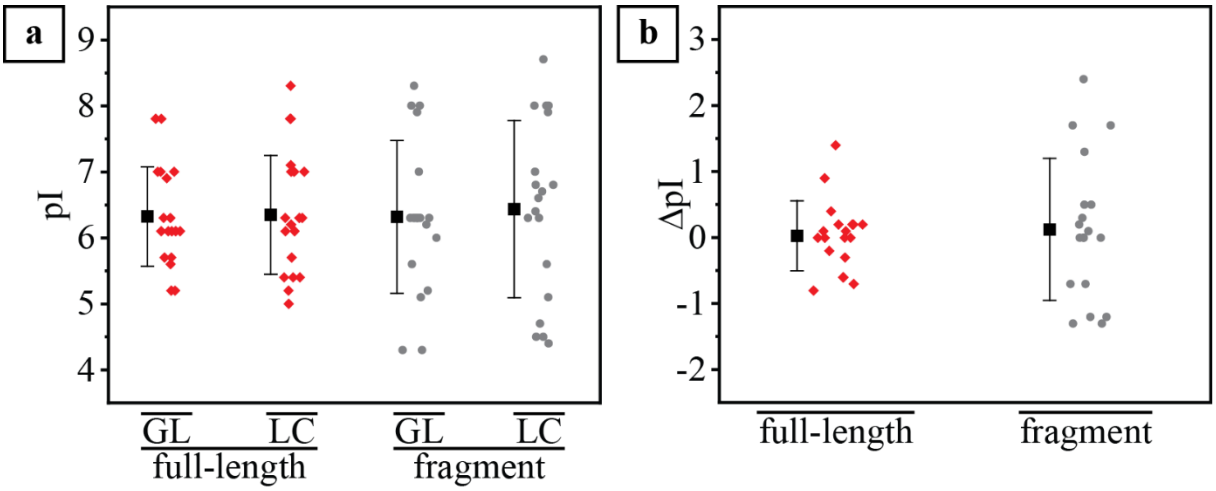


Figure 54: pI of full-length GLs and LCs (red diamonds) as well as GL and LC fragments (gray circles). a) pI values of the GL and LC sequences of full-length proteins and fragments. b) ΔpI values of full-length proteins and fragments. ΔpI = LC – GL.

The pI of the full-length proteins did not significantly differ from that of the fragments (GL: p = 0.987, LC: p = 0.811) based on a two-tailed t-test (Fig. 54). In addition, the mean pI of the full-length GLs did not significantly differ from the mean pI of the full-length LCs (p = 0.923), and the mean pI of the GL fragments did not significantly differ from the mean pI of the LC fragments (p = 0.768) (Fig. 54 a, App. Tab. 19, 20). However, in individual cases, the pI of the LC fragments increased substantially compared to the GL, as indicated by ΔpI values of >1.0 pH unit in four cases (FOR102, FOR137, FOR147, FOR009) (Fig. 54 b, App. Tab. 19, 20). The data suggest that the C_L domain attenuates the effect of mutations on the pI, as the ΔpI values of the full-length proteins are less distributed (SD = 0.5) and closer to 0.0 than those of the fragments (SD = 1.1). Therefore, the V_L domain and C_L domain were analyzed separately to investigate the effect of the mutations on the V_L domain without the attenuating effect of the C_L domain (Fig. 55, App. Tab. 19).

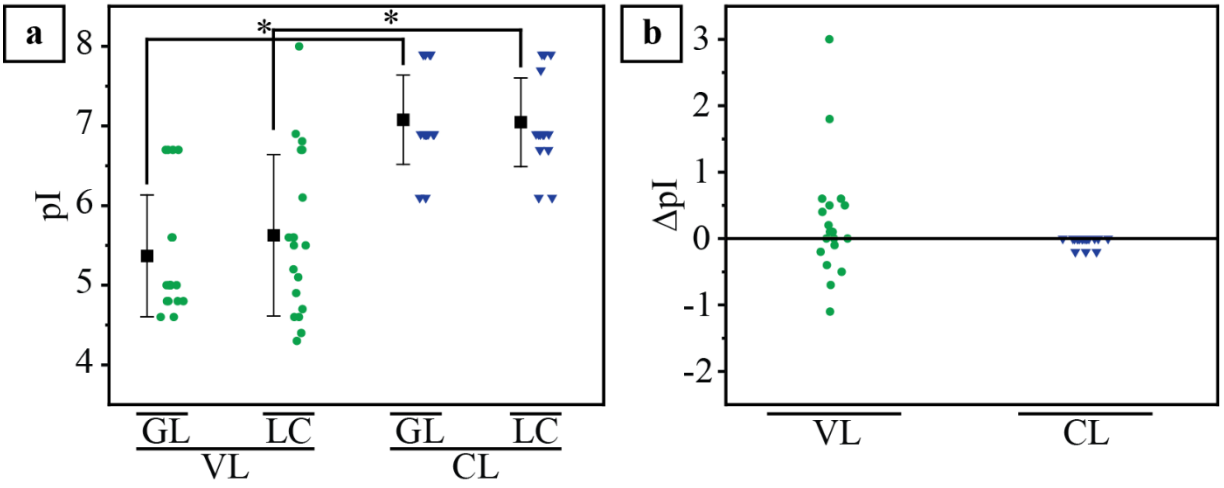


Figure 55: pI of V_L (green circles) and C_L domains (blue triangles). a) Absolut pI values for the GL and LC sequences of V_L and C_L domains. b) ΔpI values of V_L and C_L domains. ΔpI = LC – GL.

The mean pI of the V_L domains was significantly lower than the mean pI of the C_L domains (GL: $p = 4.7 \times 10^{-9}$, LC: $p = 9.7 \times 10^{-6}$) (Fig. 55 a, App. Tab. 19). Despite the absence of the C_L domain, the difference between the mean pI of the GL V_L domains and LC V_L domains was not significant ($p = 0.391$) (Fig. 55 a, App. Tab. 19). In addition, the individual ΔpI values of the V_L domain were generally closer to 0.0 than those of AL proteins, suggesting that only the C-terminal part of the C_L domain attenuates pI changes (Fig. 55 b, App. Tab. 19). The mean pI of GL and LC C_L domains did not significantly differ ($p = 0.863$), which was expected because almost all mutations that could affect the pI of the LCs are located in the V_L domain (Fig. 55, App. Tab. 19).

In summary, the pI of LC and GL sequences did not differ in general but did in individual cases, indicating that the pI affects amyloidogenicity only in some cases. The most prominent pI changes occurred in the fragments, suggesting that the C-terminus of the C_L domain attenuates the effect of mutations on the pI of LCs, while the N-terminus of the C_L domain enhances the effect of the mutations on pI. There is evidence that the combination of V_L and C_L domains is important for LC destabilization because the pI is closer to the physiological pH in full-length

proteins and fragments. Mutations could further reduce the discrepancy between pI and physiological pH, thus lowering the net charge and the solubility of complete LCs and AL proteins.

4.3.2.4.2 Investigation of the solubility

Solubility was calculated for the full-length LCs and the AL proteins using the CamSol web server [120]. In the 16 cases where the assignment of monoisotopic masses identified multiple AL proteins, the solubility was determined for all AL proteins and averaged. AL proteins are referred to as LC fragments in the bioinformatic analysis to avoid misunderstandings. The same calculations were conducted for the full-length GLs and GL fragments (Fig. 56, App. Tab 21, 22).

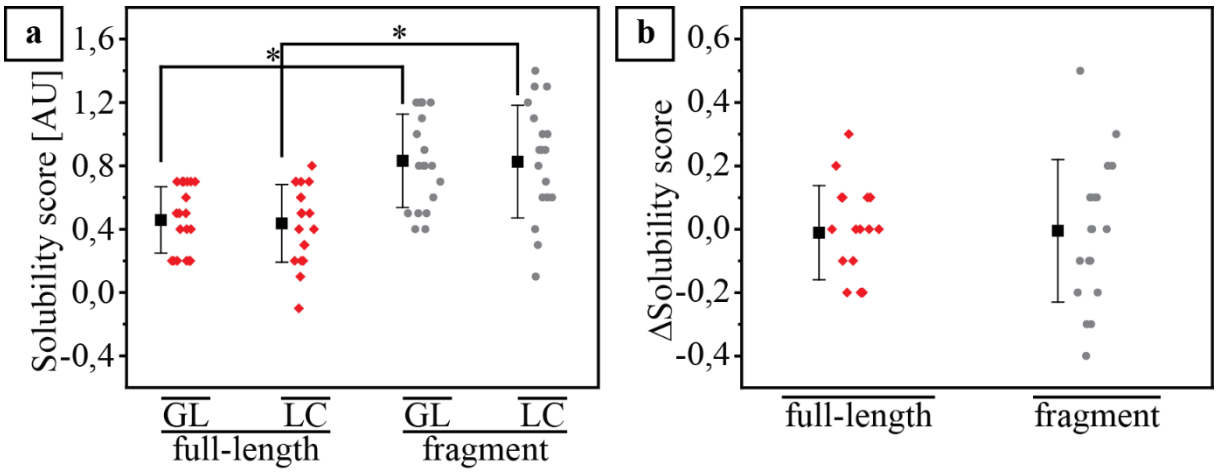


Figure 56: Solubility score of full-length GLs and LCs (red diamonds) as well as GL and LC fragments (gray circles). a) Absolut solubility score (in arbitrary units) for the GL and LC sequences of full-length proteins and fragments. b) Δsolubility values of full-length proteins and fragments. ΔpI = LC – GL.

The calculations showed a significant difference in solubility between the full-length proteins and the fragments (GL: $p = 8.1 \times 10^{-5}$, LC: $p = 4.3 \times 10^{-4}$), indicating that full-length proteins are less soluble on average than the fragments (Fig. 56 a, App. Tab. 21, 22). However, only

scores smaller than -1.0 characterize a protein as poorly soluble [120], suggesting that the difference in solubility is irrelevant because the full-length proteins and fragments have solubility scores above -1.0 and are thus both soluble. The mean solubility of full-length GLs did not differ significantly from the mean solubility of the full-length LCs (p = 0.778), and the mean solubility of GL fragments did not differ significantly from the mean solubility of the LC fragments (p = 0.961) (Fig. 56 a, App. Tab. 21, 22). As a next step, V_L and C_L domains were analyzed separately to investigate whether the effect of mutations on solubility is higher in isolated V_L domains (Fig. 57, App. Tab. 21).

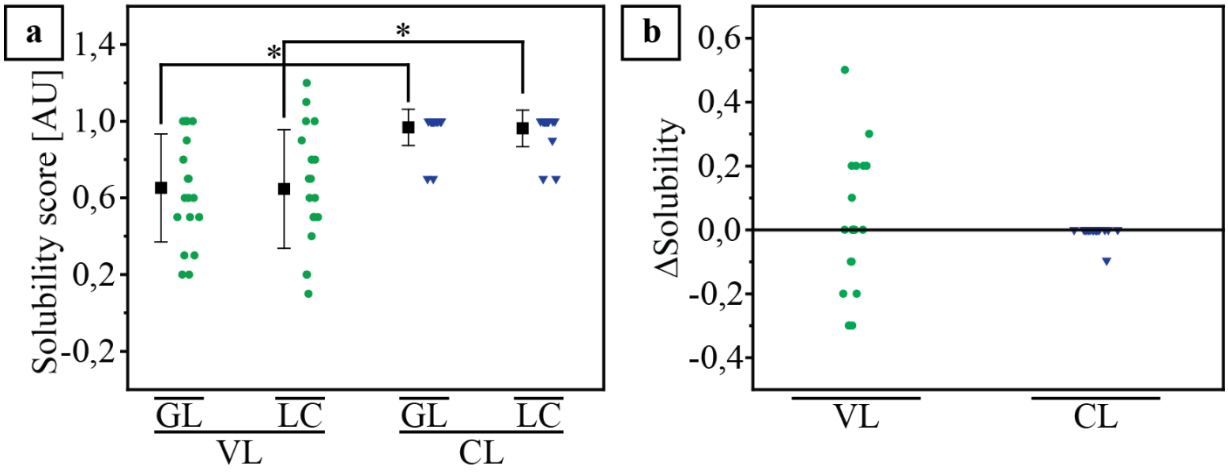


Figure 57: Solubility score of the V_L (green circles) and C_L domains (blue triangles). a) Absolut solubility scores (in arbitrary units) for the GL and LC sequences of the V_L and C_L domains. b) Δsolubility values of V_L and C_L domains. ΔpI = LC – GL.

The mean solubility of the V_L domains was significantly lower than the solubility of the C_L domains (GL: $p = 1.3 \times 10^{-4}$, LC: $p = 3.6 \times 10^{-4}$) (Fig. 57 a, App. Tab. 21). However, the mean solubility of the V_L differed not significantly between GL and LC sequences (p = 0.956), suggesting that the effect of mutations on the solubility is not substantial even in isolated V_L domains (Fig. 57 a, App. Tab. 21).

In summary, these data show that mutations do not substantially reduce the solubility of amyloidogenic LCs. Instead, the solubility increased when parts of the C_L domain were missing, suggesting that solubility does not generally affect the amyloidogenicity of LCs.

4.3.2.4.3 Investigation of the hydrophobicity

Changes in hydrophobicity were investigated by determining the GRAVY using the ProtParam tool of ExPASy [117]. The GRAVY was calculated by the sum of all hydropathy values of an amino acid sequence divided by the total number of amino acids [117,118] for full-length LCs and AL proteins. In the 16 cases where the assignment of monoisotopic masses identified multiple AL proteins, the GRAVY was determined for all AL proteins and averaged. AL proteins are referred to as LC fragments in the bioinformatic analysis to avoid misunderstandings. The same calculations were conducted for the full-length GLs and GL fragments (Fig. 58, App. Tab 23, 24).

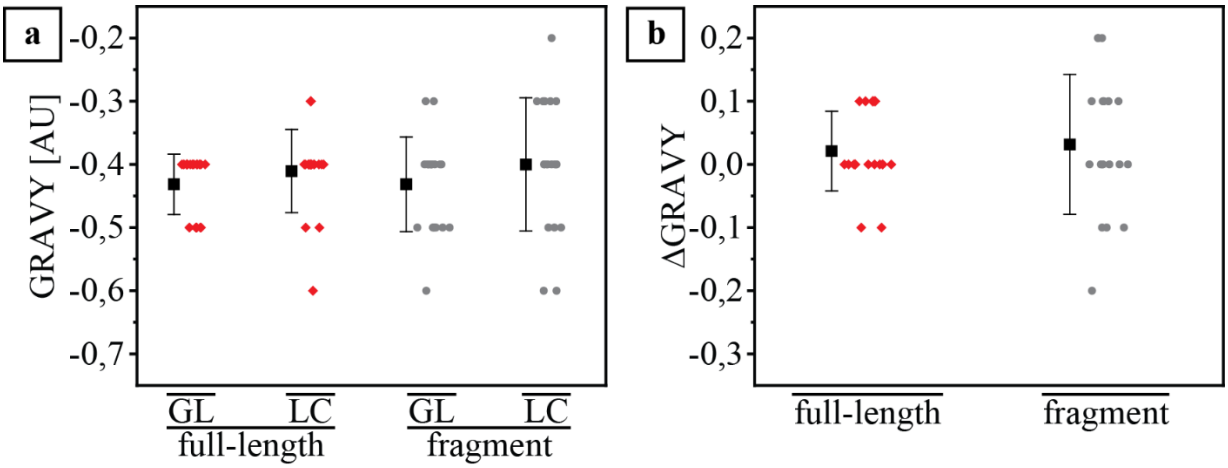


Figure 58: GRAVY of full-length GLs and LCs (red diamonds) as well as GL and LC fragments (gray circles). a) Absolut GRAVY values (in arbitrary units) for the GL and LC sequences of full-length proteins and fragments. b) ΔGRAVY values of full-length proteins and fragments. ΔpI = LC – GL.

The mean GRAVY of full-length proteins and fragments did not differ significantly (GL: p = 1.000, LC: p = 0.714) (Fig. 58 a, App. Tab. 23, 24). Furthermore, the mean GRAVY

of full-length GLs did not differ significantly from the mean GRAVY of full-length LCs ($p = 0.267$), and the mean GRAVY of GL fragments did not differ significantly from the mean GRVAY of LC fragments ($p = 0.295$) (Fig. 58 a, App. Tab 23, 24). This observation suggests that neither LC truncation nor mutations substantially increase the hydrophobicity. As a next step, V_L and C_L domains were analyzed separately to investigate whether the effect of the mutations on the hydrophobicity is higher in the individual domains (Fig. 59, App. Tab. 23).

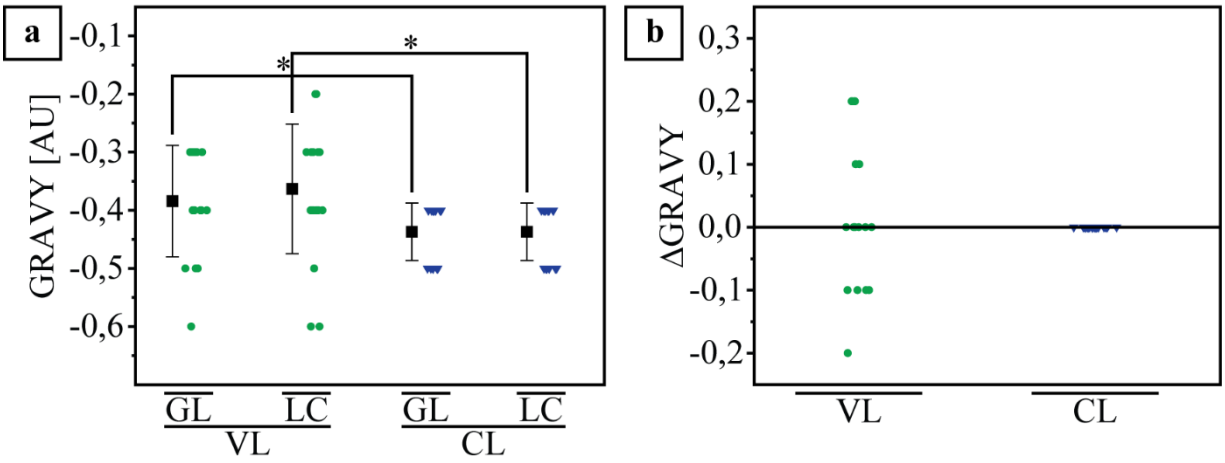


Figure 59: GRAVY of the V_L (green circles) and C_L domains (blue triangles). a) Absolut GRAVY score (in arbitrary units) for the GL and LC sequences of the V_L and C_L domains. b) ΔGRAVY score of the V_L and C_L domains. ΔpI = LC – GL.

The mean GRAVY of the V_L domains differed significantly from that of the C_L domains (GL: $p = 0.043$, LC: $p = 0.014$), indicating that the V_L domain is more hydrophobic than the C_L domain (Fig. 59 a, App. Tab. 23). The mean GRAVY of GL and LC V_L domains, and the mean GRAVY of GL and LC C_L domains, did not differ significantly (V_L: $p = 0.537$, C_L: $p = 1.000$) (Fig. 59 a, App. Tab. 23).

In conclusion, the GRAVY values did not differ significantly between GL and LC sequences suggesting that mutations have no substantial effect on hydrophobicity. However, the hydrophobicity was significantly higher in V_L domains compared to C_L domains, which might be one reason that AL proteins consist mainly of the V_L domain.

4.3.2.4.4 Investigation of the aggregation propensity

The aggregation propensity of the different proteins was calculated using TANGO. The algorithm provided an AGG score, which was normalized in this work by dividing it by the sum of amino acids of the corresponding protein. Aggregation propensity was calculated for full-length LCs and AL proteins. In the 16 cases where the assignment of monoisotopic masses identified multiple AL proteins, the aggregation propensity was determined for all AL proteins and averaged. AL proteins are referred to as LC fragments in the bioinformatic analysis to avoid misunderstandings. The same calculations were conducted for the full-length GLs and GL fragments (Fig. 60, App. Tab. 25, 26).

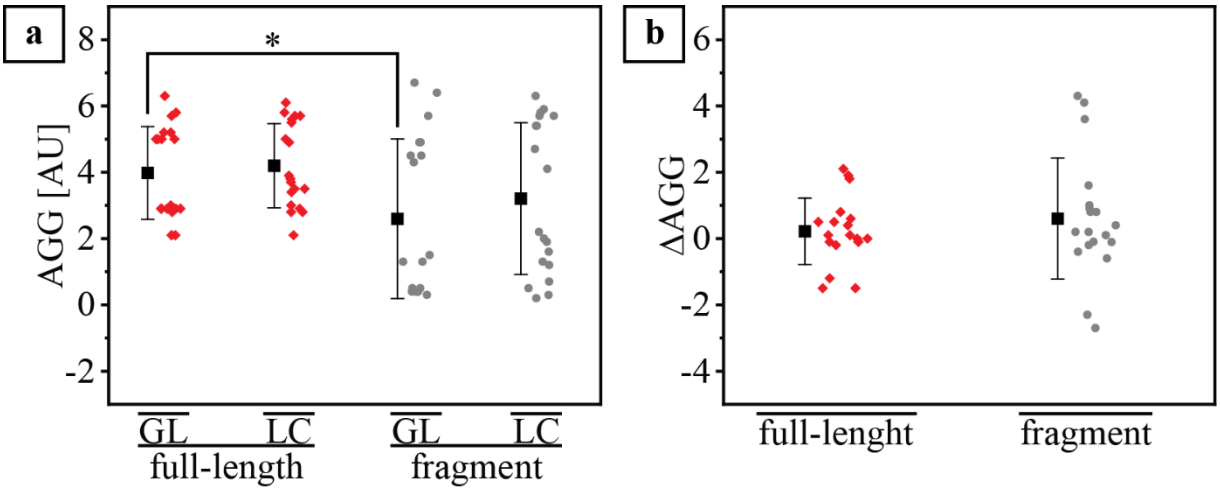


Figure 60: Aggregation score (AGG) of full-length GLs and LCs (red diamonds) as well as GL and LC fragments (gray circles). a) Absolut normalized AGG values (in arbitrary units) for the GL and LC sequences of full-length proteins and fragments. b) ΔAGG values of full-length proteins and fragments. ΔpI = LC – GL.

The mean AGG score of full-length GLs was significantly higher than the mean AGG score of GL fragments ($p = 0.039$) (Fig. 60 a, App. Tab. 25, 26), which suggests that the C_L domain has a higher aggregation propensity than the V_L domain. However, if the C_L domain has a higher aggregation propensity than the V_L domain, one would expect fibrils to consist of the C_L domain. The fact that V_L domains and not C_L domains are the major part of the fibril-forming AL proteins suggests that several factors besides aggregation propensity affect fibril formation.

In addition, this observation suggests that proteolytic cleavage of the C_L domain could happen after fibril formation. The mean AGG score of full-length GLs did not differ significantly from the mean AGG score of full-length LCs (p = 0.622), and the mean AGG score of the GL fragments did not differ significantly from the mean AGG score of the LC fragments (p = 0.428) (Fig. 60 a, App. Tab. 25, 26). This observation indicates no general effect of mutations on aggregation propensity. However, in individual cases, the ΔAGG score of the LC fragments increased substantially compared to the GL, as indicated by ΔAGG values of >4.0 in two cases (FOR137, FOR1001) (Fig. 60 b, App. Tab. 25, 26). Furthermore, the ΔAGG values of the full-length proteins were less distributed (SD = 1.0) and closer to 0.0 than those of the fragments (SD = 1.8), providing evidence that the effect of mutations on the aggregation propensity is higher in the absence of the C_L. Therefore, V_Ls and C_Ls were analyzed separately to investigate the effect of the C_L domain on the aggregation propensity (Fig. 61, App. Tab. 25).

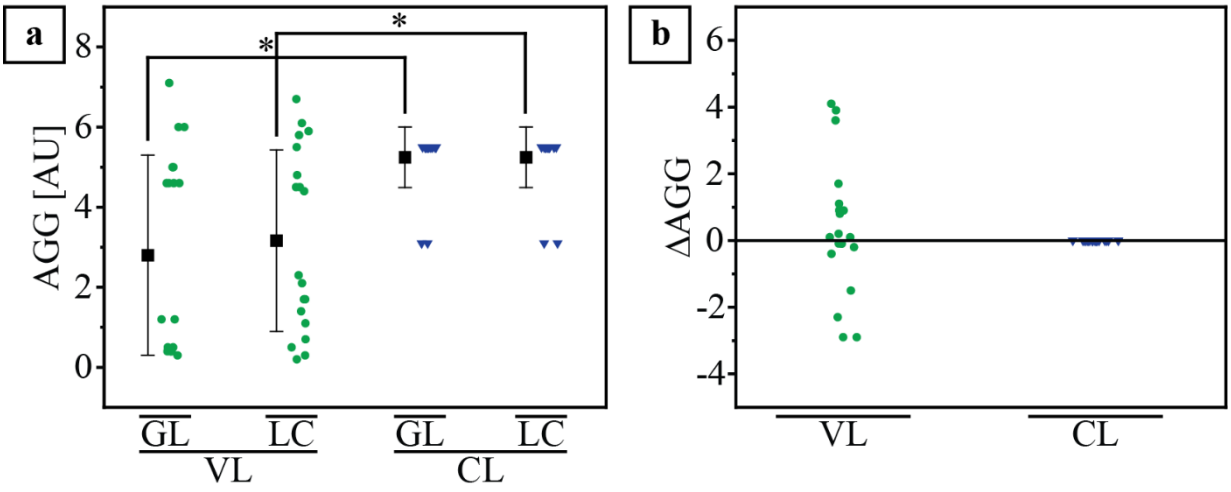


Figure 61: Aggregation score (AGG) of the V_L (green circles) and C_L domains (blue triangles). a) Absolut normalized AGG score (in arbitrary units) for the GL and LC sequences of the V_L and C_L domains. b) ΔAGG score of the V_L and C_L domains. ΔpI = LC – GL.

The mean AGG score was significantly higher for C_L domains than for V_L domains (GL: p = 5.3 x 10^{-4}, LC: p = 0.001) (Fig. 61 a, App. Tab. 25). The mean AGG score of the GL and LC V_L domains as well as the mean AGG score of the GL and LC C_L domains, did not

differ significantly (V_L: p = 0.637, C_L: p = 1.000) (Fig. 61 a, App. Tab. 25). The ΔAGG values of the V_L domains were similar to those of the AL proteins (mean ΔAGG: fragments = 0.6 ± 1.8, V_Ls = 0.4 ± 2.0) (Fig. 61 b, App. Tab. 25), supporting the assumption that the C_L attenuates the effect of mutations on the aggregation propensity.

In summary, there was no general difference between the AGG score of GL and LC sequences, indicating that mutations increase aggregation propensity only in individual cases. There is evidence that the C_L domain attenuates changes in aggregation propensity because the ΔAGG values of full-length proteins were less distributed and closer to 0.0 than those of the fragments and VL domains. The CL domains had the highest aggregation propensity on average, which contradicts the fact that fibrils are composed of AL proteins containing mainly the V_L domain. A possible explanation for this observation is that other parameters besides aggregation propensity affect and cause fibril formation. Furthermore, it suggests that proteolysis might occur after fibril formation and that the C_L domain contributes to LC aggregation.

4.3.2.4.5 GL-specific correlations between biochemical parameters and fibril formation

Since the biochemical parameters generally did not differ between GL and LC sequences, it was investigated whether it is GL-dependent which parameter changes upon mutation. For this purpose, cases were compared expressing LCs with the same GL V segment. Because AL proteins mostly showed greater changes in the biochemical properties than full-length LCs, the analysis focused on the AL proteins. In the 16 cases where the assignment of monoisotopic masses to LC fragments identified multiple AL proteins, parameters were determined for all AL proteins and averaged. AL proteins are referred to as LC fragments to avoid misunderstandings. The same calculations were performed for the GL fragments (Fig. 62, 63, App. Tab. 20, 22, 24, 26).

GL-specific changes were observed for IGLV1-44-derived LC fragments (Fig. 62, Tab. 20), as the mean pI of the IGLV1-44-derived LC fragments was significantly higher than that of the GL fragments (p = 0.032) (Fig. 62 a, App. Tab. 20). Five of six IGLV1-44-derived fragments (FOR102, FOR107, FOR137, FOR142, FOR147) had ΔpI values of 0.5 or higher (Fig. 62 b, App. Tab. 20), suggesting that mutations increase the pI of IGLV1-44-derived LCs and thus increase amyloidogenicity. However, there is evidence that other parameters also affect amyloidogenicity in IGLV1-44 derived LC fragments, as mutations did not affect the net charge in FOR128 (Fig. 62, App. Tab. 20).

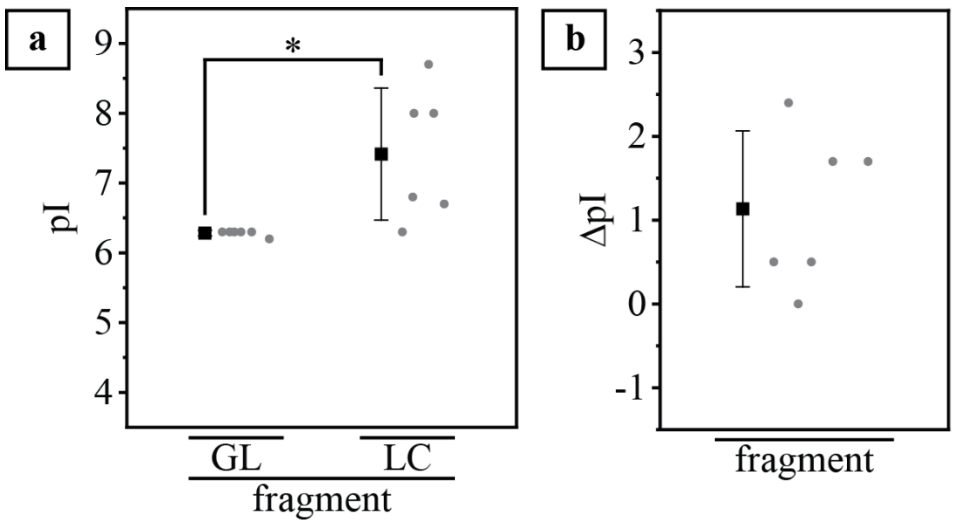


Figure 62: pI of GL and LC fragments (gray circles) derived from IGLV1-44. a) pI values of the GL and LC fragments. b) ΔpI values of the fragments. ΔpI = LC – GL.

GL-specific changes were also observed for IGLV3-19-derived LC fragments (Fig. 63 a, App. Tab. 26). All IGLV3-19-derived AL proteins had a 0.8 higher AGG score than the corresponding GL (Fig. 63 b, App. Tab. 26), suggesting that the aggregation propensity affects the amyloidogenicity of IGLV3-19-derived AL proteins. No other GL-dependent changes in biochemical properties were identified that influence aggregation and fibril formation.

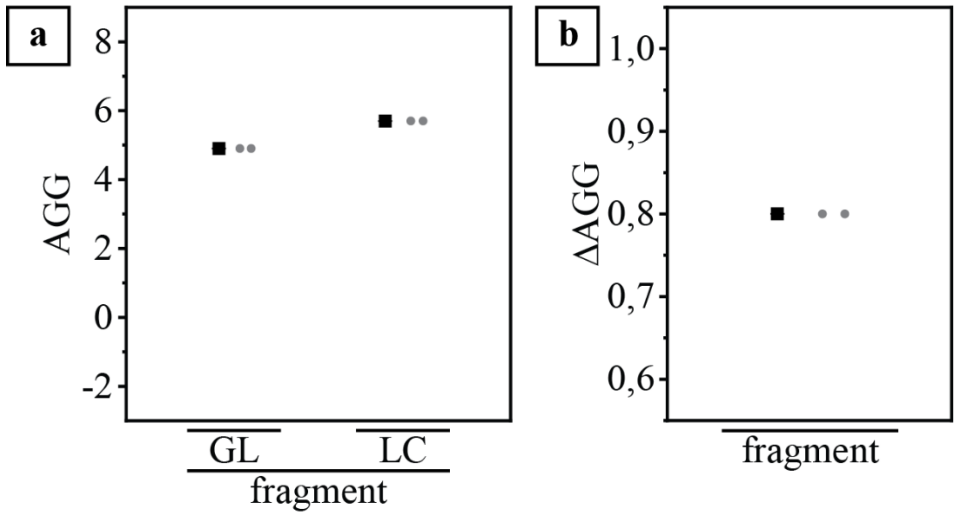


Figure 63: Aggregation score (AGG) of GL and LC fragments (gray circles) derived from IGLV3-19. a) Absolut normalized AGG values (in arbitrary units) for the GL and LC fragments. b) ΔAGG values of the fragments. ΔpI = LC – GL.

In summary, the bioinformatic analyses suggest that no parameter generally changes upon mutation, and it is case-dependent whether a parameter increases amyloidogenicity. Even at the GL level, it is unlikely that a single biochemical parameter fosters aggregation and fibril formation. However, the cohort of each GL V segment was small, making general conclusions difficult. In two cases (FOR012, FOR1007), none of the investigated parameters changed substantially due to mutations in the GL sequence (App. Tab. 19 – 26), indicating that other biochemical parameters or structural changes foster aggregation and fibril formation here.

5. Discussion

The aim of this work was to analyze LCs and AL proteins from different patients to identify commonalities in mutations, PTMs, and truncations that provide evidence of what makes a LC amyloidogenic. To this end, methods for purifying LCs from urine with high and low albumin loads and AL proteins from blood were developed to provide information on LC truncation and the timing of proteolysis. Subsequently, AL proteins were extracted from abdominal fat and heart tissue, and their primary structure was identified using MS and cDNA from Heidelberg. For this purpose, a method was established to extract AL protein from low amounts of abdominal fat, and to identify the amino acid sequence of the AL proteins by MS-based protein sequencing and total mass analysis. Primary structures of the AL proteins could be determined in 19 cases and were compared concerning mutations, PTMs, and truncations. In addition, a bioinformatic analysis was performed to identify the effect of mutations on net charge, solubility, hydrophobicity, and aggregation propensity.

5.1 Characterization of LCs and their derivatives from serum and urine

5.1.1 Purification of FLCs from urine and investigation of their oligomeric state

LCs are formed in the bone marrow, released into the blood, and excreted in the urine as BJPs [6–8]. In systemic AL amyloidosis, some LCs remain in the body and form amyloid fibrils [5]. However, these amyloid deposits consist only of LC fragments [59,60,79,127], and it is still controversial when fragmentation occurs and whether it plays a role in the amyloidogenicity of LCs. Morgan et al. suggested that a disrupted dimer interface makes LCs more susceptible to proteolysis [51], indicating that proteolysis can occur before fibril formation if the BJPs in the urine of AL amyloidosis patients do not form dimers. However, analyzing BJPs is hampered

by albumin, especially in patients with albuminuria, as albumin is tightly bound to the LCs and difficult to remove [125]. To address this problem, in this work, a workflow combining ammonium sulfate precipitation and IEC was established that allows separating LCs from albumin in urine with high and low albumin loads. LCs were purified from the urine of two AL amyloidosis patients U1 (λ LC, low albumin load) and U2 (κ LC, high albumin load), and analyzed for their oligomeric state by SEC. Only dimers were found in the λ case with low albumin load (U1), which were stabilized via a disulfide bond. In the κ case with high albumin load (U2), dimers and monomers were found, and the dimers were not stabilized via a disulfide bond. These data suggest that stable λ LC dimers are present in urine with a low albumin load, whereas unstable κ LC dimers are present in urine with a high albumin load. Since unstable dimers are susceptible to proteolysis [51], the data indirectly suggests that proteolysis can occur before fibril formation. The results are consistent with previous studies showing that λ LCs have a higher tendency to form dimers than κ LCs and that λ LCs mostly form covalent dimers, while κ LCs mostly form non-covalent dimers [144,145]. Although those studies indicated that the LC isotype affects dimerization, the present study raised evidence that dimerization is also affected by the albumin load. Additional evidence was provided by Andrich et al., as they exclusively found dimers when investigating the oligomerization state of λ and κ BJPs from urine with low albumin load [125]. A high albumin load in the urine of AL amyloidosis patients could therefore be indicative of LCs with a disrupted dimer interface and the presence of monomers. Thus, there might also be a link between renal involvement and a disrupted dimer interface, as albuminuria is characteristic of AL amyloidosis patients with renal involvement [146,147]. However, more cases need to be investigated to draw general conclusions about the correlation between albumin load and a disrupted dimer interface. The present work provides the basis for such studies by establishing a purification strategy to separate BJPs from albumin.

5.1.2 Analysis of LC truncations in serum

Since the investigation of the BJPs only indirectly suggested that LC fragmentation could occur before fibril formation, more concrete evidence regarding the timing of proteolysis was sought in the blood. Previous studies at the Institute of Protein Biochemistry (Ulm University) indicated that AL proteins are present in the blood, suggesting proteolysis before fibril deposition. However, these results from a master's thesis could not be reproduced. Thus, the experiments were repeated in this work by establishing a purification method for AL proteins from the serum of four AL amyloidosis cases. The method was based on first purifying AL protein from cardiac tissue by IEC and RPC, then applying the same purification protocol to serum from the same AL amyloidosis patient, and analyzing the RPC fractions by MS. Special care was taken not to contaminate the serum sample with AL protein from cardiac tissue by cleaning the IEC and RPC columns after each purification. In addition, a blank run was always performed between the MS analyses of the heart and serum fractions to avoid cross-contamination. The investigations provided no evidence that the AL proteins found in the fibril deposits of the heart are also present in the blood, suggesting that proteolysis occurs only in the tissue of fibril deposition. However, those results have to be viewed critically, as it is also possible that the purification strategy was not able to detect AL protein in the blood. The prerequisite for the applied method is that AL proteins from fibril deposits and serum behave the same way during purification, that is, the AL proteins in the heart and serum have the same biochemical and biophysical properties. However, considering that the environmental conditions are very different in heart tissue and blood and that corresponding AL proteins might differ in the number of amino acids, they could behave differently during purification. Nevertheless, full-length LCs in fibril deposits were found in other studies [58,66,148], supporting that proteolysis occurs after fibril formation. Furthermore, no one has been able to

confirm significant populations of AL proteins in serum so far, which is also supportive of proteolysis after fibril deposition [149–151].

Although the investigation of body fluids did not provide clear information on the timing of proteolysis, there is evidence that proteolysis generally occurs after fibril formation. AL proteins are most likely due to proteases that cleave unstructured parts of the LCs protruding from the fibril core [58]. However, since it has been demonstrated that VL domains are more amyloidogenic than the full-length LCs [51,57] and that some amyloidogenic LCs have a disrupted dimer interface favoring proteolysis [51], proteolysis possibly precedes fibril formation in individual cases.

5.2 Establishment of a universal protocol to determine the primary structure of AL proteins extracted from abdominal fat and heart tissue

5.2.1 Extraction of amyloid fibrils for MS analysis

In addition to body fluids, the primary structure of AL proteins was investigated in this work to obtain further information on LC truncation and other properties that possibly influence fibril formation. As in any seldom disease, patient samples are extremely valuable and rare when studying AL amyloidosis. Therefore, a method had to be established to extract amyloid fibrils from small amounts of tissue without vital function that can be easily obtained. Abdominal fat is a tissue that can be obtained from all AL amyloidosis patients without aggravating disease progression and without complex surgery [129]. Because fibrils from abdominal fat have already been shown to be representative of fibril deposits in different organs of the same patient [14], it is a good alternative for cardiac or renal biopsies. However, the amount of abdominal fat varies considerably from patient to patient and affects the amount of aspirate and the degree of contamination. Sometimes, only small biopsies of a few micrograms were

obtained with a high degree of blood contamination. Therefore, finding a protocol that efficiently extracts pure amyloid fibrils from small amounts of abdominal fat was necessary. A prerequisite for reliable information on the AL proteins is evidence that the extracted LC fragments constitute the deposited amyloid fibrils and are no soluble degradation products. Therefore, intact amyloid fibrils had to be extracted, which can be verified by electron microscopy. The only extraction method that met all the requirements was the water extraction protocol of Annamalai et al. [14], which was adapted to increase the concentration of the extracted amyloid fibrils. The strength, but at the same time, the potential limitation of that extraction method is the use of collagenase to digest the connective tissue around the amyloid fibrils. Using collagenase was the most efficient method to extract substantial amounts of pure and intact amyloid fibrils, as non-enzymatic extraction often requires large amounts of patient tissue [152], requires the use of denaturing agents [66,73,103,153], and results in low amyloid levels and heavy contamination [73]. All attempts to expose the amyloid fibrils with organic solvents, detergents or mechanical homogenization instead of collagenase failed. The potential disadvantage of using collagenase is that it is controversial whether enzymatic fibril extraction leads to artifactual degradation of amyloid fibrils and thus affects the primary structure of AL proteins. However, an argument against artifactual degradation is that enzymatic water extraction has been applied to various types of amyloidosis in mice and humans. In all cases, MS identified a major fibril protein that extended well beyond the ordered fibril core and corresponded to either the full-length sequence or a fragment mainly described in the literature. Examples include other amyloidogenic proteins such as human Aβ 1-40 [130], murine and human serum amyloid A [131], and human transthyretin [132], but also AL fibrils with larger fragments than what constitutes the fibril core [79,127]. Many of the mentioned fibril structures have additional electron densities around the fibril core that may reflect unstructured regions of the amyloidogenic protein that would be accessible for enzymatic degradation. Consequently, enzymatic water extraction was considered reliable for obtaining intact amyloid fibrils from

various organs and biopsies. Fibrils were extracted from heart tissue using the common water extraction protocol [63] and from abdominal fat using the adapted water extraction protocol of the present work.

5.2.2 Primary structure determination

After extracting the amyloid fibrils, a method was established to determine the primary structure of the fibril-forming AL proteins. For this purpose, the Heidelberg University Hospital provided cDNA-derived amino acid sequences of the precursor LCs. In the past, most primary structures of AL proteins have been derived solely from cDNA without considering protein information such as truncations or other PTMs [22,37,38,66,69–72]. However, I could show that cDNA information often lacks N- or C-terminal information, especially in or adjacent to regions where primers anneal during sequencing. In addition, many cDNA-derived sequences contained amino acid uncertainties due to overlapping signals caused by oligoclonal differences in plasma cell populations [154,155]. Thus, it was mandatory to use protein-based information to confirm the cDNA-derived sequence, identify missing and uncertain amino acids, and provide information on PTMs. To this end, the monoisotopic mass of the extracted AL proteins was determined by MS and assigned to the cDNA-derived sequence provided by Heidelberg University Hospital. Incorrect assignments were avoided by only allowing a deviation of 0.1 Da or less between the theoretical mass of the cDNA-derived amino acid sequence and the experimental mass of the AL protein and assuming as few PTMs as possible. The assignments identified uncertain amino acids of the cDNA-derived sequence, as the experimental mass matched the theoretical mass only when assuming the correct amino acid. However, the amino acid uncertainties became a problem the more amino acids were possible at one position. In addition, the number of amino acid combinations grew exponentially as the number of amino acid uncertainties in the cDNA-derived sequence increased. Considering that two amino acids

are possible at five positions, this results in 32 amino acid combinations. If all 20 amino acids are possible at those five positions, the number of amino acid combinations increases to 3.2×10^6. Besides the uncertainties, it turned out that some cDNA sequences contained errors, as the monoisotopic mass of the AL proteins often did not match the theoretical mass of the sequence. Thus, it was necessary to become independent of the cDNA to expand the number of cases with resolved primary structure of the AL proteins. For this purpose, the amino acid sequence of the AL proteins was determined using MS-based protein *de novo* sequencing, replacing the cDNA-derived sequence. The principle of protein sequencing was that peptide fragments of the AL proteins were prepared and assigned to an amino acid sequence based on their mass and assuming different PTMs. The AL proteins used for protein sequencing were cut from a polyacrylamide gel, minimizing contamination. However, it is possible that some peptide fragments did not originate from the AL proteins and were thus incorrectly assigned to an amino acid sequence. In addition, some peptide fragments may have been overfitted by assuming too many PTMs, resulting in errors in the amino acid sequence. Therefore, each MS-based protein sequence was assigned to the monoisotopic mass of the corresponding AL protein to verify that the amino acid sequence was correct. Although MS-based protein sequencing expanded the number of identified primary structures of AL proteins, a limiting factor remained the amount of extracted AL protein from abdominal fat. Less than 50 % of all cases with abdominal fat eventually yielded sufficient AL protein for reliable identification of their primary structure. The reason for low AL protein concentrations was either the availability of too little abdominal fat or too little amyloid deposits in the tissue. Consequently, heart tissue remained the best tissue for extracting amyloid fibrils and determining the primary structure of the corresponding AL proteins by MS. Further studies are necessary to optimize the extraction methods from abdominal fat and to increase the efficiency of primary structure determination of AL proteins.

5.3 Determination and analysis of AL proteins' primary structures to identify parameters determining amyloidogenicity

5.3.1 Truncations in AL proteins

The primary structure determination of the AL proteins was successful in 19 cases and showed that all AL proteins consist of precursor LC fragments. In some cases, the MS analysis identified multiple AL proteins in the amyloid deposits, as multiple monoisotopic masses could be assigned to LC fragments. Sometimes, one monoisotopic mass could be assigned to multiple LC fragments, if amino acids with the same molecular mass were present at the N- and C-terminus of the fragments. Whether a monoisotopic mass resulted from different fragments or only from one AL protein remained unclear. All AL proteins had a molecular weight of 10 to 15 kDa. Almost all AL proteins consisted of the V_L domain and small parts of the C_L domain with heterogeneity at their N- and C-terminus, which has also been reported in previous MS studies [58,73]. The AL proteins of the κ cases (FOR1001, FOR1007) were shorter and consisted of only the V_L domain, which is in contrast to previous studies that reported κ AL proteins to consist exclusively of the C_L domain [67,156]. In agreement with previous studies [58,66,73], these results indicate the presence of ubiquitous truncation regions in the LCs. Comparing the identified AL proteins with the published fibril structures [59,60,79,127] shows that the AL proteins mainly correspond to the fibril core. This observation suggests that the fibril structure could be affected by mutations, as they occur predominantly in the V_L domain, which is present in all identified AL proteins. The fact that the AL proteins form the fibril core further suggests that the N- and C-terminal heterogeneity of AL proteins could be due to a structural effect of the amyloid fibrils. In other words, the parts of the AL proteins that reach out of the protective fibril core might be accessible to exoproteases and digested. Evidence in support is that the λ6-derived AL proteins of case FOR140 show no N-terminal heterogeneity, as it has been shown that the N-terminus of λ6-derived fibril proteins is protected in the core of

the amyloid fibrils [60]. The fact that no enzyme could be identified that cleaves exclusively in the ubiquitous truncation regions of the AL proteins suggests that the proteolytic cleavage of the C_L domain is also due to the structural effect of the amyloid fibrils. Thus, this data indicates that proteolysis of the precursor LCs occurs after fibril formation. Typically, V_L domains are more amyloidogenic than full-length LCs, and it has been suggested that proteolytic cleavage of the protective C_L domain is necessary for unfolding and aggregation [51,157,158]. Therefore it has been proposed that proteolytic digestion precedes fibril formation [43]. However, other studies indicated that full-length LCs are amyloidogenic if the C_L domain is destabilized and does not form interdomain interactions [57,123,124]. As a result, proteolysis is not mandatory for aggregation and fibril formation, which is consistent with other studies that found full-length LCs in fibril deposits [58,66,148].

In summary, this work confirms that the amyloid fibrils are composed of LC fragments, which have a ubiquitous truncation region and consist of the V_L and, in most cases, a small part of the C_L domain. In all investigated κ cases, the AL proteins terminate before the C_L, which might be a special feature of the κ isotype. There is evidence that N- and C-terminal variability, as well as proteolytic cleavage of the C_L domain, is caused by the structural effect of the amyloid fibrils, suggesting that proteolytic events occur after fibril formation.

5.3.2 PTMs in AL proteins

In addition to the truncation pattern, the primary structure of the AL proteins also provided information about other PTMs. All AL proteins of this work contained the conserved disulfide bond of a natively folded V_L domain and, in some cases, pyroglutamate formation and oxidation. In contrast, other studies found more PTMs, such as disulfide bonds, deamidation, oxidation, phosphorylation, glycosylation, and pyroglutamate formation [19,58,61,66,127]. A

problem with considering many PTMs during the assignment of the primary structure of AL proteins is that it can lead to overfitting the data. Therefore, it indicates the reliability of the present data that very few PTMs were necessary to identify the primary structure of the AL proteins in this work. Although glycosylation is a common PTM in κ LCs [53], there is no evidence of glycosylation in AL proteins of this study. The most interesting PTM found in this study is the consistently present disulfide bond of the V_L domain, which has also been found in all cryo-EM structures of AL fibrils published so far [59,60,79,127]. The conserved disulfide bond in AL proteins indicates that the precursor LCs adopt their native fold after synthesis, which is also supported by the fact that LCs were found in urine, as they cannot pass the quality control in the ER unless they are natively folded [81]. Thus, the disulfide bond suggests a universal mechanism for fibril formation with reduced conformational dynamics, as unfolding and refolding of the precursor LC must occur around the intact disulfide bond. Pyroglutamate formation only occurred in AL proteins with an intact N-terminus and Gln at position 1. There is evidence that pyroglutamate formation increases the aggregation propensity in Aβ because three charges are lost during the lactam ring formation [159,160]. However, the present data do not indicate that pyroglutamate formation affects fibril formation in AL amyloidosis, as the Gln was cleaved off in many AL proteins. Met oxidation occurred only rarely, suggesting that it also does not affect fibril formation. There is evidence that Met oxidation results from ESI [139] and could thus be an artifact of MS. However, this is contradicted by the rare occurrence of Met oxidation, as artifactual oxidation would be expected in all AL proteins containing Met. Consequently, Met oxidation may affect amyloidogenicity. Although there is evidence that glycosylation affects fibril formation [53,127], it seems to be very rare.

In summary, this work provides reliable information on the PTMs of AL proteins. The only evidence that PTMs affect fibril formation was provided by the conserved disulfide bond of natively folded V_L domains, as it is present in all AL proteins and suggests a central mechanism

of unfolding and refolding around the disulfide bond. There is no clear evidence that pyroglutamate formation or Met oxidation affects fibril formation.

5.3.3 Mutations affecting amyloidogenicity

5.3.3.1 GL assignment and identification of mutations

The primary structure of AL proteins also provided information about mutations in the amino acid sequence. In order to identify these mutations, GL segments were assigned to the precursor LC sequences serving as non-mutated references. The GL segment with the least deviation from the LC sequence was selected for the GL assignment so that no differences between GL and LC were mistakenly identified as mutations. Nevertheless, it is possible, although very unlikely, that a LC sequence is derived from a GL segment that does not share the highest similarity with the LC sequence and has been mutated more frequently. However, the probability that mutations are missed in such a case is much lower than the risk of identifying too many mutations by mistake. The GL assignment identified seven different V segments (IGLV1-40, IGLV1-44, IGLV2-14, IGLV3-1, IGLV3-19, IGLV6-57, IGLK1-33) in the precursor LCs of the 19 cases associated with AL amyloidosis [36–38,41,161,162]. Although previous studies have reported prevalent organ involvement for different GL V segments, this work provided no information regarding this issue because of the small cohort size of each GL V segment. Consistent with other publications, mutations of the precursor LCs cluster within or close to CDR1, CDR2, and CDR3 [22]. None of these mutations occurred consistently in the different cases, suggesting that most mutations are “noise” or, in other words, have no effect on the amyloidogenicity of the LCs. The data indicate that different mutations, depending on their position and the biochemical properties of the amino acid, affect the native structure of the LCs, their aggregation propensity, or the stability of the fibrils. Visualization of the mutations in the

native structure of a LC suggested that primarily FR mutations destabilize the LCs by either disrupting intramolecular interactions or altering the spatial orientation of the V_L and C_L domains. Previous studies have identified many mutations that increase the amyloidogenicity of LCs, which are located in the FRs [43,47–50,55,56]. However, some mutations that increase amyloidogenicity are also located in the CDRs [44,57], reflecting the diversity of mutations affecting fibril formation. Therefore, mutations from previous studies were compared with the mutations of LCs from the present work to identify mutations that increase their amyloidogenicity.

5.3.3.2 Mutations affecting amyloidogenicity

Mutations do not affect one general property, which determines the amyloidogenicity of LCs. Instead, multiple parameters have been identified in the past contributing to fibril formation. Many studies found mutations that reduce the thermodynamic stability of the LCs and increase their amyloidogenic potential [43,44,50]. Some of these reported positions are also mutated in the LCs of this study, for example, position 52 (FOR140, FOR157) [50], 53 (FOR010, FOR103) [44], 73 (FOR1001) [50], and 89 (FOR142) [50]. The mutation of Tyr to Phe at position 52 was previously reported in κ LCs [50]. In this study, however, the same mutation was found in two λ LCs (FOR140, FOR157), suggesting an effect on thermodynamic stability in both κ and λ LCs. The other mutations from this study did not correspond to the substitutions that affected the thermodynamic stability of the LCs. At position 89, however, the mutation of Ala to Thr was previously reported as destabilizing because it places a hydrophilic group in a tightly packed hydrophobic environment [50]. Since the substitution of Ala to Ser at position 89 in FOR142 has the same effect, there is evidence that both mutations reduce the thermodynamic stability of the LC.

A parameter closely related to decreased thermodynamic stability is the alteration of LC interactions by disrupting the dimer interface [33,49]. Three of the conserved amino acids that stabilize the dimer interface [25] are mutated in the precursor LCs of this study. The Leu at position 49 stabilizes the dimer interface via hydrophobic interactions [25] and is mutated to Ile in FOR138 and Phe in FOR107. However, the mutations do not change the biochemical properties of the amino acids, indicating no effect on the dimer interface. The Gln at position 41 stabilizes the dimer interface via a hydrogen bond [25] and is mutated in four cases (FOR102, FOR107, FOR137, FOR142) to His. There is no evidence that this mutation destabilizes the dimer interface, as His can also form hydrogen bonds. Tyr at position 92 also stabilizes the dimer interface via a hydrogen bond [25] and is mutated in FOR010 to Phe. The fact that Phe does not form hydrogen bonds suggests that this mutation destabilizes the dimer interface and increases the amyloidogenicity of LCs. It has been previously shown that the substitution of Asn by Ile at position 37, which is part of the dimer interface, also reduces the stability of LCs [49]. FOR1001 contained an amino acid substitution of Asn to Valine (Val) at this position, which is similar to the previously described mutation, suggesting that both mutations destabilize LCs. Mutating an amino acid at position 46 to Ala also disrupts the dimer interface and has been reported to be a key determinant of LC toxicity [33]. FOR140 contained a mutation from Ser to Ala at position 46, which may thus disrupt the dimer interface and contribute to LC toxicity.

Another parameter that affects amyloidogenicity is increasing conformational dynamics in regions maintaining the dimer interface of LCs [22,43,44,57]. There is evidence that the orientation of the loop between β-strand C and C' affects dimerization [22,49]. In this study, two LCs (FOR103, FOR1007) contained a mutation at the start of that loop at position 43 from Pro to Ser. Pro prevents α-helix and β-strand formation [143], suggesting that the mutation

could lead to an elongated C β-strand, which displaces the loop and disrupts the dimer interface of the precursor LCs.

In summary, amyloidogenicity is closely related to the destabilization of the LCs. Mutations destabilize the LCs either by directly reducing their thermodynamic stability, disrupting their dimer interface, or increasing conformational dynamics. Although mutations of 7 LCs (FOR010, FOR103, FOR140, FOR142, FOR157, FOR1001, FOR1007) could be correlated with reduced stability of the LCs, 12 LCs remain where it is unknown which mutations affect amyloidogenicity. For this reason, the LCs must contain other mutations that have not yet been associated with reduced stability but affect the amyloidogenicity of LCs.

5.3.3.3 Junctional diversity and other rare mutations

There is evidence that mutations in the VJ junction could affect amyloidogenicity, as they occurred in more than 40 % of all 19 cases. Junctional diversity leads to insertions or deletions in the CDR3, suggesting that those mutations alter the structure of the CDR3 loop. These structural changes could disrupt intramolecular interactions, destabilize the LCs, and even alter binding interfaces. Evidence in support is that the CDRs have been shown to be important for the correct folding of the V_L domain and the interaction with V_H domains [163]. The proximity of the VJ junction to the linker region between the V_L and C_L domain suggests that amino acid deletions in the VJ junction could also destabilize the LCs by shortening the linker region and changing the relative orientation of the V_L and C_L domain. However, until now, it has only been reported that an extended linker region alters the relative orientation of the V_L and C_L domain and destabilizes the LC [57]. It is unlikely that the insertion of non-encoded amino acids in the V-J junction elongates the linker because the polypeptide chain is more likely to extend into the CDR3 loop. However, this does not apply to the unique insertion at positions 112 and 113 in

FOR1001, as it is closer to the linker region than to the CDR3. Thus, the insertion suggests an extension of the linker region, leading to a destabilization of the LC. Another mutation located in the linker region, which occurred in three cases (FOR012, FOR138, FOR140), was the substitution of Gly by Ser at the first position of the C_L domain. The fact that Ser can form hydrogen bonds suggests new intramolecular interactions that alter the relative orientation of the V_L and C_L domains and destabilize the LCs. Weber et al. showed that the Arg at the first position of the C_L domain in κ LCs is important to stabilize the linker region and thus regulates the orientation of the V_L and C_L domains [57]. Although the first amino acid of the C_L domain differs in κ and λ LCs, the present data suggest that the first position of the C_L domain in λ LCs is also important for the stability of the linker region. Apart from mutations in the linker region and the VJ junction, there are several other positions where mutations have altered biochemical properties, which have not yet been associated with destabilization. However, these mutations could also increase amyloidogenicity, as it has been shown that even rare mutations destabilize LCs [50]. Examples are the substitution of Gln to Arg at position 17 (FOR124) and the substitution of Gly to Pro at position 108 (FOR161).

In summary, there is evidence that mutations in the VJ junction or the linker region can destabilize the LCs by disrupting interdomain interactions. In addition, many other mutations in conserved regions of the LCs lead to biochemical changes, but it is not yet known whether and how they affect the stability of LCs.

5.3.3.4 GL-specific mutations

In addition, certain mutations only occurred in LCs with the same V segment. All IGKV1-33-derived precursor LCs contained two mutations, the substitution of Ser by Gly at position 33 and Lys by Asn at position 48. The mutations of Ser to Gly at position 26, Gln to

His at position 41, and any amino acid to Pro at position 104 were exclusively observed in IGLV1-44 derived LCs. In all IGLV3-1-derived LCs, Cys was replaced by Ser at position 37. The mutations of the IGKV1-33-derived and the IGLV1-44-derived LCs change the biochemical properties at a given position, suggesting that they reduce LC stability by affecting intramolecular interactions. The fact that the mutation to Pro at position 104 in IGLV1-44-derived LCs is located in a C-terminal β-strand of the V_L domain further suggests that it increases conformational dynamics by breaking the β-strand. Supposing that the Cys at position 37 in IGLV3-1-derived LCs could form a disulfide bond with other LCs, the mutation to Ser could disrupt intermolecular interactions and thus destabilize the LCs. These observations suggest that some mutations increase amyloidogenicity in a GL-dependent manner, which is consistent with previous studies that found mutations accumulating at different positions depending on the GL [22].

In summary, these results suggest that the unfolding of LCs is caused by different mechanisms and mutations. The mutations are GL-specific or occur at random positions in conserved regions or CDRs. Depending on their position and property, these mutations destabilize the LCs [123,164], alter the dimer interface [45], or alter inter-domain interactions [57]. Insertions and deletions occurring either in the V-J junction or in parts of the linker region could also destabilize the LCs by either changing the relative orientation of the V_L and C_L domain or affecting the native fold of LCs. However, *in vitro* studies analyzing the stability of LC mutants are necessary to determine which mutations destabilize LCs and increase amyloidogenicity.

5.3.4 The effect of biochemical parameters on the amyloidogenicity of LCs

In addition to the direct effect of mutations on the structure of LCs or their intermolecular interactions, it is suggested that mutations also affect biochemical properties such as net charge,

solubility, hydrophobicity, and aggregation propensity [165–167]. Therefore, LC and GL sequences of this work were analyzed for these biochemical parameters using bioinformatic algorithms. In some cases, the bioinformatic analyses have already been published [168]. The published results indicated that mutations significantly increase the net charge of the LCs and destabilize them [168]. However, after analyzing all 19 cases, no significant correlation between mutations and an increased net charge was found. The mutations also had no significant effect on solubility, hydrophobicity, and aggregation propensity. Only in individual cases, the parameters differed substantially between GL and LC. At the GL level, the net charge increased significantly from the IGLV1-44 GL to the IGLV1-44-derived LC sequences, and the aggregation propensity increased substantially from the IGLV3-19 GL to the IGLV3-19-derived LC sequence. The fragments had a significantly higher solubility and a significantly lower aggregation propensity than the full-length sequence, suggesting that the full-length LCs form amyloid fibrils and proteolytic cleavage occurs after deposition. Altogether, the biochemical properties varied substantially in the different cases, which is consistent with previous studies [60,66,149,169] and indicates that mutations do not generally affect one of the investigated parameters. Although there is evidence that mutations affect the parameters in individual cases or specific GLs, it is questionable whether these changes can trigger LC unfolding and fibril formation. Instead, it appears that AL amyloidosis is comparable to protein folding. According to the Levinthal paradox, protein folding never occurs when the protein undergoes all conformations. Instead, there are folding pathways that a protein goes through until it reaches the energetically favored native fold. The same applies to the factors of amyloidogenicity in AL amyloidosis, as mutations affect a plethora of parameters of precursor LCs. In addition, an interplay of mutations could also contribute to amyloidogenicity. Consequently, global parameters such as net charge, solubility, hydrophobicity, or aggregation propensity may not be sufficient to explain the amyloidogenicity of LCs. Instead, a detailed

investigation of the individual mutations is needed to identify pathways, as in protein folding, that will finally solve the mystery of amyloidogenicity and fibril formation in AL amyloidosis.

5.4 Conclusion

In this study, one central aim was to characterize LCs and their derivatives from body fluids and provide information on the timing of LC proteolysis. For this purpose, a protocol to separate BJPs and albumin was successfully developed, which allowed the investigation of the oligomerization state of free LCs. Furthermore, a purification method was established to investigate the existence of AL proteins in serum. Although the experiments provided no clear evidence on the timing of LC proteolysis, the data suggest that proteolytic digestion occurs after fibril formation, although in individual cases, proteolysis may precede it.

Another aim was to establish a workflow to determine the primary structure of AL proteins. Despite low AL protein yields and errors in the provided cDNA sequences from Heidelberg, it was possible to achieve this aim by establishing a protocol for efficient amyloid fibril extraction from abdominal fat and combining MS-based protein and cDNA information for the primary structure determination. The application of this workflow identified the primary structure of the AL proteins in 19 different cases and demonstrated the importance of verifying cDNA sequences with MS.

After determining the primary structure of the AL proteins, the aim was to identify parameters determining amyloidogenicity. Analyzing the truncations of the AL proteins suggested that LC proteolysis occurs due to a structural effect of the fibrils, takes place after fibril formation, and has no impact on amyloidogenicity. The presence of a conserved disulfide bond in all identified AL proteins suggested that disulfide bonding affects fibril formation by reducing conformational dynamics during the unfolding and misfolding of the LCs. The study provided no evidence that other PTMs affect the amyloidogenicity of LCs. Analyzing the precursor LC

sequences identified several mutations possibly contributing to amyloidogenicity by destabilizing the LCs, disrupting intermolecular interactions, or increasing conformational dynamics. The mutations had no general effect on the net charge, solubility, hydrophobicity, and aggregation propensity, indicating that such global parameters are not sufficient to explain the amyloidogenicity of LCs.

6. Summary

Systemic AL amyloidosis is a rare disease that originates from an abnormally proliferating plasma cell producing an excess of a patient-specific LC. These free LCs are secreted into the blood and excreted as BJPs in the urine. In systemic AL amyloidosis, however, some LCs deposit as amyloid fibrils in different organs leading to diverse pathology. Studies have shown that the fibril-forming AL proteins are fragments of the precursor LCs containing PTMs and mutations. Until today it is unknown whether proteolysis of the LCs occurs before or after fibril formation and if fragmentation affects amyloidogenicity. There is evidence that primary structural elements destabilize the LCs and create aggregation-prone regions that eventually lead to fibril formation. However, most information on the primary structure of LCs and AL proteins is derived from cDNA sequences, which are often incomplete and lack information about truncation sites and other PTMs of the AL proteins. Furthermore, only a few mutations have been related to an increased amyloidogenicity of LCs. Consequently, it is still largely unknown which concrete properties and mutations increase the amyloidogenicity of LCs.

This study addresses the question of the timing of LC proteolysis by analyzing the body fluids of AL amyloidosis patients for LCs and their fragments. For this purpose, a method was established to purify BJPs from urine with high and low albumin loads and separate the LCs from albumin. The LCs were analyzed for their oligomeric state, as a previous study suggested that a disrupted dimer interface makes LCs more susceptible to proteolysis, indicating that proteolysis can occur before fibril formation if the LCs in the urine do not form dimers. The fact that monomeric LCs were found in urine with high albumin load suggested that albuminuria could be indicative of disrupted LC dimers and that LC proteolysis can precede fibril formation. As a next step, a method was established to investigate the existence of AL protein in serum. Detecting no AL protein in the serum of four AL amyloidosis patients provided evidence that proteolysis does not occur in the blood but after fibril deposition. In combination with the urine

analysis, the results suggested that proteolysis might occur before fibril formation in LCs with a disrupted dimer interface but generally occurs after fibril formation.

This study also addresses the question of what makes a LC amyloidogenic by establishing a method to extract amyloid fibrils from abdominal fat more efficiently and to determine the primary structure of AL proteins by MS and cDNA. The primary structure of the AL proteins could be identified in 17 λ and 2 κ cases, despite some problems with low yields of extracted AL protein and errors in the provided cDNA from Heidelberg University Hospital. The determination of the primary structures showed that cDNA was often incomplete and erroneous, highlighting the importance of cDNA validation by MS. The analysis of the AL proteins identified ubiquitous truncation sites and heterogeneity at their N- and C-terminus. Almost all AL proteins terminated in the N-terminal β-strand of the C_L domain, while especially the κ AL proteins were shorter and terminated in the V_L domain. An intact N-terminus was observed in AL proteins from a λ6-derived LC, indicating that N- and C-terminal heterogeneity is due to a structural effect of the amyloid fibrils because it has been shown previously that the N-terminus of λ6-derived fibril proteins is protected in the fibril core. The fact that no enzyme could be identified, cleaving the LCs exclusively in the ubiquitous truncation regions, suggested that proteolytic processing is also due to the structural effects of the amyloid fibrils and thus occurs after fibril formation. The PTM analysis of the AL proteins revealed the presence of the conserved disulfide bond of a natively folded V_L domain, oxidation, and pyroglutamate formation. The disulfide bond occurred in all AL proteins suggesting a general mechanism of fibril formation with reduced conformational dynamics, as unfolding and misfolding must occur around the intact disulfide bond.

In the next step, mutations in the LC sequences were identified by assigning a GL sequence as a non-mutated reference. The number of mutations differed among the cases suggesting that it is the position and properties of the amino acid rather than the number of mutations affecting

amyloidogenicity. No common mutation site could be identified in all LC sequences but in LCs derived from the same GL V segment. The GL-specific mutations and other mutations in the FRs, CDRs, and VJ junction suggested that they increase the amyloidogenicity of LCs by reducing thermodynamic stability, disrupting the dimer interface, or altering inter-domain interactions. A bioinformatic analysis of the GL and LC sequences was conducted to investigate whether mutations also affect biochemical properties such as net charge, solubility, hydrophobicity, or aggregation propensity and thus increase the amyloidogenicity of LCs. However, the parameters differed not significantly between the GL and LC sequences suggesting that mutations have no general effect. Only in individual cases, the parameters changed substantially upon mutation. These data suggest that global parameters such as net charge, solubility, hydrophobicity, or aggregation propensity are not sufficient to explain the amyloidogenicity of LCs and that it is a combination of several factors affecting fibril formation in a diverse way.

7. References

[1] Perfetti V, Vignarelli MC, Casarini S, et al. Biological features of the clone involved in primary amyloidosis (AL). *Leuk 2001 152*. 2001;15:195–202.

[2] Quock TP, Yan T, Chang E, et al. Epidemiology of AL amyloidosis: a real-world study using US claims data. *Blood Adv*. 2018;2:1046–1053.

[3] Shimazaki C. [Current diagnosis and treatment of AL amyloidosis in Japan: a nationwide epidemiological survey]. *Rinsho Ketsueki*. 2019;60:973–978.

[4] Wisniowski B, McLeod DSA, Adams R, et al. The epidemiology of amyloidosis in Queensland, Australia. *Br J Haematol*. 2019;186:829–836.

[5] Blancas-Mejia LM, Misra P, Dick CJ, et al. Immunoglobulin light chain amyloid aggregation. *Chem Commun*. 2018;54:10664–10674.

[6] Abraham RS, Clark RJ, Katzmann JA, et al. Quantitative analysis of serum free light chains. A new marker for the diagnostic evaluation of primary systemic amyloidosis. *Am J Clin Pathol*. 2003;119:274–278.

[7] Dittrich T, Kimmich C, Hegenbart U, et al. Prognosis and Staging of AL Amyloidosis. *Acta Haematol*. 2020;143:388–400.

[8] Deegan MJ. Bence jones proteins: nature, metabolism, detection and significance. *Ann Clin Lab Sci*. 1976;6:38–46.

[9] Merlini G, Bellotti V. Molecular Mechanisms of Amyloidosis. *N Engl J Med*. 2003;349:583–596.

[10] Blancas-Mejía LM, Ramirez-Alvarado M. Systemic amyloidoses. *Annu Rev Biochem*. 2013;82:745–774.

[11] Benson MD, Buxbaum JN, Eisenberg DS, et al. Amyloid nomenclature 2020: update and recommendations by the International Society of Amyloidosis (ISA) nomenclature committee. *Amyloid*. 2020;27:217–222.

[12] Buxbaum JN, Dispenzieri A, Eisenberg DS, et al. Amyloid nomenclature 2022: update, novel proteins, and recommendations by the International Society of Amyloidosis (ISA) Nomenclature Committee. *Amyloid*. 2022;29:213–219.

[13] Devata S, Hari P, Markelova N, et al. Detection of amyloid in abdominal fat pad aspirates in early amyloidosis: Role of electron microscopy and Congo red stained cell block sections. *Cytojournal*. 2011;8:11.

[14] Annamalai K, Liberta F, Vielberg MT, et al. Common Fibril Structures Imply Systemically Conserved Protein Misfolding Pathways In Vivo. *Angew Chemie - Int Ed*. 2017;56:7510–7514.

[15] Dispenzieri A, Gertz MA, Kyle RA, et al. Serum cardiac troponins and N-terminal pro-brain natriuretic peptide: a staging system for primary systemic amyloidosis. *J Clin Oncol*. 2004;22:3751–3757.

[16] Lilleness B, Ruberg FL, Mussinelli R, et al. Development and validation of a survival staging system incorporating BNP in patients with light chain amyloidosis. *Blood*. 2019;133:215–223.

[17] Merlini G, Stone MJ. Dangerous small B-cell clones. *Blood*. 2006;108:2520–2530.

[18] Wechalekar AD, Schonland SO, Kastritis E, et al. A European collaborative study of treatment outcomes in 346 patients with cardiac stage III AL amyloidosis. *Blood*. 2013;121:3420–3427.

[19] Bellotti V, Mangione P, Merlini G. Review: immunoglobulin light chain amyloidosis--

the archetype of structural and pathogenic variability. 2000;130:280–289.

[20] Lefranc MP, Lefranc G. Immunoglobulins or Antibodies: IMGT® Bridging Genes, Structures and Functions. *Biomedicines*. 2020;8:.

[21] Bork P, Holm L, Sander C. The immunoglobulin fold: Structural classification, sequence patterns and common core. *J Mol Biol*. 1994;242:309–320.

[22] Poshusta TL, Sikkink LA, Leung N, et al. Mutations in Specific Structural Regions of Immunoglobulin Light Chains Are Associated with Free Light Chain Levels in Patients with AL Amyloidosis. *PLoS One*. 2009;4:e5169.

[23] Clarke J, Cota E, Fowler SB, et al. Folding studies of immunoglobulin-like β-sandwich proteins suggest that they share a common folding pathway. *Structure*. 1999;7:1145–1153.

[24] Kaplan B, Livneh A, Sela BA. Immunoglobulin free light chain dimers in human diseases. *ScientificWorldJournal*. 2011;11:726–735.

[25] Schiffer M. Molecular anatomy and the pathological expression of antibody light chains. *Am J Pathol*. 1996;148:1339.

[26] Bourne PC, Ramsland PA, Shan L, et al. Three-dimensional structure of an immunoglobulin light-chain dimer with amyloidogenic properties. *Acta Crystallogr D Biol Crystallogr*. 2002;58:815–823.

[27] Lefranc MP. Nomenclature of the human immunoglobulin kappa (IGK) genes. *Exp Clin Immunogenet*. 2001;18:161–174.

[28] LeFranc MP. Nomenclature of the human immunoglobulin lambda (IGL) genes. *Exp Clin Immunogenet*. 2001;18:242–254.

[29] Mikocziova I, Greiff V, Sollid LM. Immunoglobulin germline gene variation and its

impact on human disease. *Genes Immun*. 2021;22:205–217.

[30] Market E, Papavasiliou FN. V(D)J Recombination and the Evolution of the Adaptive Immune System. *PLoS Biol*. 2003;1:.

[31] Gellert M. V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu Rev Biochem*. 2002;71:101–132.

[32] Hwang JK, Alt FW, Yeap L-S. Related Mechanisms of Antibody Somatic Hypermutation and Class Switch Recombination. *Microbiol Spectr*. 2015;3:.

[33] Garofalo M, Piccoli L, Romeo M, et al. Machine learning analyses of antibody somatic mutations predict immunoglobulin light chain toxicity. *Nat Commun 2021 121*. 2021;12:1–10.

[34] Herrera GA, Teng J, Turbat-Herrera EA, et al. Understanding Mesangial Pathobiology in AL-Amyloidosis and Monoclonal Ig Light Chain Deposition Disease. *Kidney Int Reports*. 2020;5:1870.

[35] Merlini G, Dispenzieri A, Sanchorawala V, et al. Systemic immunoglobulin light chain amyloidosis. *Nat Rev Dis Prim*. 2018;4:.

[36] Comenzo RL, Zhang Y, Martinez C, et al. The tropism of organ involvement in primary systemic amyloidosis: Contributions of Ig V L germ line gene use and clonal plasma cell burden. *Blood*. 2001;98:714–720.

[37] Abraham RS, Geyer SM, Price-Troska TL, et al. Immunoglobulin light chain variable (V) region genes influence clinical presentation and outcome in light chain-associated amyloidosis (AL). *Blood*. 2003;101:3801–3808.

[38] Berghaus N, Schreiner S, Granzow M, et al. Analysis of the complete lambda light chain germline usage in patients with AL amyloidosis and dominant heart or kidney

involvement. *PLoS One*. 2022;17:e0264407.

[39] Perfetti V, Casarini S, Palladini G, et al. Analysis of V(lambda)-J(lambda) expression in plasma cells from primary (AL) amyloidosis and normal bone marrow identifies 3r (lambdaIII) as a new amyloid-associated germline gene segment. *Blood*. 2002;100:948–953.

[40] Del Pozo-Yauner L, Becerril B, Ochoa-Leyva A, et al. The Structural Determinants of the Immunoglobulin Light Chain Amyloid Aggregation. *Phys Biol Proteins Pept Theory, Exp Simul*. 2015;1–28.

[41] Kourelis T V., Dasari S, Theis JD, et al. Clarifying immunoglobulin gene usage in systemic and localized immunoglobulin light-chain amyloidosis by mass spectrometry. *Blood*. 2017;129:299–306.

[42] Glenner GG, Ein D, Eanes ED, et al. Creation of "amyloid" fibrils from Bence Jones proteins in vitro. *Science*. 1971;174:712–714.

[43] Kazman P, Vielberg MT, Pulido Cendales MD, et al. Fatal amyloid formation in a patient's antibody light chain is caused by a single point mutation. *Elife*. 2020;9:.

[44] Rottenaicher GJ, Weber B, Rührnößl F, et al. Molecular mechanism of amyloidogenic mutations in hypervariable regions of antibody light chains. *J Biol Chem*. 2021;296:.

[45] Baden EM, Owen BAL, Peterson FC, et al. Altered Dimer Interface Decreases Stability in an Amyloidogenic Protein. *J Biol Chem*. 2008;283:15853.

[46] Del Pozo-Yauner L, Ortiz E, Sánchez R, et al. Influence of the germline sequence on the thermodynamic stability and fibrillogenicity of human lambda 6 light chains. *Proteins*. 2008;72:684–692.

[47] Del Pozo-Yauner L, Wall JS, González Andrade M, et al. The N-terminal strand

modulates immunoglobulin light chain fibrillogenesis. *Biochem Biophys Res Commun*. 2014;443:495–499.

[48] Rennella E, Morgan GJ, Yan N, et al. The Role of Protein Thermodynamics and Primary Structure in Fibrillogenesis of Variable Domains from Immunoglobulin Light Chains. *J Am Chem Soc*. 2019;141:13562–13571.

[49] Baden EM, Randles EG, Aboagye AK, et al. Structural Insights into the Role of Mutations in Amyloidogenesis. *J Biol Chem*. 2008;283:30950.

[50] Hurle MR, Helms LR, Li L, et al. A role for destabilizing amino acid replacements in light-chain amyloidosis. *Proc Natl Acad Sci U S A*. 1994;91:5446.

[51] Morgan GJ, Kelly JW. The Kinetic Stability of a Full-Length Antibody Light Chain Dimer Determines whether Endoproteolysis Can Release Amyloidogenic Variable Domains. *J Mol Biol*. 2016;428:4280–4297.

[52] van der Kant R, Karow-Zwick AR, Van Durme J, et al. Prediction and Reduction of the Aggregation of Monoclonal Antibodies. *J Mol Biol*. 2017;429:1244–1261.

[53] Stevens FJ. Four structural risk factors identify most fibril-forming kappa light chains. *Amyloid*. 2000;7:200–211.

[54] Oberti L, Rognoni P, Barbiroli A, et al. Concurrent structural and biophysical traits link with immunoglobulin light chains amyloid propensity. *Sci Rep*. 2017;7:.

[55] Luna-Martínez OD, Hernández-Santoyo A, Villalba-Velázquez MI, et al. Stabilizing an amyloidogenic λ6 light chain variable domain. *FEBS J*. 2017;284:3702–3717.

[56] González-Andrade M, Becerril-Luján B, Sánchez-Lõpez R, et al. Mutational and genetic determinants of λ6 light chain amyloidogenesis. *FEBS J*. 2013;280:6173–6183.

[57] Weber B, Hora M, Kazman P, et al. The Antibody Light-Chain Linker Regulates Domain

Orientation and Amyloidogenicity. *J Mol Biol*. 2018;430:4925–4940.

[58] Mazzini G, Ricagno S, Caminito S, et al. Protease-sensitive regions in amyloid light chains: what a common pattern of fragmentation across organs suggests about aggregation. *FEBS J*. 2021;289:494–506.

[59] Radamaker L, Lin Y-HH, Annamalai K, et al. Cryo-EM structure of a light chain-derived amyloid fibril from a patient with systemic AL amyloidosis. *Nat Commun*. 2019;10:1–8.

[60] Swuec P, Lavatelli F, Tasaki M, et al. Cryo-EM structure of cardiac amyloid fibrils from an immunoglobulin light chain AL amyloidosis patient. *Nat Commun*. 2019;10:.

[61] Dispenzieri A, Larson DR, Rajkumar S V., et al. N-glycosylation of monoclonal light chains on routine MASS-FIX testing is a risk factor for MGUS progression. *Leuk 2020 3410*. 2020;34:2749–2753.

[62] Glenner G, Harada M, Isersky C, et al. Human amyloid protein: diversity and uniformity. *Biochem Biophys Res Commun*. 1970;41:1013–1019.

[63] Annamalai K, Gührs KH, Koehler R, et al. Polymorphism of Amyloid Fibrils in Vivo. *Angew Chemie Int Ed*. 2016;55:4822–4825.

[64] Sletten K, Natvig JB, Husby G, et al. The complete amino acid sequence of a prototype immunoglobulin-lambda light-chain-type amyloid-fibril protein AR. *Biochem J*. 1981;195:561–572.

[65] Solomon A, Frangione B, Franklin EC. Bence Jones proteins and light chains of immunoglobulins. Preferential association of the V(λVI) subgroup of human light chains with amyloidosis AL(λ). *J Clin Invest*. 1982;70:453–460.

[66] Lavatelli F, Perlman DH, Spencer B, et al. Amyloidogenic and associated proteins in

systemic amyloidosis proteome of adipose tissue. *Mol Cell Proteomics*. 2008;7:1570–1583.

[67] Solomon A, Weiss DT, Murphy CL, et al. Light chain-associated amyloid deposits comprised of a novel κ constant domain. *Proc Natl Acad Sci U S A*. 1998;95:9547–9551.

[68] Engvig JP, Olsen KE, Gislefoss RE, et al. Constant region of a kappa III immunoglobulin light chain as a major AL-amyloid protein. *Scand J Immunol*. 1998;48:92–98.

[69] Prokaeva T, Spencer B, Kaut M, et al. Soft tissue, joint, and bone manifestations of AL amyloidosis: clinical presentation, molecular features, and survival. *Arthritis Rheum*. 2007;56:3858–3868.

[70] Sikkink LA, Ramirez-Alvarado M. Biochemical and aggregation analysis of Bence Jones proteins from different light chain diseases. *Amyloid* . 2008;15:29.

[71] Comenzo RL, Wally J, Kica G, et al. Clonal immunoglobulin light chain variable region germline gene use in AL amyloidosis: association with dominant amyloid-related organ involvement and survival after stem cell transplantation. *Br J Haematol*. 1999;106:744–751.

[72] Wally J, Kica G, Zhang Y, et al. Identification of a novel substitution in the constant region of a gene coding for an amyloidogenic kappa1 light chain. *Biochim Biophys Acta*. 1999;1454:49–56.

[73] Lavatelli F, Mazzini G, Ricagno S, et al. Mass spectrometry characterization of light chain fragmentation sites in cardiac AL amyloidosis: insights into the timing of proteolysis. *J Biol Chem*. 2020;295:16572–16584.

[74] Hartl FU, Hayer-Hartl M. Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol*. 2009;16:574–581.

[75] DuBay KF, Pawar AP, Chiti F, et al. Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains. *J Mol Biol*. 2004;341:1317–1326.

[76] Breydo L, Redington JM, Uversky VN. Effects of Intrinsic and Extrinsic Factors on Aggregation of Physiologically Important Intrinsically Disordered Proteins. *Int Rev Cell Mol Biol*. 2017;329:145–185.

[77] Kazman P, Absmeier RM, Engelhardt H, et al. Dissection of the amyloid formation pathway in AL amyloidosis. *Nat Commun*. 2021;12:.

[78] He S, Scheres SHW. Helical reconstruction in RELION. *J Struct Biol*. 2017;198:163–176.

[79] Radamaker L, Baur J, Huhn S, et al. Cryo-EM reveals structural breaks in a patient-derived amyloid fibril from systemic AL amyloidosis. *Nat Commun*. 2021;12:875.

[80] Absmeier RM, Rottenaicher GJ, Svilenov HL, et al. Antibodies gone bad - the molecular mechanism of light chain amyloidosis. *FEBS J*. 2022;

[81] Chen JJ, Genereux JC, Wiseman RL. Endoplasmic reticulum quality control and systemic amyloid disease: Impacting protein stability from the inside out. *IUBMB Life*. 2015;67:404–413.

[82] Marin-Argany M, Lin Y, Misra P, et al. Cell Damage in Light Chain Amyloidosis: FIBRIL INTERNALIZATION, TOXICITY AND CELL-MEDIATED SEEDING. *J Biol Chem*. 2016;291:19813–19825.

[83] Teng J, Turbat-Herrera EA, Herrera GA. Extrusion of amyloid fibrils to the extracellular space in experimental mesangial AL-amyloidosis: transmission and scanning electron microscopy studies and correlation with renal biopsy observations. *Ultrastruct Pathol*. 2014;38:104–115.

[84] Selkoe DJ. Alzheimer's disease: genes, proteins, and therapy. *Physiol Rev*. 2001;81:741–766.

[85] Haass C. Take five—BACE and the γ-secretase quartet conduct Alzheimer's amyloid β-peptide generation. *EMBO J*. 2004;23:483.

[86] Sethi S, Vrana JA, Theis JD, et al. Laser microdissection and mass spectrometry-based proteomics aids the diagnosis and typing of renal amyloidosis. *Kidney Int*. 2012;82:226–234.

[87] Glish GL, Vachet RW. The basics of mass spectrometry in the twenty-first century. *Nat Rev Drug Discov*. 2003;2:140–150.

[88] Heck AJR, Van Den Heuvel RHH. Investigation of intact protein complexes by mass spectrometry. *Mass Spectrom Rev*. 2004;23:368–389.

[89] Ho CS, Lam CWK, Chan MHM, et al. Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications. *Clin Biochem Rev*. 2003;24:3.

[90] Kebarle P, Tang L. From Ions in Solution to Ions in the Gas Phase: The Mechanism of Electrospray Mass Spectrometry. *Anal Chem*. 1993;65:.

[91] Bruins AP. Mass spectrometry with ion sources operating at atmospheric pressure. *Mass Spectrom Rev*. 1991;10:53–77.

[92] Paul W, Steinwedel H. Notizen: Ein neues Massenspektrometer ohne Magnetfeld. *Zeitschrift Für Naturforsch*. 1953;A8:448–450.

[93] Dawson PH. Quadrupole mass analyzers: performance, design and some recent applications. *Mass Spectrom Rev*. 1986;5:1–37.

[94] Olsen J V., de Godoy LMF, Li G, et al. Parts per Million Mass Accuracy on an Orbitrap Mass Spectrometer via Lock Mass Injection into a C-trap. *Mol Cell Proteomics*.

2005;4:2010–2021.

[95] Scigelova M, Makarov A. Orbitrap mass analyzer--overview and applications in proteomics. *Proteomics*. 2006;6 Suppl 2:16–21.

[96] Makarov A. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal Chem*. 2000;72:1156–1162.

[97] McLafferty FW. Tandem mass spectrometry. *Science*. 1981;214:280–287.

[98] Coon JJ, Syka JEP, Shabanowitz J, et al. Tandem mass spectrometry for peptide and protein sequence analysis. *Biotechniques*. 2005;38:519–523.

[99] Holmes JL. Assigning structures to ions in the gas phase. *Org Mass Spectrom*. 1985;20:169–183.

[100] Olsen J V., Macek B, Lange O, et al. Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods*. 2007;4:709–712.

[101] Hazenberg BPC, Bijzet J, Limburg PC, et al. Diagnostic performance of amyloid A protein quantification in fat tissue of patients with clinical AA amyloidosis. *Amyloid*. 2007;14:133–140.

[102] Bligh EG, Dyer WJ. A rapid method of total lipid extraction and purification. *Can J Biochem Physiol*. 1959;37:911–917.

[103] Olsen KE, Sletten K, Westermark P. Extended analysis of AL-amyloid protein from abdominal wall subcutaneous fat biopsy: kappa IV immunoglobulin light chain. *Biochem Biophys Res Commun*. 1998;245:713–716.

[104] Khudyakov JI, Deyarmin JS, Hekman RM, et al. A sample preparation workflow for adipose tissue shotgun proteomics and proteogenomics. *Biol Open*. 2018;7:.

[105] Cañas B, Piñeiro C, Calvo E, et al. Trends in sample preparation for classical and second generation proteomics. *J Chromatogr A*. 2007;1153:235–258.

[106] Johnson M. Detergents: Triton X-100, Tween-20, and More. *Mater Methods*. 2013;3:.

[107] Pasing Y, Schniers A, Hansen T. Straightforward Protocol for Gel-Free Proteomic Analysis of Adipose Tissue. *Methods Mol Biol*. 2018;1788:289–296.

[108] Cladera J, Rigaud JL, Villaverde J, et al. Liposome solubilization and membrane protein reconstitution using Chaps and Chapso. *Eur J Biochem*. 1997;243:798–804.

[109] Goedert M, Spillantini MG, Cairns NJ, et al. Tau proteins of Alzheimer paired helical filaments: abnormal phosphorylation of all six brain isoforms. *Neuron*. 1992;8:159–168.

[110] van Dongen JA, Stevens HP, Parvizi M, et al. The fractionation of adipose tissue procedure to obtain stromal vascular fractions for regenerative purposes. *Wound Repair Regen*. 2016;24:994–1003.

[111] Wu Z, Roberts DS, Melby JA, et al. MASH Explorer: A Universal Software Environment for Top-Down Proteomics. *J Proteome Res*. 2020;19:3867–3876.

[112] Tran NH, Rahman MZ, He L, et al. Complete De Novo Assembly of Monoclonal Antibody Sequences. *Sci Rep*. 2016;6:.

[113] Niedermeyer THJ, Strohalm M. mMass as a Software Tool for the Annotation of Cyclic Peptide Tandem Mass Spectra. *PLoS One*. 2012;7:.

[114] Gasteiger E, Gattiker A, Hoogland C, et al. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*. 2003;31:3784–3788.

[115] Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–410.

[116] Giudicelli V, Chaume D, Lefranc MP. IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Res*. 2005;33:.

[117] Gasteiger E, Hoogland C, Gattiker A, et al. Protein Identification and Analysis Tools on the ExPASy Server. *Proteomics Protoc Handb*. 2005;571–607.

[118] Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol*. 1982;157:105–132.

[119] Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, et al. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotechnol*. 2004;22:1302–1306.

[120] Sormanni P, Aprile FA, Vendruscolo M. The CamSol Method of Rational Design of Protein Mutants with Enhanced Solubility. *J Mol Biol*. 2015;427:478–490.

[121] Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac Symp Biocomput*. 2002;310–322.

[122] Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera - A visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25:1605–1612.

[123] Weber B, Hora M, Kazman P, et al. Domain Interactions Determine the Amyloidogenicity of Antibody Light Chain Mutants. *J Mol Biol*. 2020;432:6187–6199.

[124] Rennella E, Morgan GJ, Kelly JW, et al. Role of domain interactions in the aggregation of full-length immunoglobulin light chains. *Proc Natl Acad Sci U S A*. 2019;116:854–863.

[125] Andrich K, Hegenbart U, Kimmich C, et al. Aggregation of Full-length Immunoglobulin Light Chains from Systemic Light Chain Amyloidosis (AL) Patients Is Remodeled by

Epigallocatechin-3-gallate. *J Biol Chem*. 2017;292:2328–2344.

[126] Fanali G, Di Masi A, Trezza V, et al. Human serum albumin: from bench to bedside. *Mol Aspects Med*. 2012;33:209–290.

[127] Radamaker L, Karimi-Farsijani S, Andreotti G, et al. Role of mutations and post-translational modifications in systemic AL amyloidosis studied by cryo-EM. *Nat Commun*. 2021;12:.

[128] Pepys MB, Rademacher TW, Amatayakul-Chantler S, et al. Human serum amyloid P component is an invariant constituent of amyloid deposits and has a uniquely homogeneous glycostructure. *Proc Natl Acad Sci U S A*. 1994;91:5602.

[129] Van Gameren II, Hazenberg BPC, Bijzet J, et al. Diagnostic accuracy of subcutaneous abdominal fat tissue aspiration for detecting systemic amyloidosis and its utility in clinical practice. *Arthritis Rheum*. 2006;54:2015–2021.

[130] Kollmer M, Close W, Funk L, et al. Cryo-EM structure and polymorphism of Aβ amyloid fibrils purified from Alzheimer's brain tissue. *Nat Commun*. 2019;10:.

[131] Liberta F, Loerch S, Rennegarbe M, et al. Cryo-EM fibril structures from systemic AA amyloidosis reveal the species complementarity of pathological amyloids. *Nat Commun*. 2019;10:.

[132] Schmidt M, Wiese S, Adak V, et al. Cryo-EM structure of a transthyretin-derived amyloid fibril from a patient with hereditary ATTR amyloidosis. *Nat Commun*. 2019;10:.

[133] Sun G, Palmer AF. Preparation of Ultrapure Bovine and Human Hemoglobin by Anion Exchange Chromatography. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2008;867:1.

[134] Fitzpatrick AWP, Falcon B, He S, et al. Cryo-EM structures of Tau filaments from

Alzheimer's disease brain. *Nature*. 2017;547:185.

[135] Glenner GG. Amyloid Deposits and Amyloidosis — The β-Fibrilloses — (First of Two Parts). *N Engl J Med*. 1980;302:1283–1292.

[136] Luptak J, Bista M, Fisher D, et al. Antibody fragments structurally enable a drug-discovery campaign on the cancer target Mcl-1. *Acta Crystallogr Sect D Struct Biol*. 2019;75:1003–1014.

[137] Chelius D, Jing K, Lueras A, et al. Formation of pyroglutamic acid from N-terminal glutamic acid in immunoglobulin gamma antibodies. *Anal Chem*. 2006;78:2370–2376.

[138] Vogt W. Oxidation of methionyl residues in proteins: Tools, targets, and reversal. *Free Radic Biol Med*. 1995;18:93–105.

[139] Chen M, Cook KD. Oxidation Artifacts in the Electrospray Mass Spectrometry of Aβ Peptide. *Anal Chem*. 2007;79:2031.

[140] Retter I, Althaus HH, Münch R, et al. VBASE2, an integrative V gene database. *Nucleic Acids Res*. 2005;33:.

[141] Howe KL, Achuthan P, Allen J, et al. Ensembl 2021. *Nucleic Acids Res*. 2021;49:D884–D891.

[142] Aken BL, Ayling S, Barrell D, et al. The Ensembl gene annotation system. *Database (Oxford)*. 2016;2016:.

[143] Chou PY, Fasman GD. Empirical predictions of protein conformation. *Annu Rev Biochem*. 1978;47:251–276.

[144] Sölling K. Polymeric forms of free light chains in serum from normal individuals and from patients with renal diseases. *Scand J Clin Lab Invest*. 1976;36:447–452.

[145] Sølling K. Light chain polymerism in normal individuals in patients with severe proteinuria and in normals with inhibited tubular protein reabsorption by lysine. *Scand J Clin Lab Invest*. 1980;40:129–134.

[146] Gertz MA, Comenzo R, Falk RH, et al. Definition of organ involvement and treatment response in immunoglobulin light chain amyloidosis (AL): A consensus opinion from the 10th International Symposium on Amyloid and Amyloidosis. *Am J Hematol*. 2005;79:319–328.

[147] Visram A, Al Saleh AS, Parmar H, et al. Correlation between urine ACR and 24-h proteinuria in a real-world cohort of systemic AL amyloidosis patients. *Blood Cancer J 2020 1012*. 2020;10:1–8.

[148] Buxbaum J. Mechanisms of disease: monoclonal immunoglobulin deposition. Amyloidosis, light chain deposition disease, and light and heavy chain deposition disease. *Hematol Oncol Clin North Am*. 1992;6:323–346.

[149] Kaplan B, Ramirez-Alvarado M, Dispenzieri A, et al. 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*. 2008;46:335–341.

[150] Eulitz M, Linke RP. The precursor molecule of A Vλ II-immunoglobulin light chain-derived amyloid fibril protein circulates precleaved. *Biochem Biophys Res Commun*. 1993;194:1427–1434.

[151] Lavatelli F, Brambilla F, Valentini V, et al. A novel approach for the purification and proteomic analysis of pathogenic immunoglobulin free light chains from serum. *Biochim Biophys Acta*. 2011;1814:409–419.

[152] Bergström J, Gustavsson Å, Hellman U, et al. Amyloid deposits in transthyretin-derived

amyloidosis: cleaved transthyretin is associated with distinct amyloid morphology. *J Pathol*. 2005;206:224–232.

[153] Enqvist S, Sletten K, Westermark P. Fibril protein fragmentation pattern in systemic AL-amyloidosis. *J Pathol*. 2009;219:473–480.

[154] Walker BA, Wardell CP, Melchor L, et al. Intraclonal heterogeneity is a critical early event in the development of myeloma and precedes the development of clinical symptoms. *Leukemia*. 2014;28:384–390.

[155] Rasche L, Chavan SS, Stephens OW, et al. Spatial genomic heterogeneity in multiple myeloma revealed by multi-region sequencing. *Nat Commun*. 2017;8:.

[156] Connors LH, Jiang Y, Budnik M, et al. Heterogeneity in primary structure, post-translational modifications, and germline gene usage of nine full-length amyloidogenic κ1 immunoglobulin light chains. *Biochemistry*. 2007;46:14259–14271.

[157] Brumshtein B, Esswein SR, Landau M, et al. Formation of amyloid fibers by monomeric light chain variable domains. *J Biol Chem*. 2014;289:27513–27525.

[158] Brumshtein B, Esswein SR, Sawaya MR, et al. Identification of two principal amyloid-driving segments in variable domains of Ig light chains in systemic light-chain amyloidosis. *J Biol Chem*. 2018;293:19659–19671.

[159] Schilling S, Lauber T, Schaupp M, et al. On the seeding and oligomerization of pGlu-amyloid peptides (in vitro). *Biochemistry*. 2006;45:12393–12399.

[160] Jawhar S, Wirths O, Bayer TA. Pyroglutamate Amyloid-β (Aβ): A Hatchet Man in Alzheimer Disease. *J Biol Chem*. 2011;286:38825.

[161] Gertz RA, Kyle MA. Primary systemic amyloidosis: clinical and laboratory features in 474 cases. *Semin Hematol*. 1995;32:45–59.

[162] Perfetti V, Palladini G, Casarini S, et al. The repertoire of λ light chains causing predominant amyloid heart involvement and identification of a preferentially involved germline gene, IGLV1-44. *Blood*. 2012;119:144–150.

[163] Herold EM, John C, Weber B, et al. Determinants of the assembly and function of antibody variable domains. *Sci Rep*. 2017;7:.

[164] Klimtchuk ES, Gursky O, Patel RS, et al. The critical role of the constant region in thermal stability and aggregation of amyloidogenic immunoglobulin light chain. *Biochemistry*. 2010;49:9848–9857.

[165] Schmittschmitt JP, Scholtz JM. The role of protein stability, solubility, and net charge in amyloid fibril formation. *Protein Sci*. 2003;12:2374–2378.

[166] Chiti F, Taddei N, Baroni F, et al. Kinetic partitioning of protein folding and aggregation. *Nat Struct Biol*. 2002;9:137–143.

[167] Chiti F, Stefani M, Taddei N, et al. Rationalization of the effects of mutations on peptide andprotein aggregation rates. *Nat 2003 4246950*. 2003;424:805–808.

[168] Baur J, Berghaus N, Schreiner S, et al. Identification of AL proteins from 10 λ-AL amyloidosis patients by mass spectrometry extracted from abdominal fat and heart tissue. *Amyloid*. 2022;

[169] Kaplan B, Livneh A, Gallo G. Charge differences between in vivo deposits in immunoglobulin light chain amyloidosis and non-amyloid light chain deposition disease. *Br J Haematol*. 2007;136:723–728.

8. Appendix

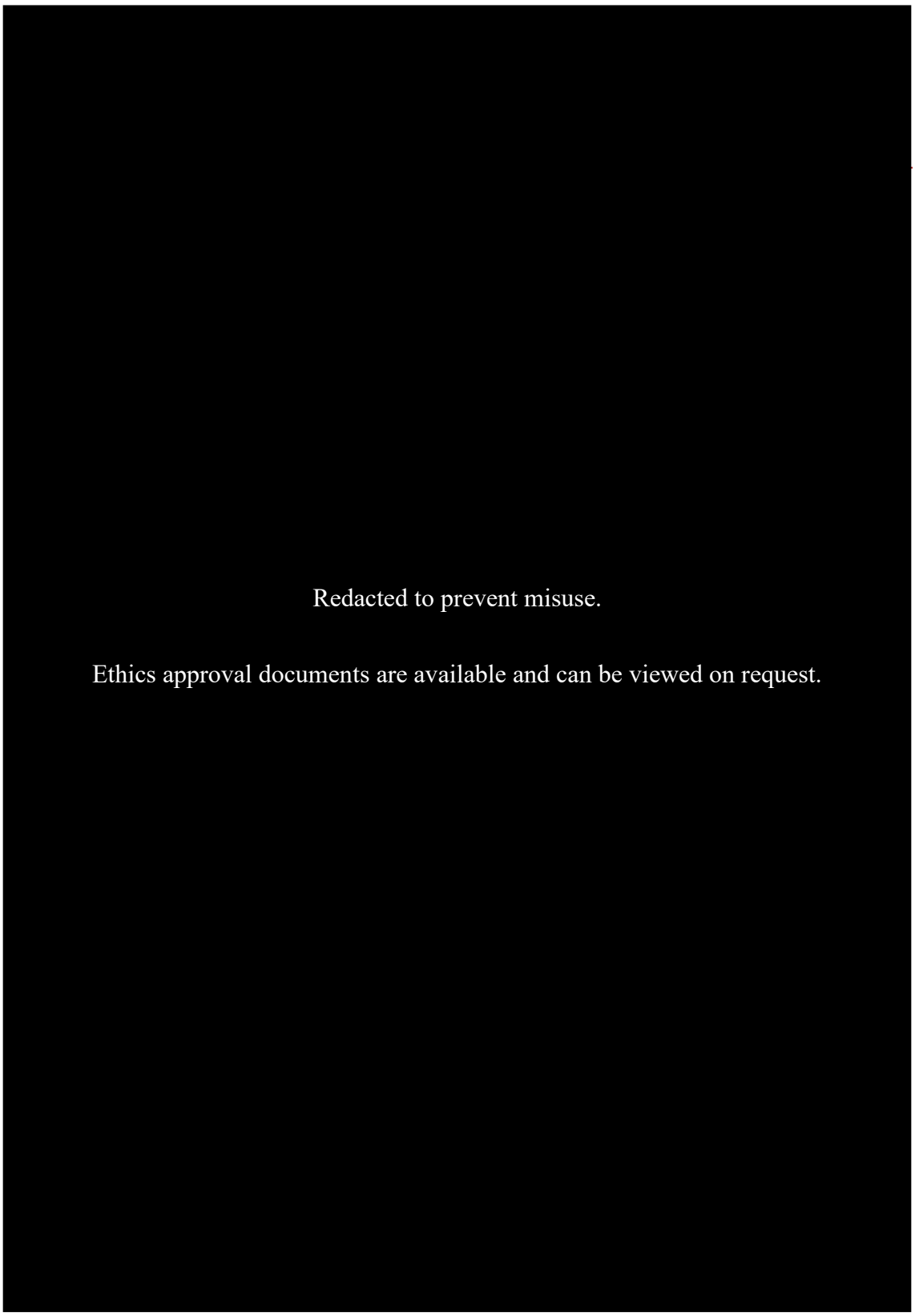


Figure 64: Ethics approval (first page) of Heidelberg Medical Faculty.



Figure 65: Ethics approval (second page) of Heidelberg Medical Faculty.



Figure 66: Ethics approval (first page) of Ulm University.

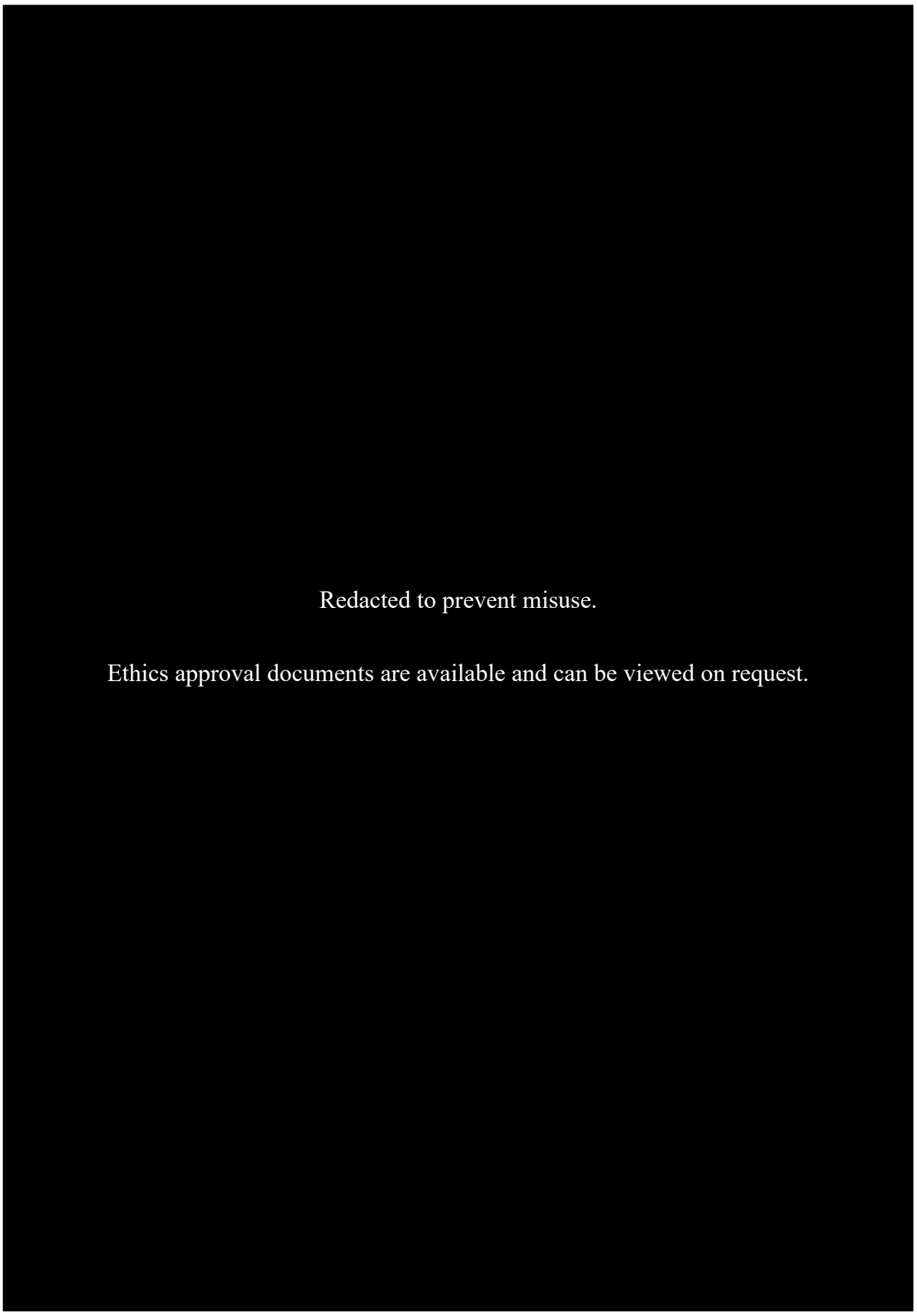


Figure 67: Ethics approval (second page) of Ulm University.

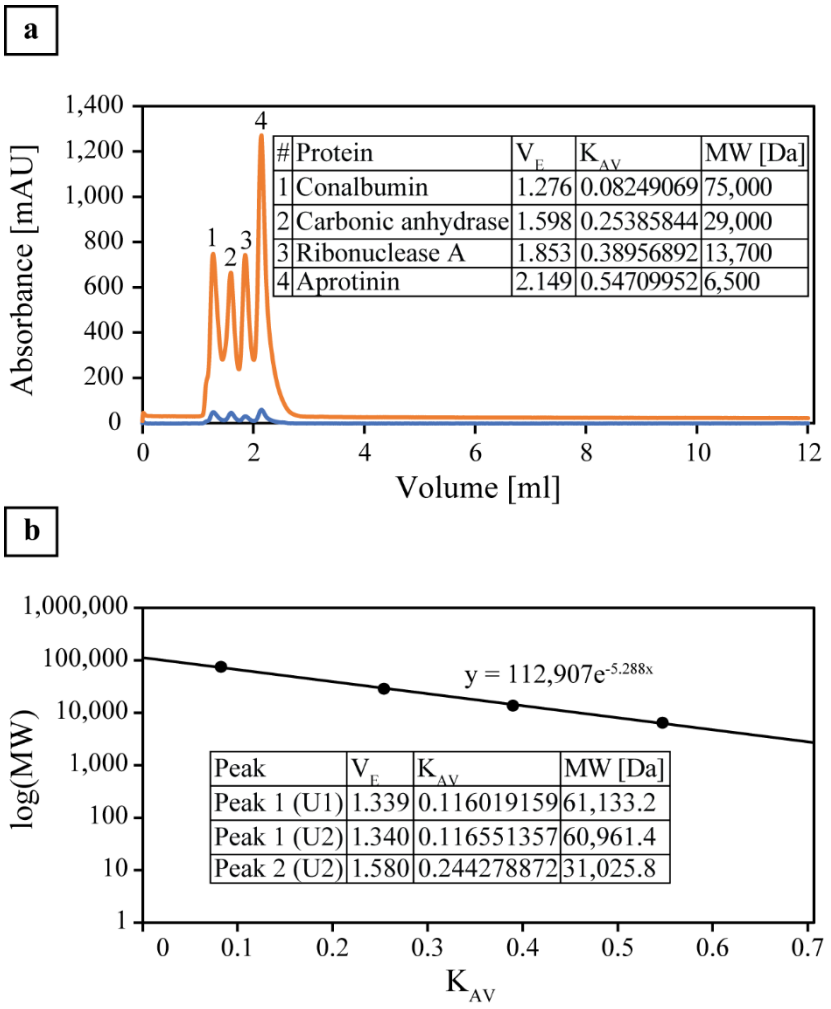


Figure 68: Mass determination using SEC. a) Elution pattern of proteins with known molecular weight. V_E, K_{AV}, and molecular weight (MW) in Dalton are illustrated for all standard proteins used. b) Calibration curve. Y-axis: decadic logarithm of the molecular weight. X-axis: K_{AV} values. The linear equation is illustrated above the curve. The corresponding calculated molecular weight of the different peaks from urine samples 1 and 2 are illustrated below the calibration curve.

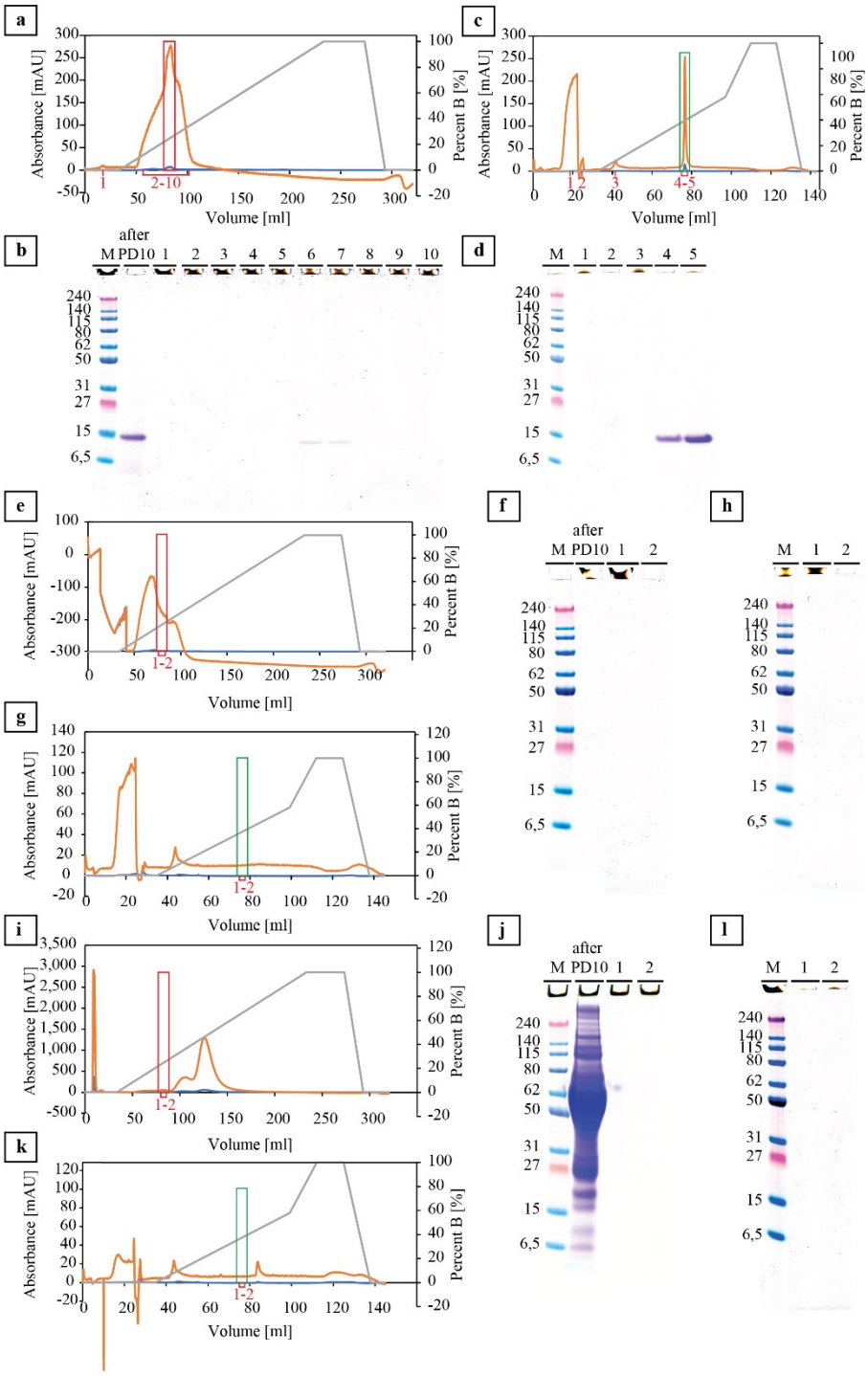


Figure 69: Purification procedure of FOR005. a) IEC of heart-extracted AL protein. Primary Y-axis: milli absorbance units (mAU) at 280 nm (blue curve) and 215 nm (orange curve). Secondary y-axis: percentage of elution buffer (gray curve). X-axis: elution volume in ml. Red rectangle: pooled fractions applied to RPC. Red numbers: fractions investigated by LDS-PAGE. b) LDS-PAGE analysis of the fractions from IEC in a). M: marker in kDa. After PD10: desalted sample before IEC. Gel columns from LDS-PAGE were removed for convenience only without changing protein bands. c) RPC purification of the pooled fractions from IEC in a). Green rectangle: pooled fractions sent to MS. d) LDS-PAGE analysis of the fractions from RPC in c). e) IEC purification of the buffer sample. f) LDS-PAGE analysis of the fractions from IEC in e). g) RPC purification of the pooled fractions from IEC in e). h) LDS-PAGE analysis of the fractions from RPC in g). i) IEC purification from blood serum. j) LDS-PAGE analysis of the fractions from IEC in i). k) RPC purification of the pooled fractions from IEC in i). l) LDS-PAGE analysis of the fractions from RPC in k).

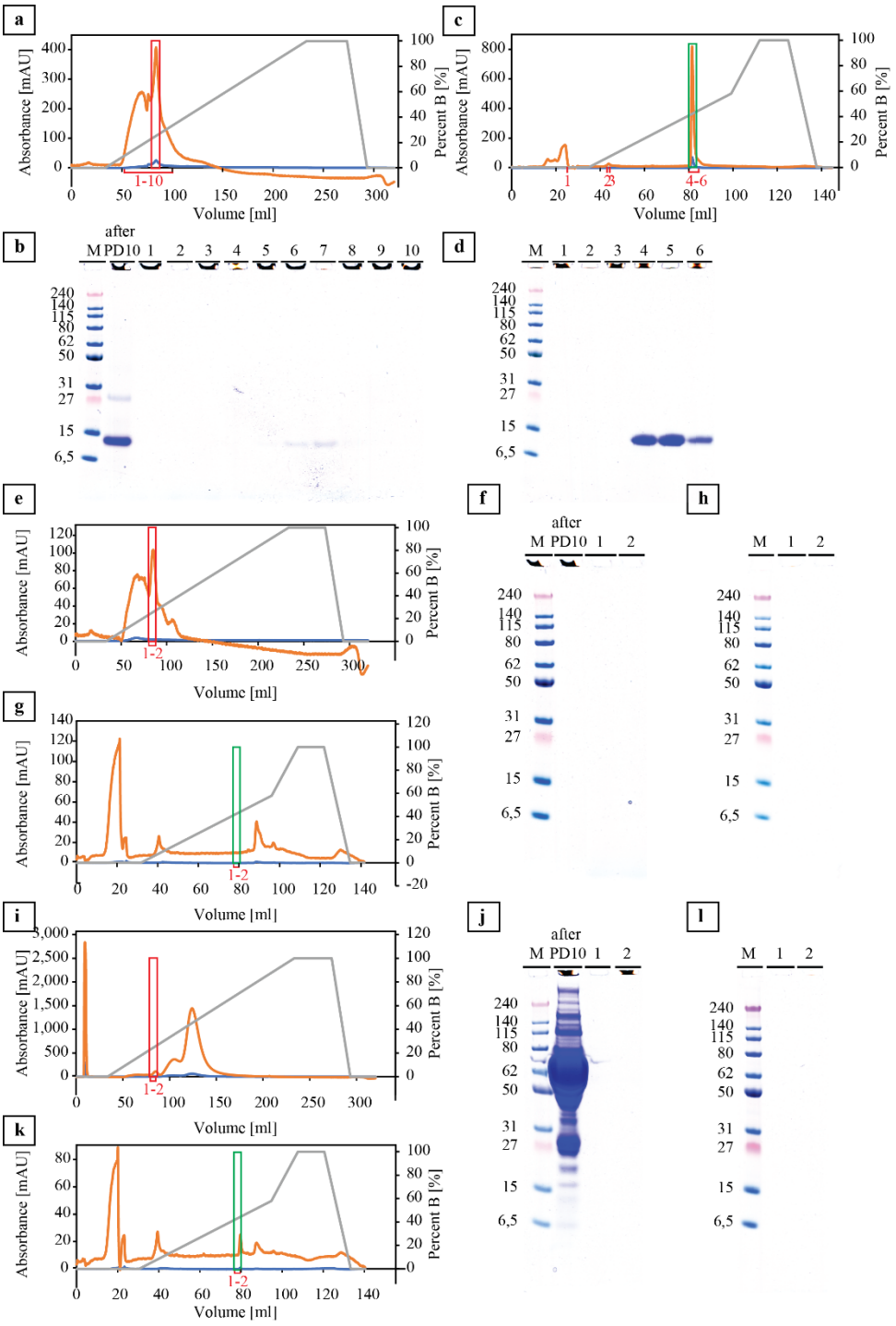


Figure 70: Purification procedure of FOR010. a) IEC of heart-extracted AL protein. Primary Y-axis: milli absorbance units (mAU) at 280 nm (blue curve) and 215 nm (orange curve). Secondary y-axis: percentage of elution buffer (gray curve). X-axis: elution volume in ml. Red rectangle: pooled fractions applied to RPC. Red numbers: fractions investigated by LDS-PAGE. b) LDS-PAGE analysis of the fractions from IEC in a). M: marker in kDa. After PD10: desalted sample before IEC. Gel columns from LDS-PAGE were removed for convenience only without changing protein bands. c) RPC purification of the pooled fractions from IEC in a). Green rectangle: pooled fractions sent to MS. d) LDS-PAGE analysis of the fractions from RPC in c). e) IEC purification of the buffer sample. f) LDS-PAGE analysis of the fractions from IEC in e). g) RPC purification of the pooled fractions from IEC in e). h) LDS-PAGE analysis of the fractions from RPC in g). i) IEC purification from blood serum. j) LDS-PAGE analysis of the fractions from IEC in i). k) RPC purification of the pooled fractions from IEC in i). l) LDS-PAGE analysis of the fractions from RPC in k).

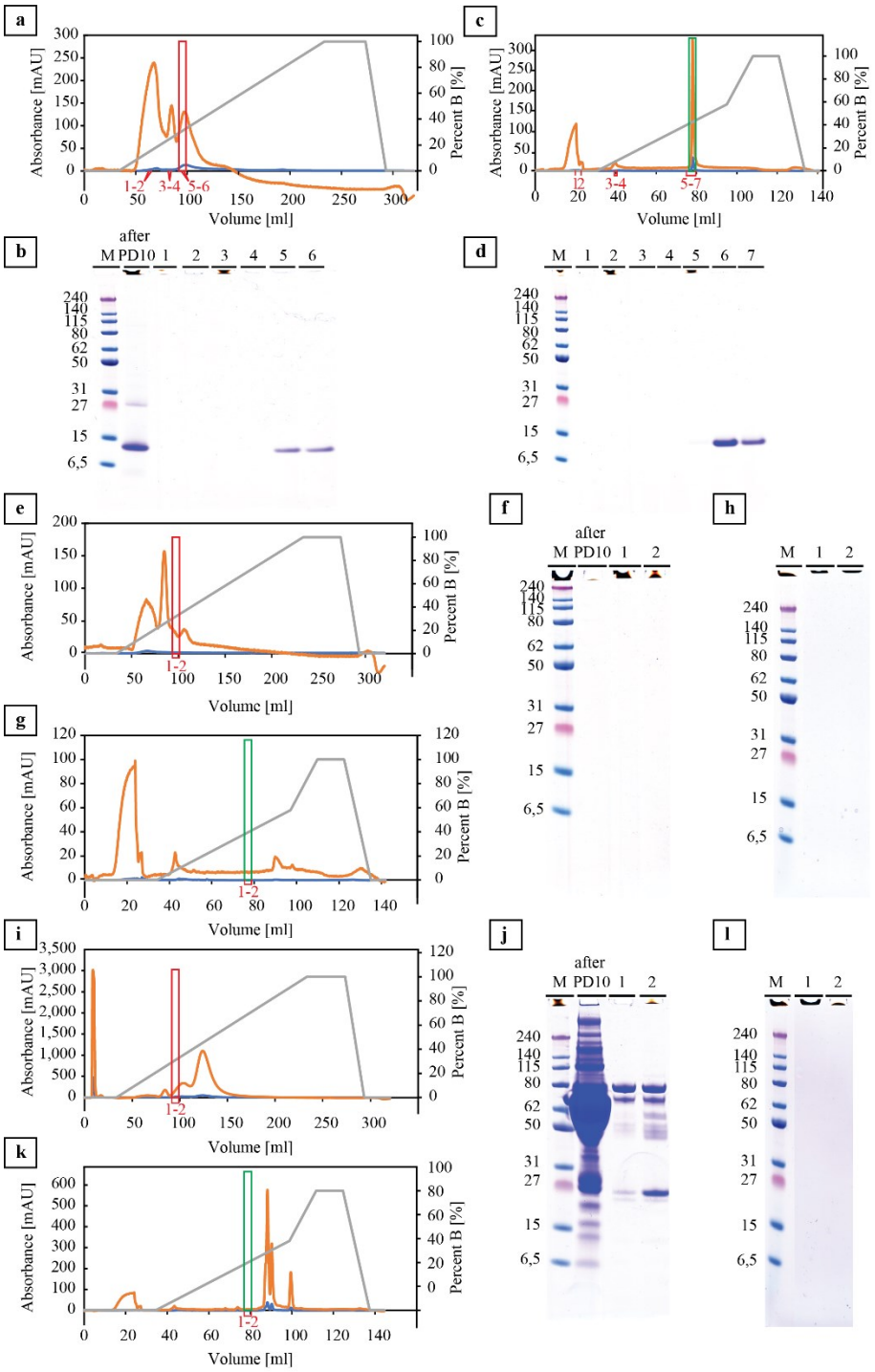


Figure 71: Purification procedure of FOR012. a) IEC of heart-extracted AL protein. Primary Y-axis: milli absorbance units (mAU) at 280 nm (blue curve) and 215 nm (orange curve). Secondary y-axis: percentage of elution buffer (gray curve). X-axis: elution volume in ml. Red rectangle: pooled fractions applied to RPC. Red numbers: fractions investigated by LDS-PAGE. b) LDS-PAGE analysis of the fractions from IEC in a). M: marker in kDa. After PD10: desalted sample before IEC. Gel columns from LDS-PAGE were removed for convenience only without changing protein bands. c) RPC purification of the pooled fractions from IEC in a). Green rectangle: pooled fractions sent to MS. d) LDS-PAGE analysis of the fractions from RPC in c). e) IEC purification of the buffer sample. f) LDS-PAGE analysis of the fractions from IEC in e). g) RPC purification of the pooled fractions from IEC in e). h) LDS-PAGE analysis of the fractions from RPC in g). i) IEC purification from blood serum. j) LDS-PAGE analysis of the fractions from IEC in i). k) RPC purification of the pooled fractions from IEC in i). l) LDS-PAGE analysis of the fractions from RPC in k).

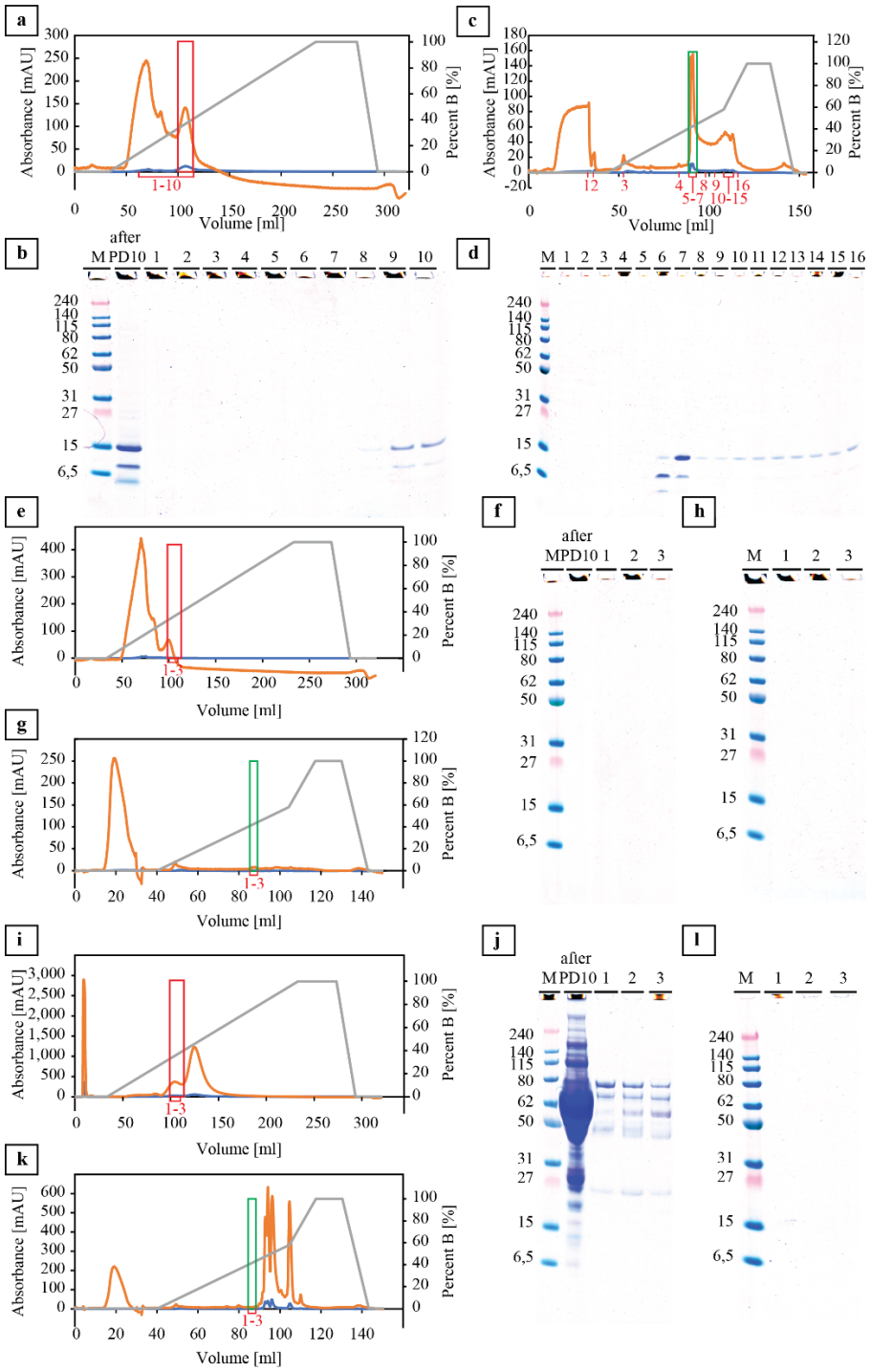


Figure 72: Purification procedure of FOR140. a) IEC of heart-extracted AL protein. Primary Y-axis: milli absorbance units (mAU) at 280 nm (blue curve) and 215 nm (orange curve). Secondary y-axis: percentage of elution buffer (gray curve). X-axis: elution volume in ml. Red rectangle: pooled fractions applied to RPC. Red numbers: fractions investigated by LDS-PAGE. b) LDS-PAGE analysis of the fractions from IEC in a). M: marker in kDa. After PD10: desalted sample before IEC. Gel columns from LDS-PAGE were removed for convenience only without changing protein bands. c) RPC purification of the pooled fractions from IEC in a). Green rectangle: pooled fractions sent to MS. d) LDS-PAGE analysis of the fractions from RPC in c). e) IEC purification of the buffer sample. f) LDS-PAGE analysis of the fractions from IEC in e). g) RPC purification of the pooled fractions from IEC in e). h) LDS-PAGE analysis of the fractions from RPC in g). i) IEC purification from blood serum. j) LDS-PAGE analysis of the fractions from IEC in i). k) RPC purification of the pooled fractions from IEC in i). l) LDS-PAGE analysis of the fractions from RPC in k).

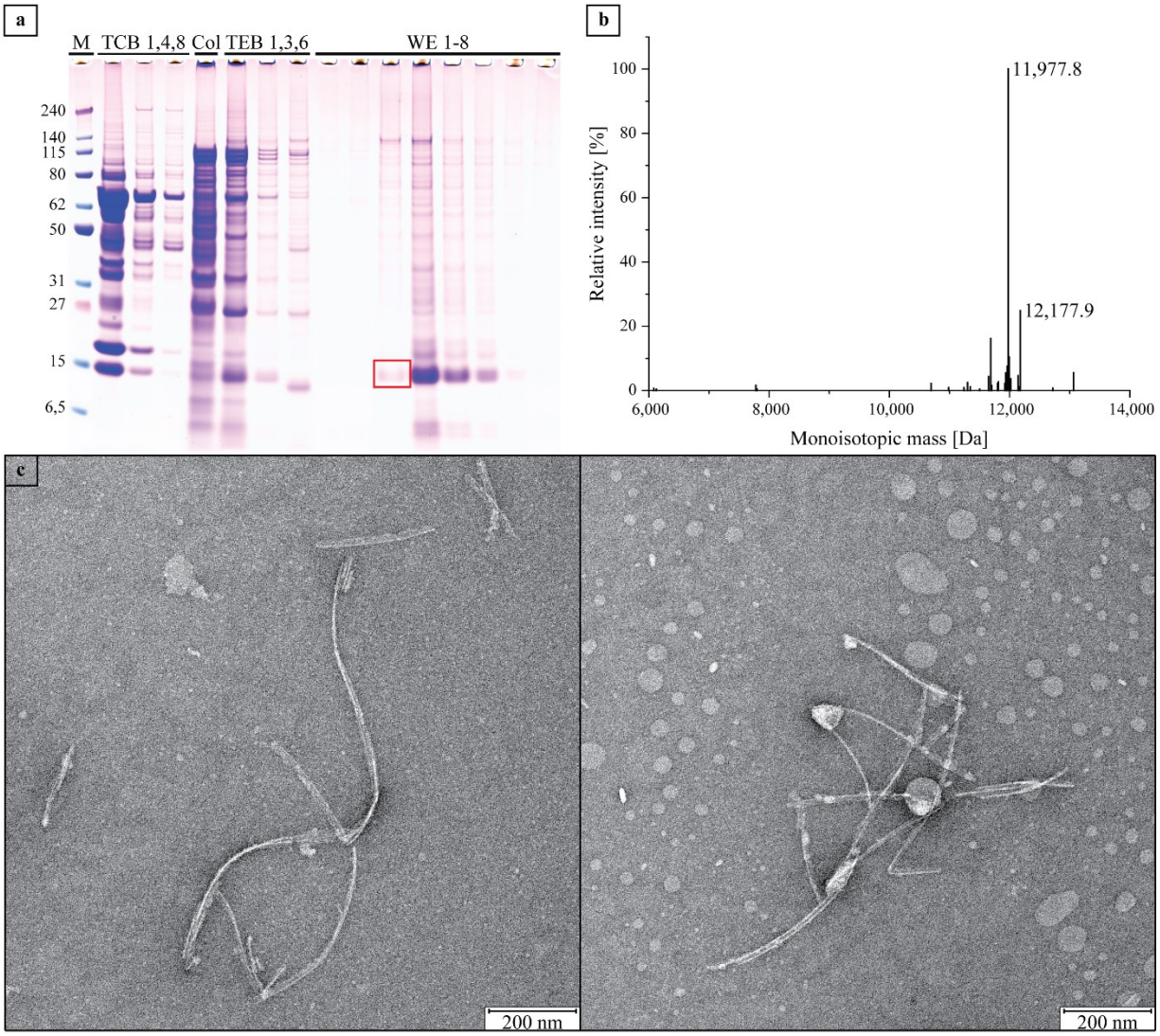


Figure 73: FOR009. a) LDS-PAGE of the water extraction. M: marker in kDa. Col: supernatant after collagenase incubation. WE: water extract. b) Deconvoluted mass spectrum of the extracted AL proteins. Y-axis: relative intensity in percent. X-axis: monoisotopic mass in Da. Number: mass assigned to a sequence. c) Negative stain TEM images (20 k magnification) of WE3 (red rectangle).

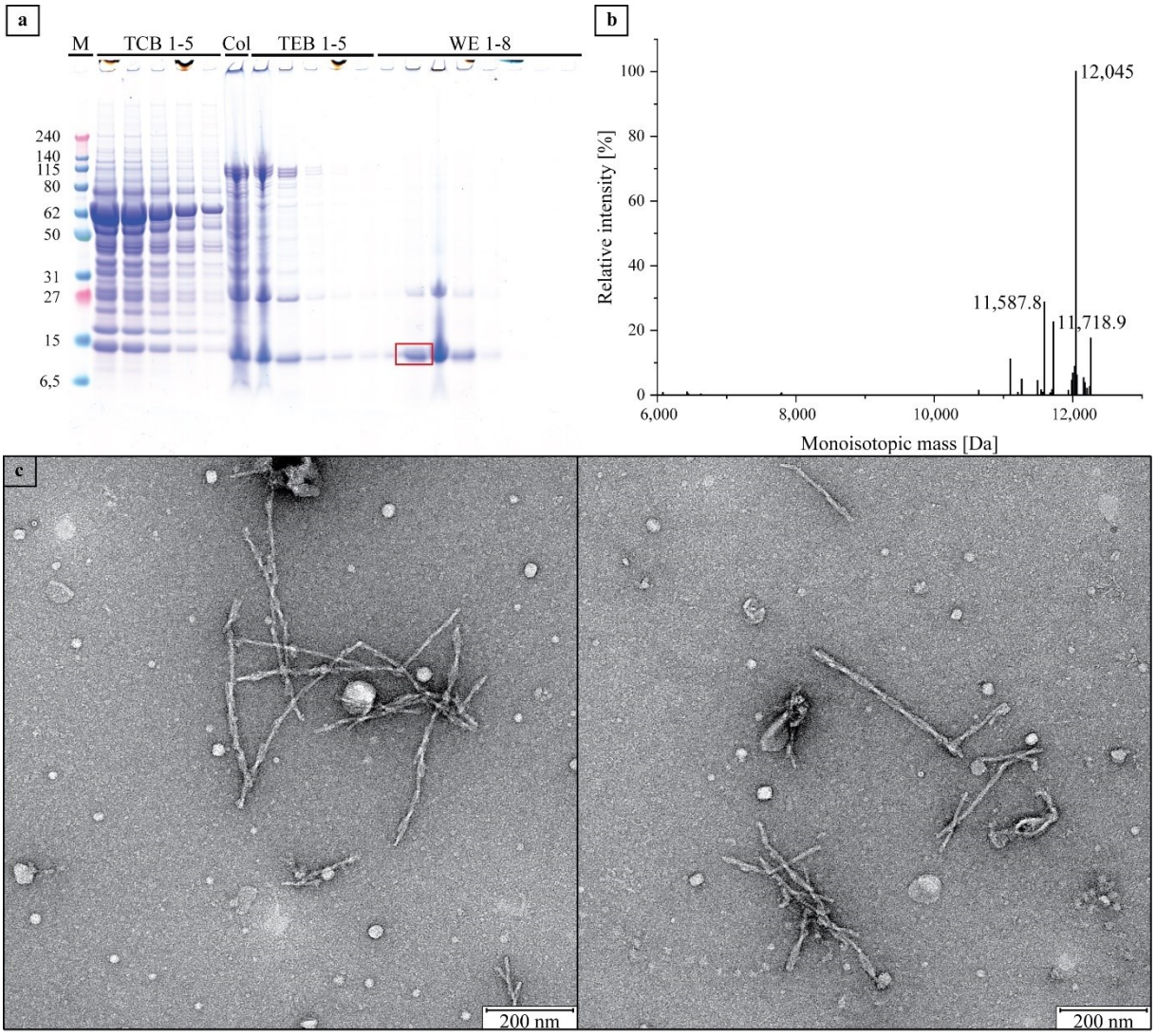


Figure 74: FOR010. a) LDS-PAGE of the water extraction. M: marker in kDa. Col: supernatant after collagenase incubation. WE: water extract. b) Deconvoluted mass spectrum of the extracted AL proteins. Y-axis: relative intensity in percent. X-axis: monoisotopic mass in Da. Number: mass assigned to a sequence. c) Negative stain TEM images (20 k magnification) of WE2 (red rectangle).

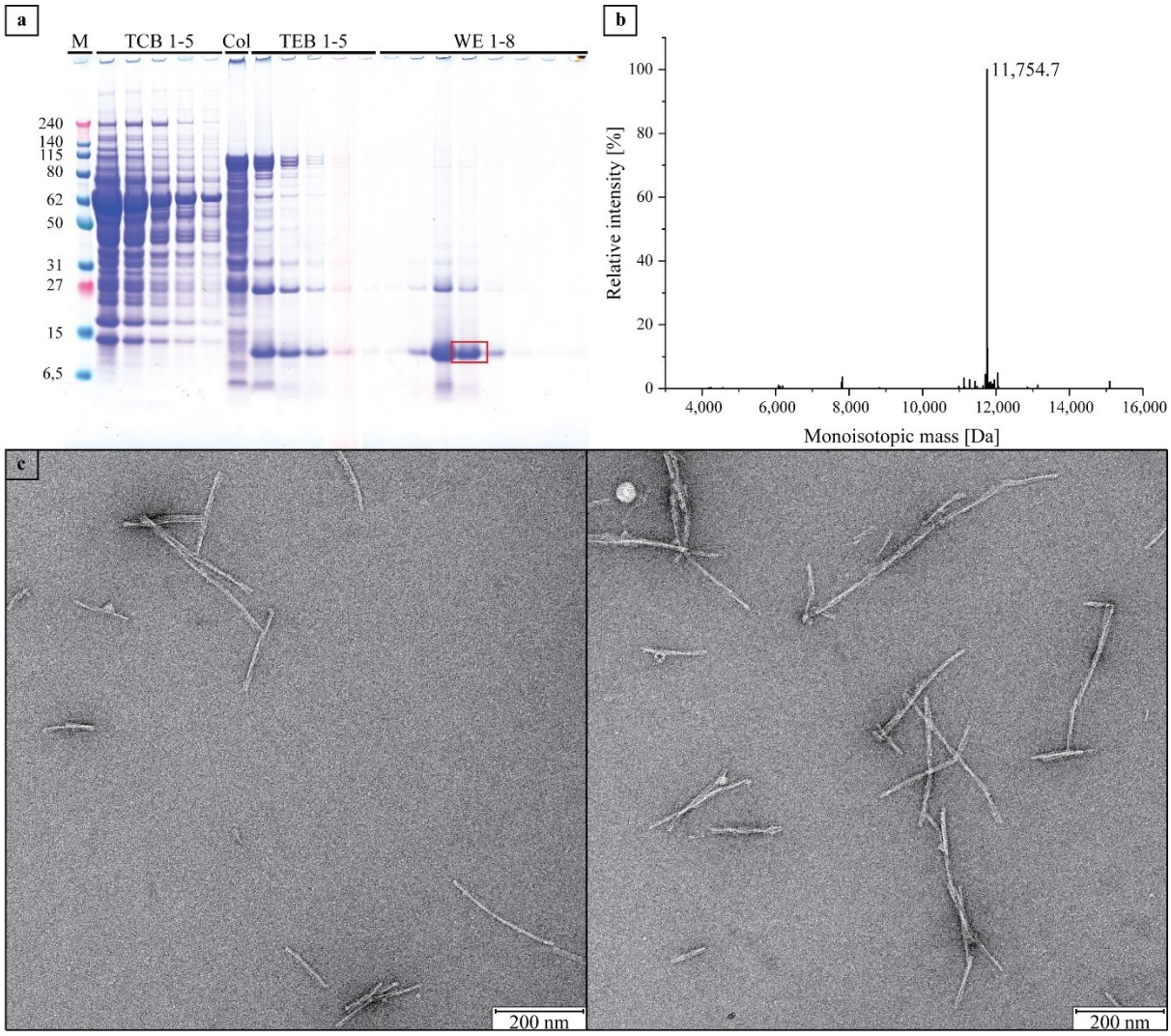


Figure 75: FOR012. a) LDS-PAGE of the water extraction. M: marker in kDa. Col: supernatant after collagenase incubation. WE: water extract. b) Deconvoluted mass spectrum of the extracted AL proteins. Y-axis: relative intensity in percent. X-axis: monoisotopic mass in Da. Number: mass assigned to a sequence. c) Negative stain TEM images (20 k magnification) of WE4 (red rectangle).

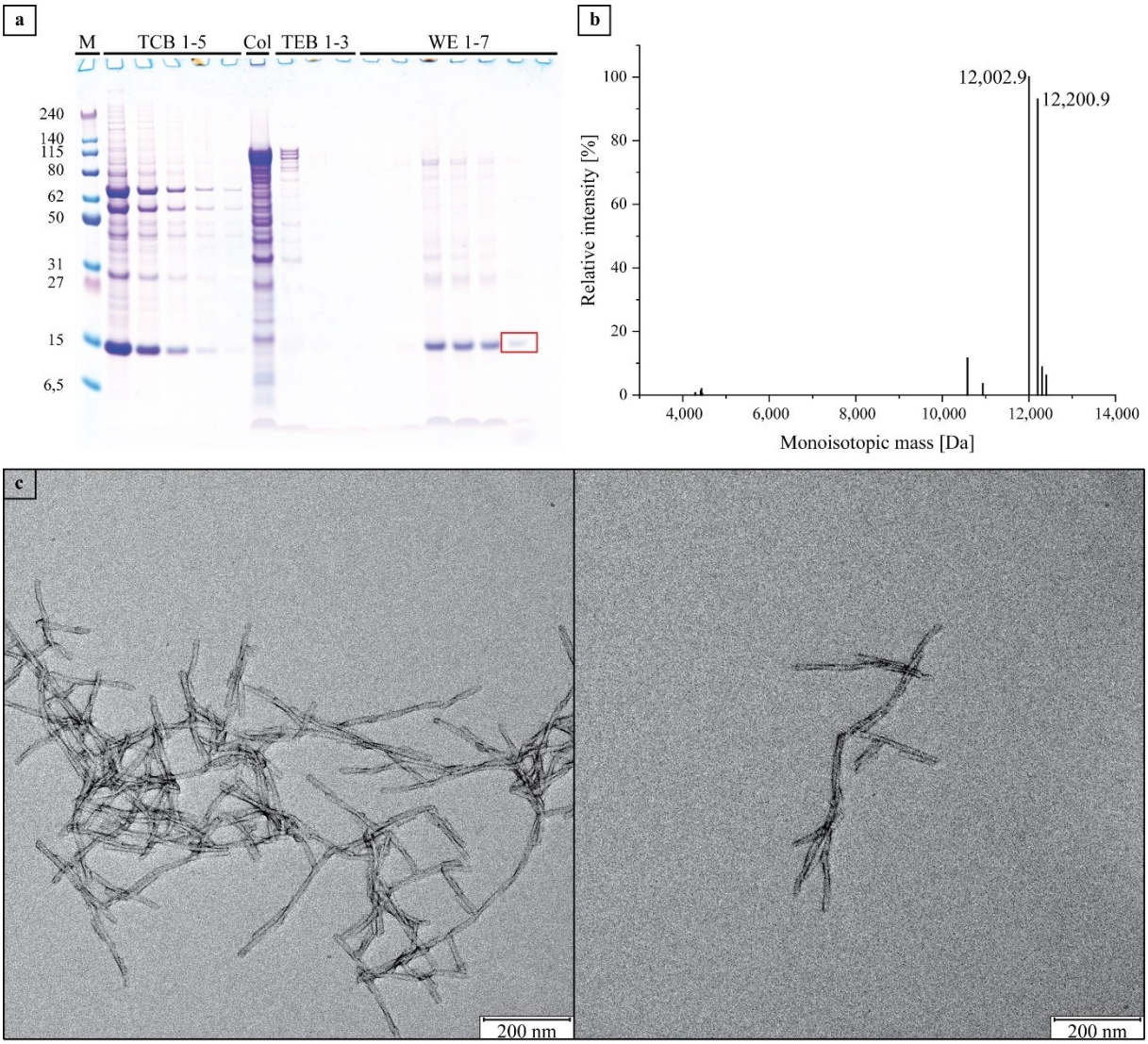


Figure 76: FOR101. a) LDS-PAGE of the water extraction. M: marker in kDa. Col: supernatant after collagenase incubation. WE: water extract. b) Deconvoluted mass spectrum of the extracted AL proteins. Y-axis: relative intensity in percent. X-axis: monoisotopic mass in Da. Number: mass assigned to a sequence. c) Negative stain TEM images (20 k magnification) of WE6 (red rectangle).

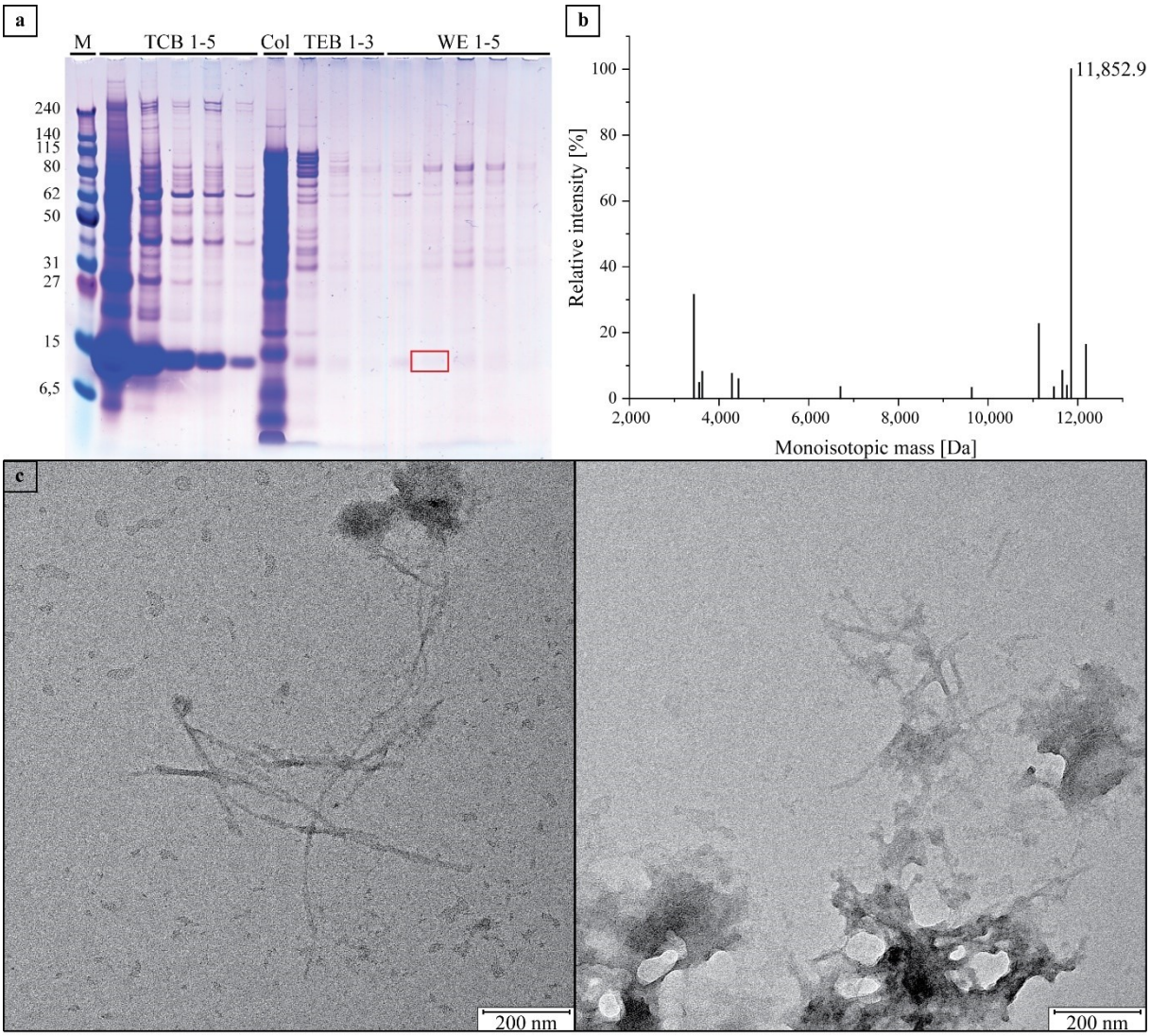


Figure 77: FOR102. a) LDS-PAGE of the water extraction. M: marker in kDa. Col: supernatant after collagenase incubation. WE: water extract. b) Deconvoluted mass spectrum of the extracted AL proteins. Y-axis: relative intensity in percent. X-axis: monoisotopic mass in Da. Number: mass assigned to a sequence. c) Negative stain TEM images (20 k magnification) of WE2 (red rectangle).

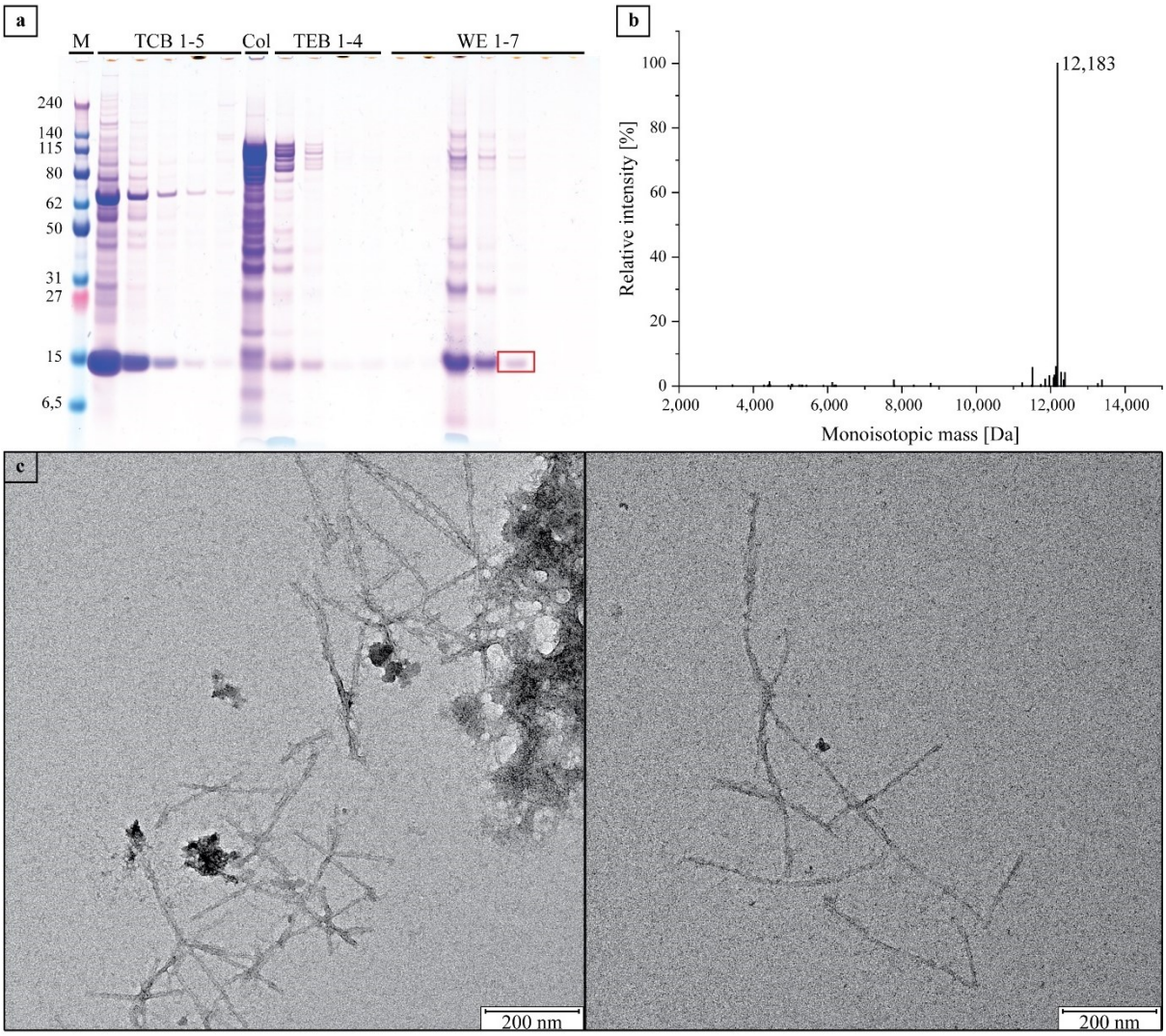


Figure 78: FOR103. a) LDS-PAGE of the water extraction. M: marker in kDa. Col: supernatant after collagenase incubation. WE: water extract. b) Deconvoluted mass spectrum of the extracted AL proteins. Y-axis: relative intensity in percent. X-axis: monoisotopic mass in Da. Number: mass assigned to a sequence. c) Negative stain TEM images (20 k magnification) of WE5 (red rectangle).

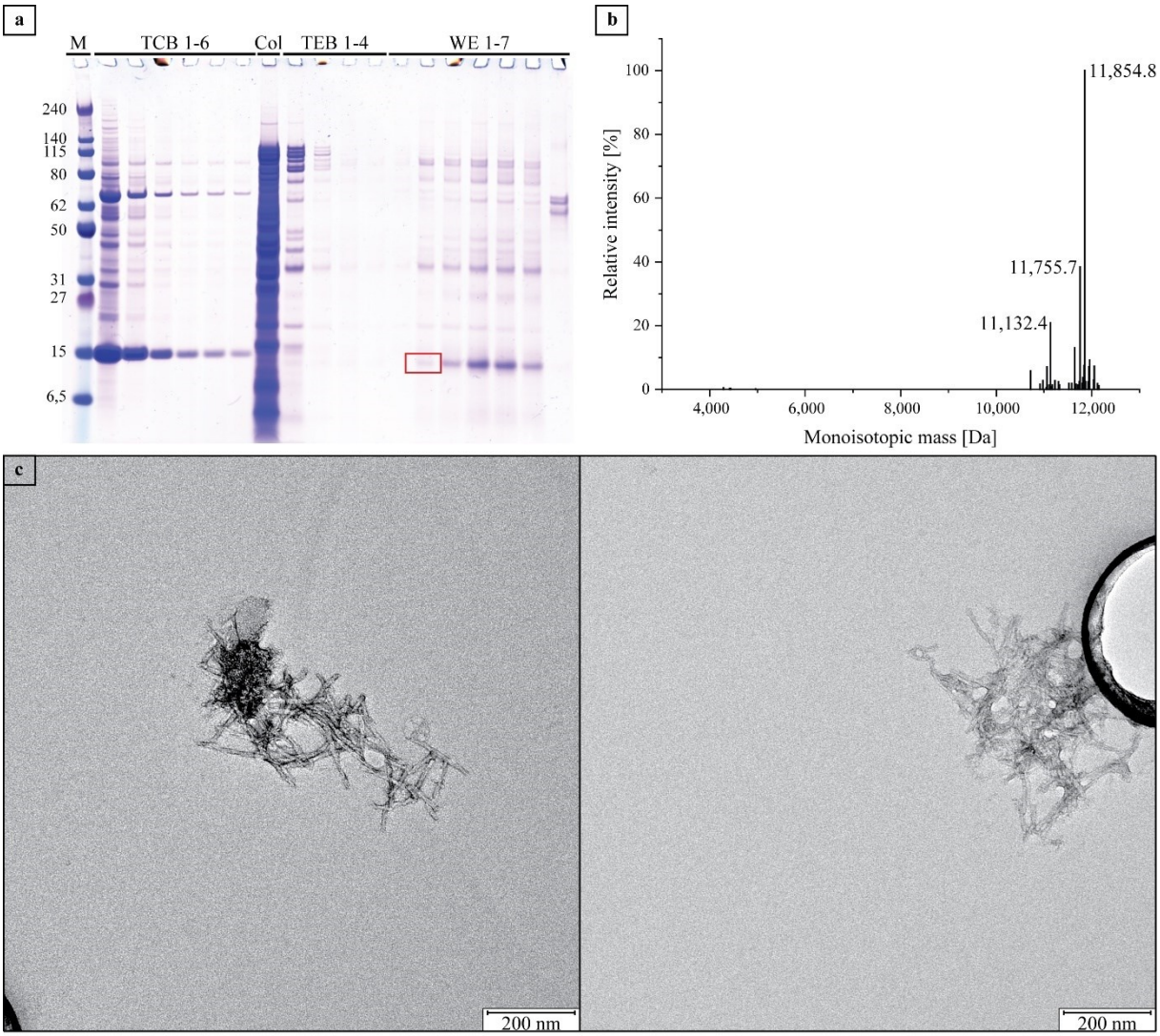


Figure 79: FOR107. a) LDS-PAGE of the water extraction. M: marker in kDa. Col: supernatant after collagenase incubation. WE: water extract. b) Deconvoluted mass spectrum of the extracted AL proteins. Y-axis: relative intensity in percent. X-axis: monoisotopic mass in Da. Number: mass assigned to a sequence. c) Negative stain TEM images (20 k magnification) of WE2 (red rectangle).

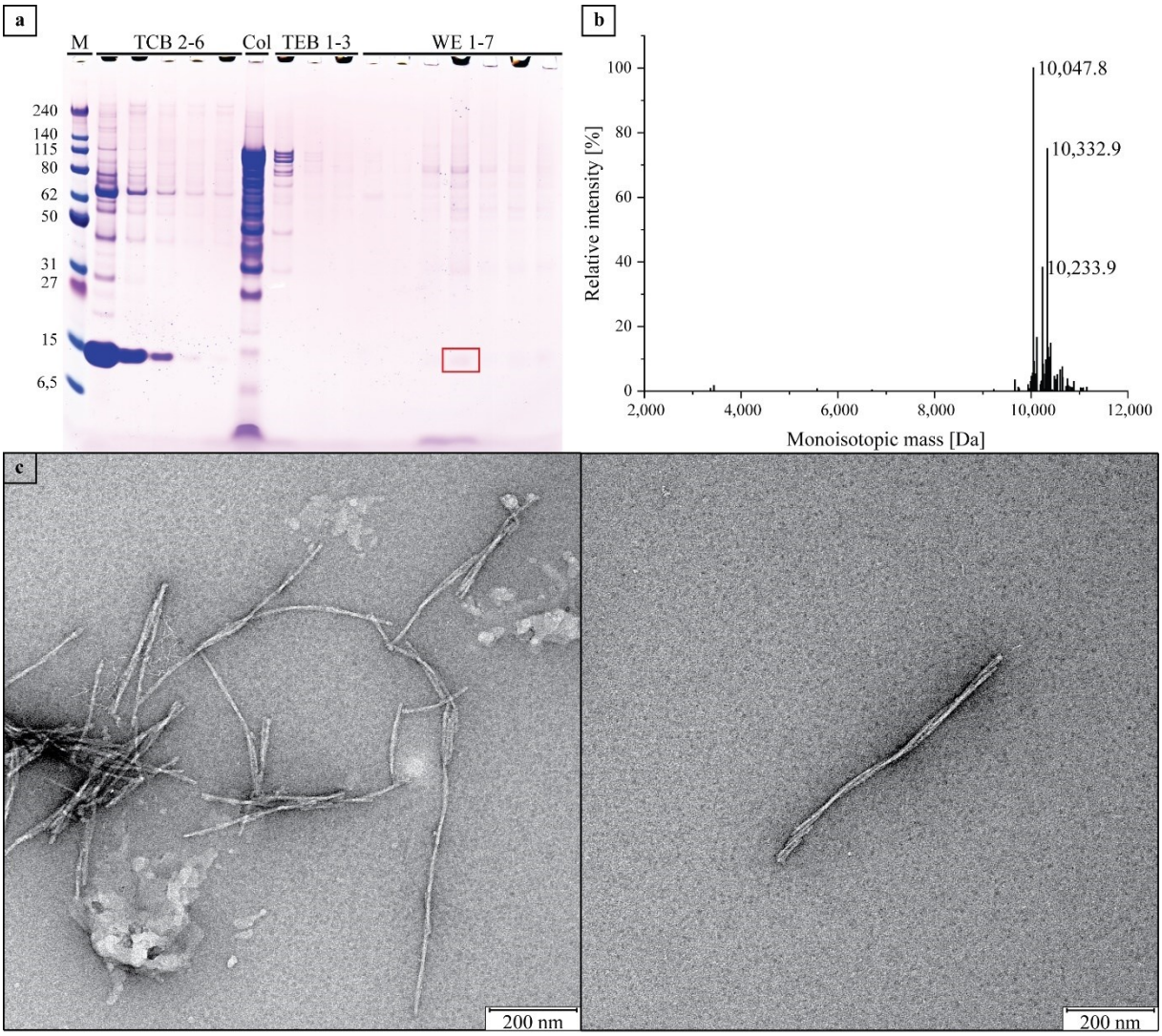


Figure 80: FOR124. a) LDS-PAGE of the water extraction. M: marker in kDa. Col: supernatant after collagenase incubation. WE: water extract. b) Deconvoluted mass spectrum of the extracted AL proteins. Y-axis: relative intensity in percent. X-axis: monoisotopic mass in Da. Number: mass assigned to a sequence. c) Negative stain TEM images (20 k magnification) of WE4 (red rectangle).

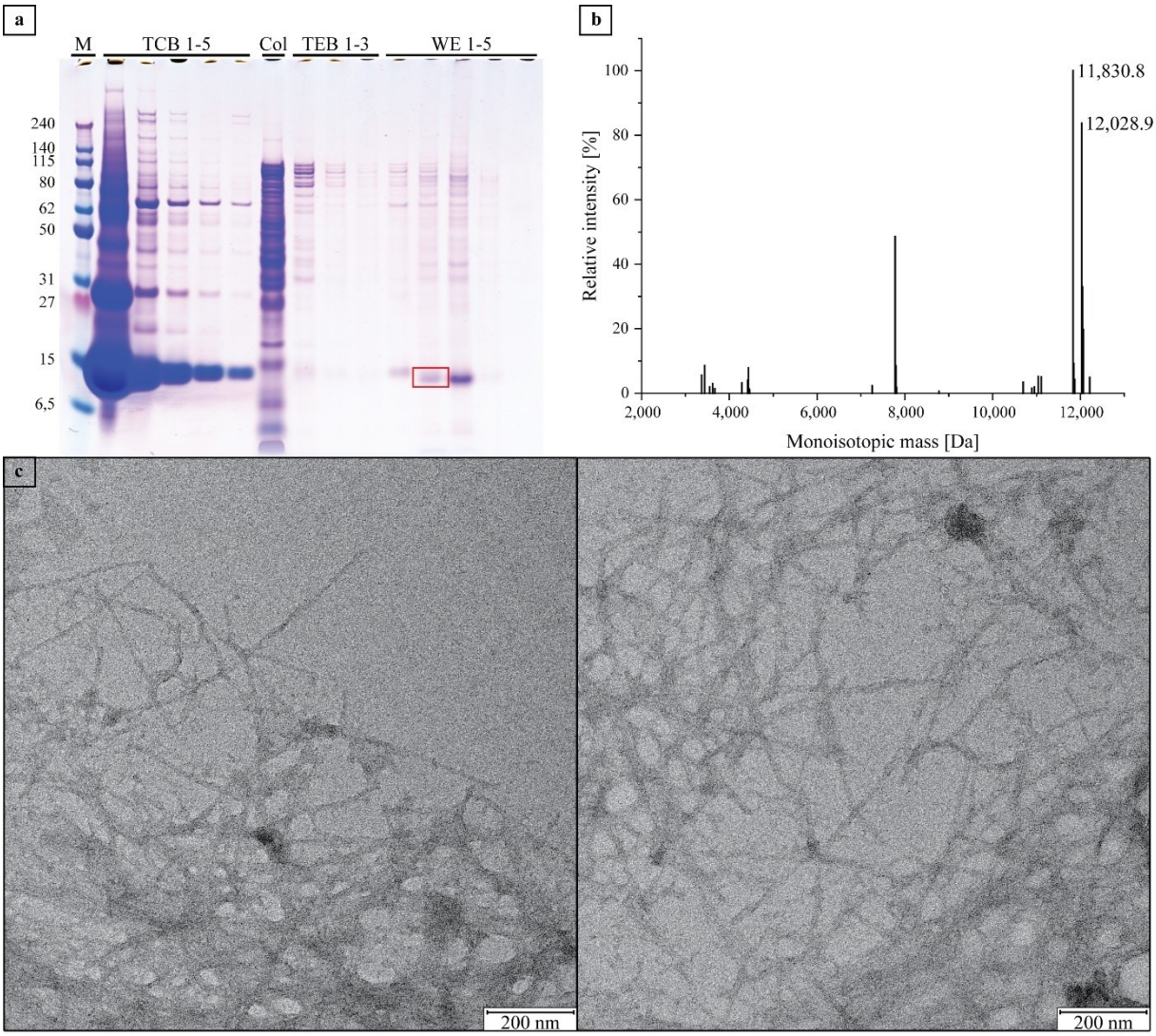


Figure 81: FOR128. a) LDS-PAGE of the water extraction. M: marker in kDa. Col: supernatant after collagenase incubation. WE: water extract. b) Deconvoluted mass spectrum of the extracted AL proteins. Y-axis: relative intensity in percent. X-axis: monoisotopic mass in Da. Number: mass assigned to a sequence. c) Negative stain TEM images (20 k magnification) of WE2 (red rectangle).

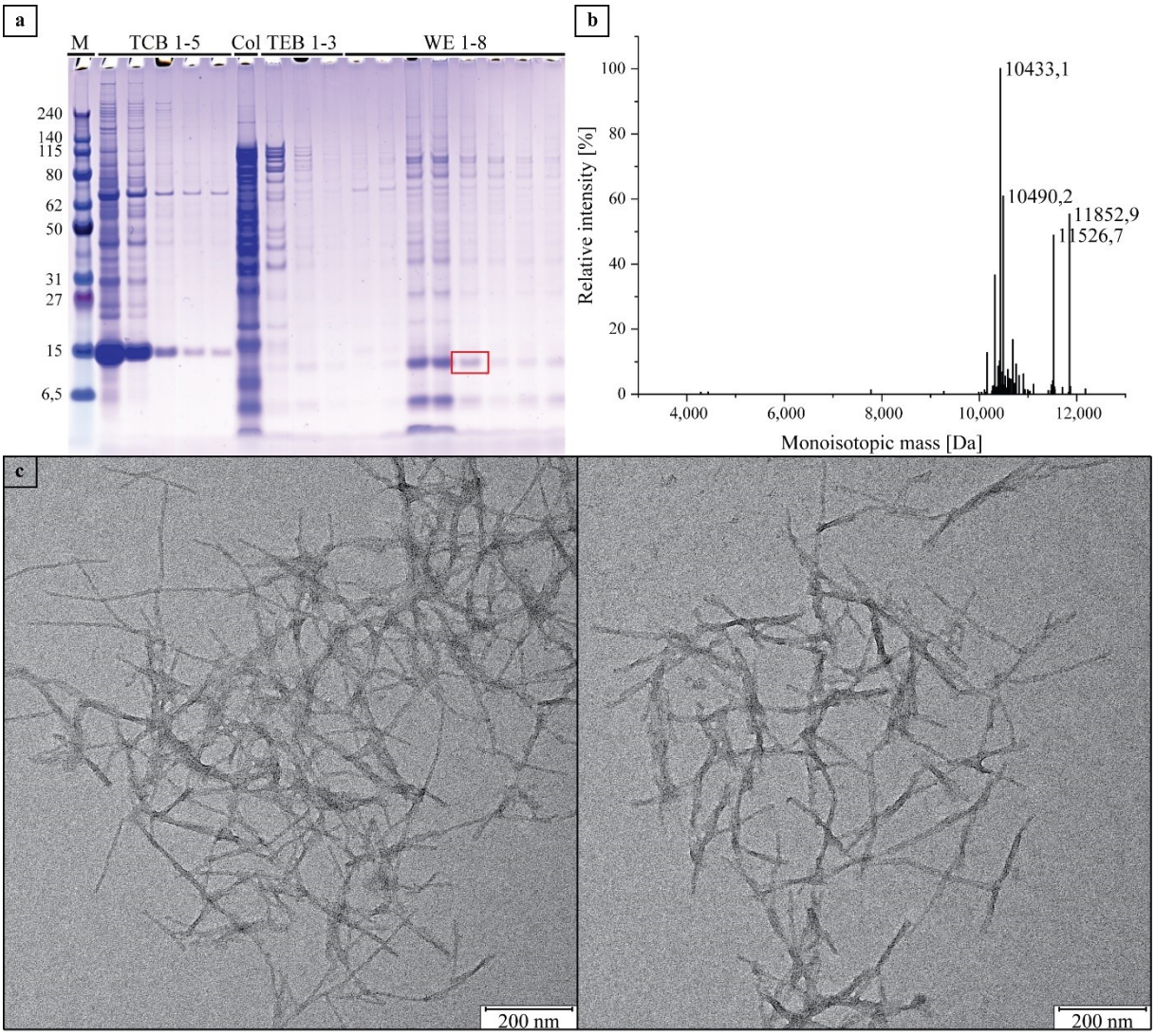


Figure 82: FOR137. a) LDS-PAGE of the water extraction. M: marker in kDa. Col: supernatant after collagenase incubation. WE: water extract. b) Deconvoluted mass spectrum of the extracted AL proteins. Y-axis: relative intensity in percent. X-axis: monoisotopic mass in Da. Number: mass assigned to a sequence. c) Negative stain TEM images (20 k magnification) of WE5 (red rectangle).

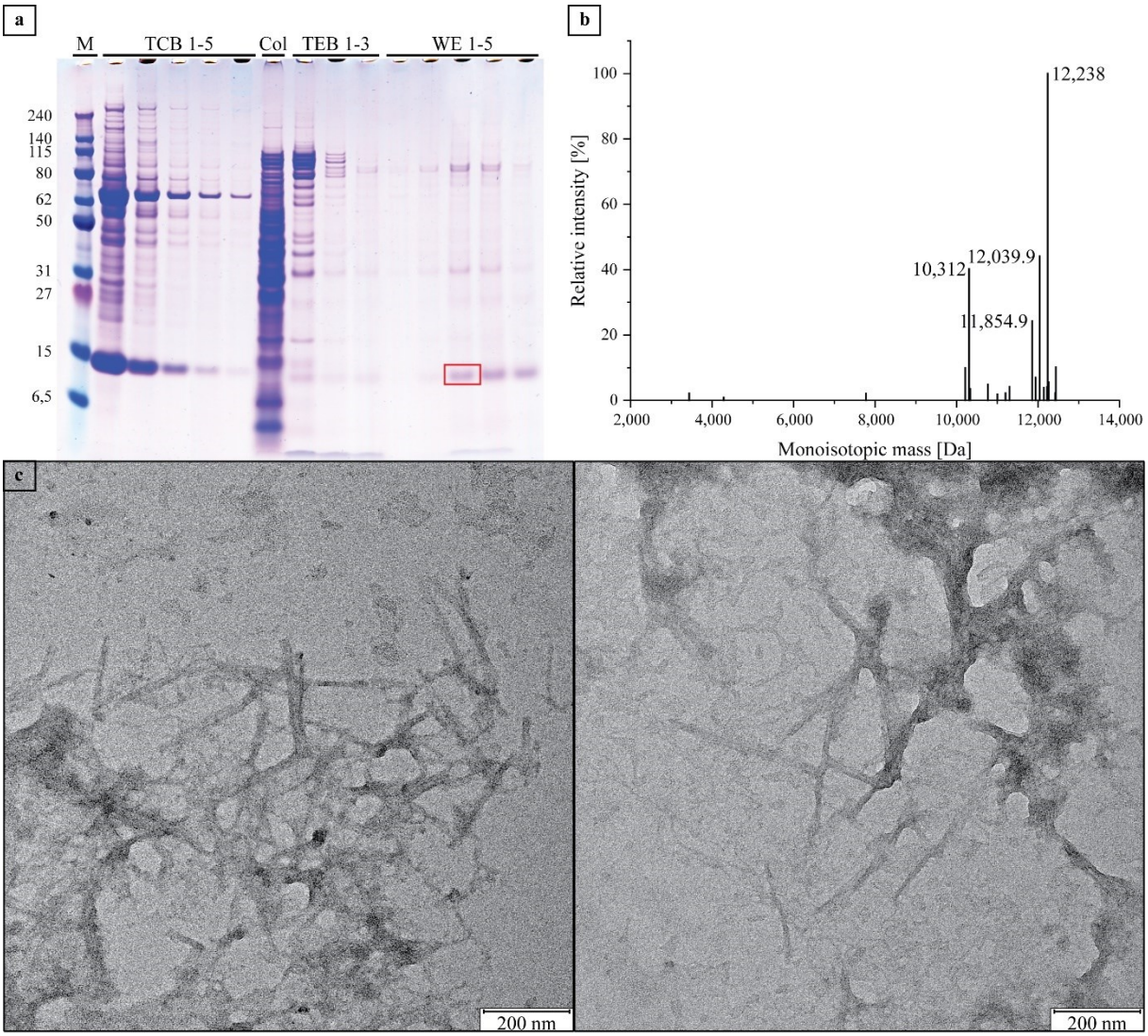


Figure 83: FOR138. a) LDS-PAGE of the water extraction. M: marker in kDa. Col: supernatant after collagenase incubation. WE: water extract. b) Deconvoluted mass spectrum of the extracted AL proteins. Y-axis: relative intensity in percent. X-axis: monoisotopic mass in Da. Number: mass assigned to a sequence. c) Negative stain TEM images (20 k magnification) of WE3 (red rectangle).

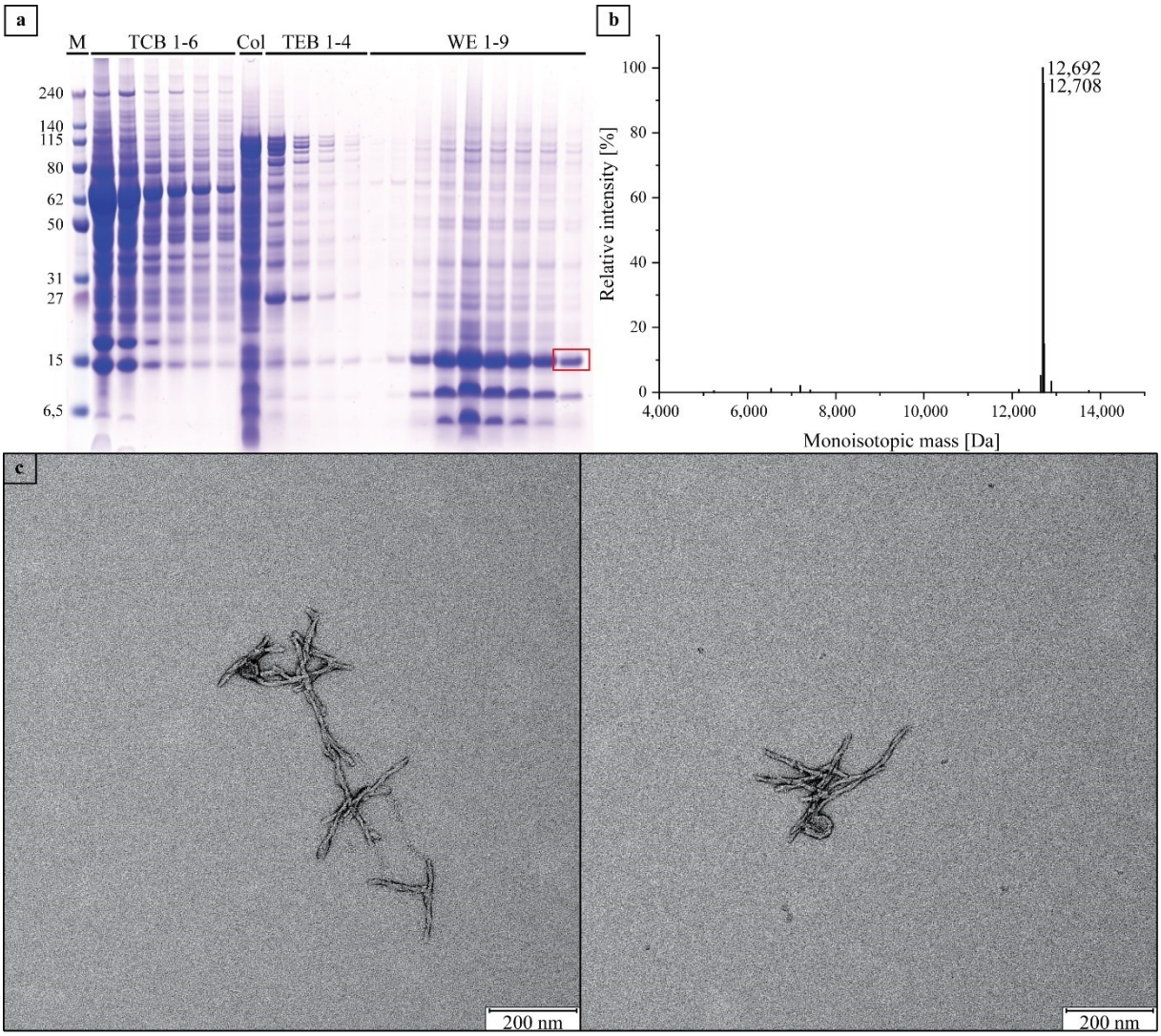


Figure 84: FOR140. a) LDS-PAGE of the water extraction. M: marker in kDa. Col: supernatant after collagenase incubation. WE: water extract. b) Deconvoluted mass spectrum of the extracted AL proteins. Y-axis: relative intensity in percent. X-axis: monoisotopic mass in Da. Number: mass assigned to a sequence. c) Negative stain TEM images (20 k magnification) of WE9 (red rectangle).

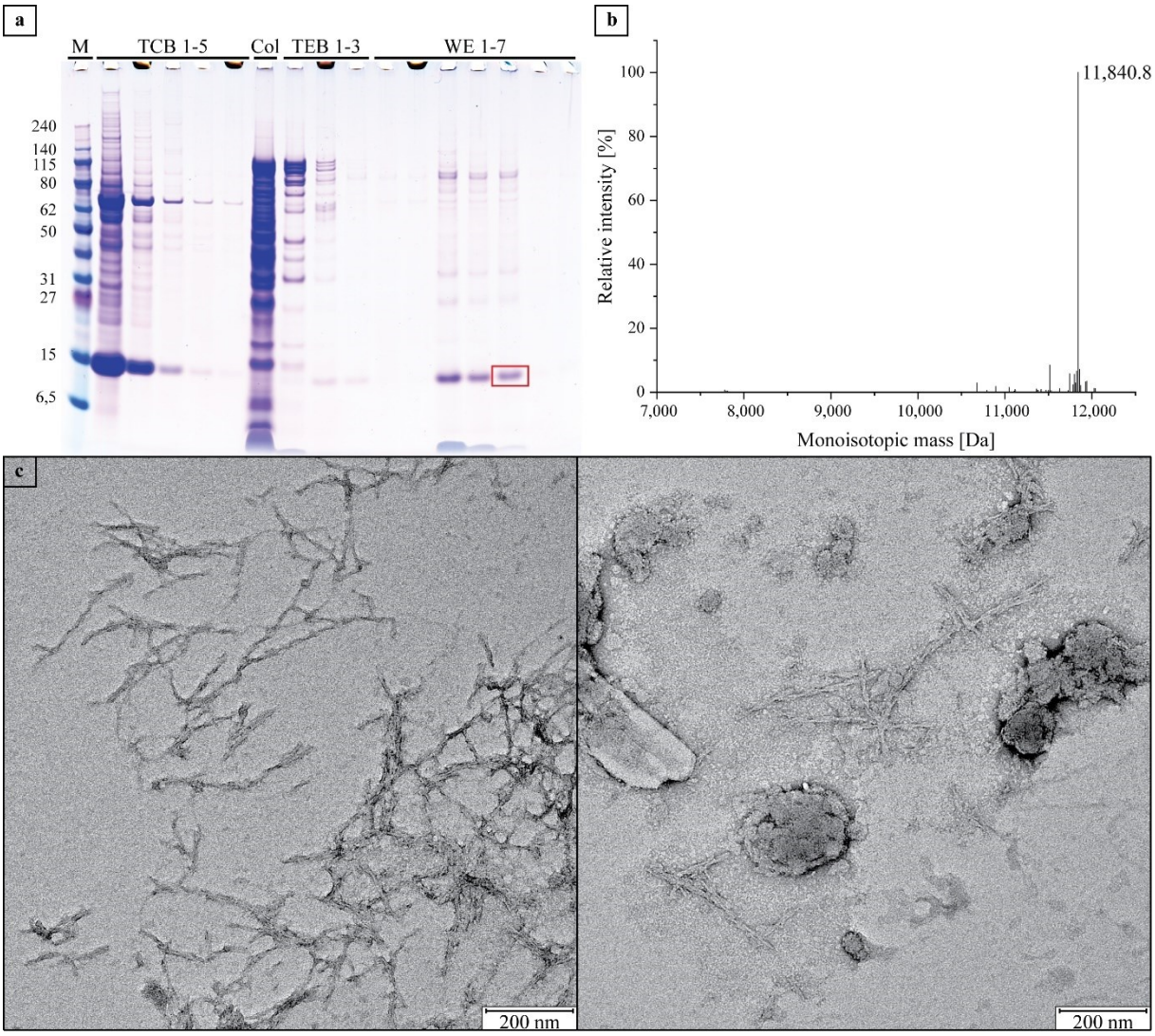


Figure 85: FOR142. a) LDS-PAGE of the water extraction. M: marker in kDa. Col: supernatant after collagenase incubation. WE: water extract. b) Deconvoluted mass spectrum of the extracted AL proteins. Y-axis: relative intensity in percent. X-axis: monoisotopic mass in Da. Number: mass assigned to a sequence. c) Negative stain TEM images (20 k magnification) of WE5 (red rectangle).

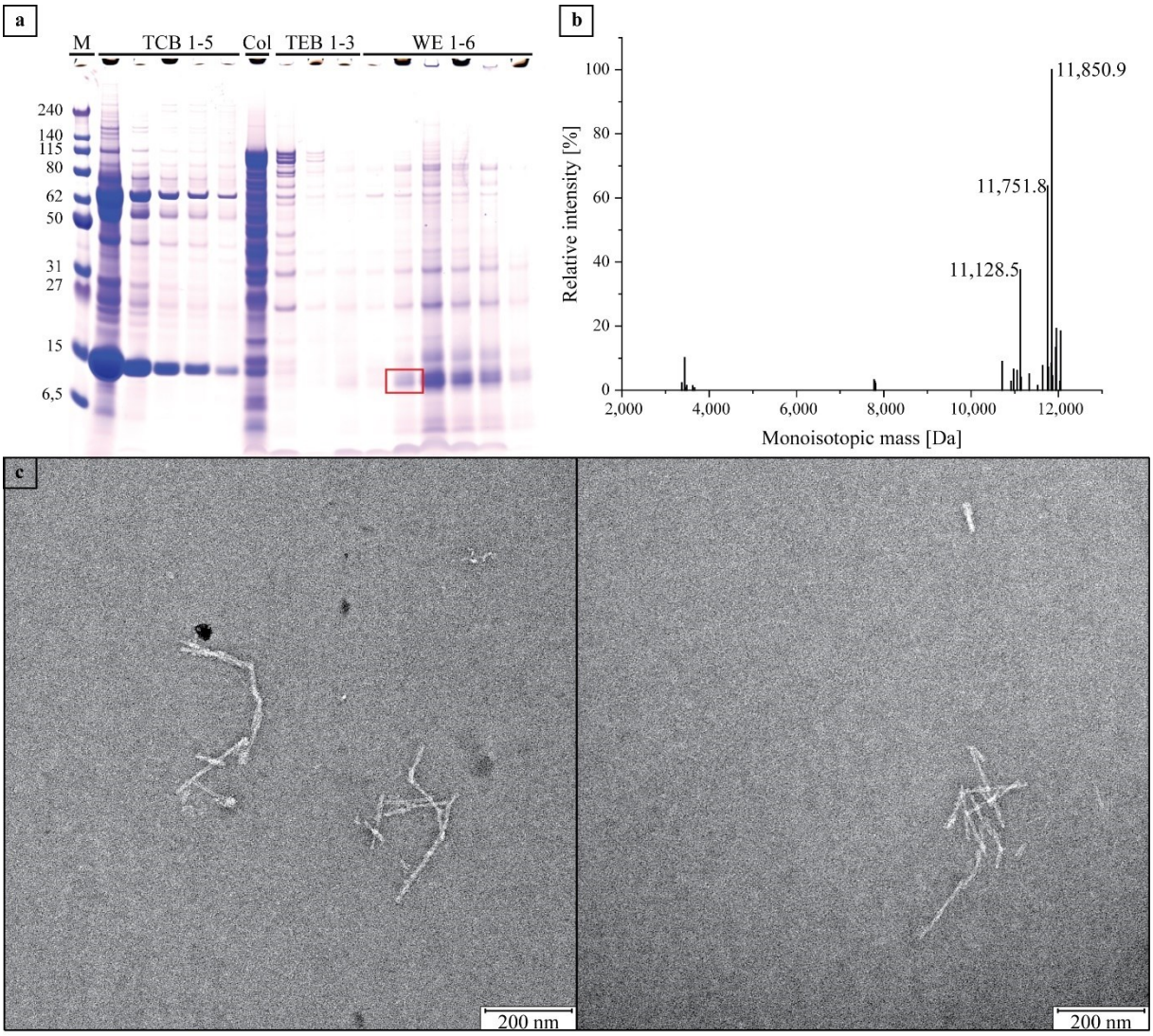


Figure 86: FOR147. a) LDS-PAGE of the water extraction. M: marker in kDa. Col: supernatant after collagenase incubation. WE: water extract. b) Deconvoluted mass spectrum of the extracted AL proteins. Y-axis: relative intensity in percent. X-axis: monoisotopic mass in Da. Number: mass assigned to a sequence. c) Negative stain TEM images (20 k magnification) of WE2 (red rectangle).

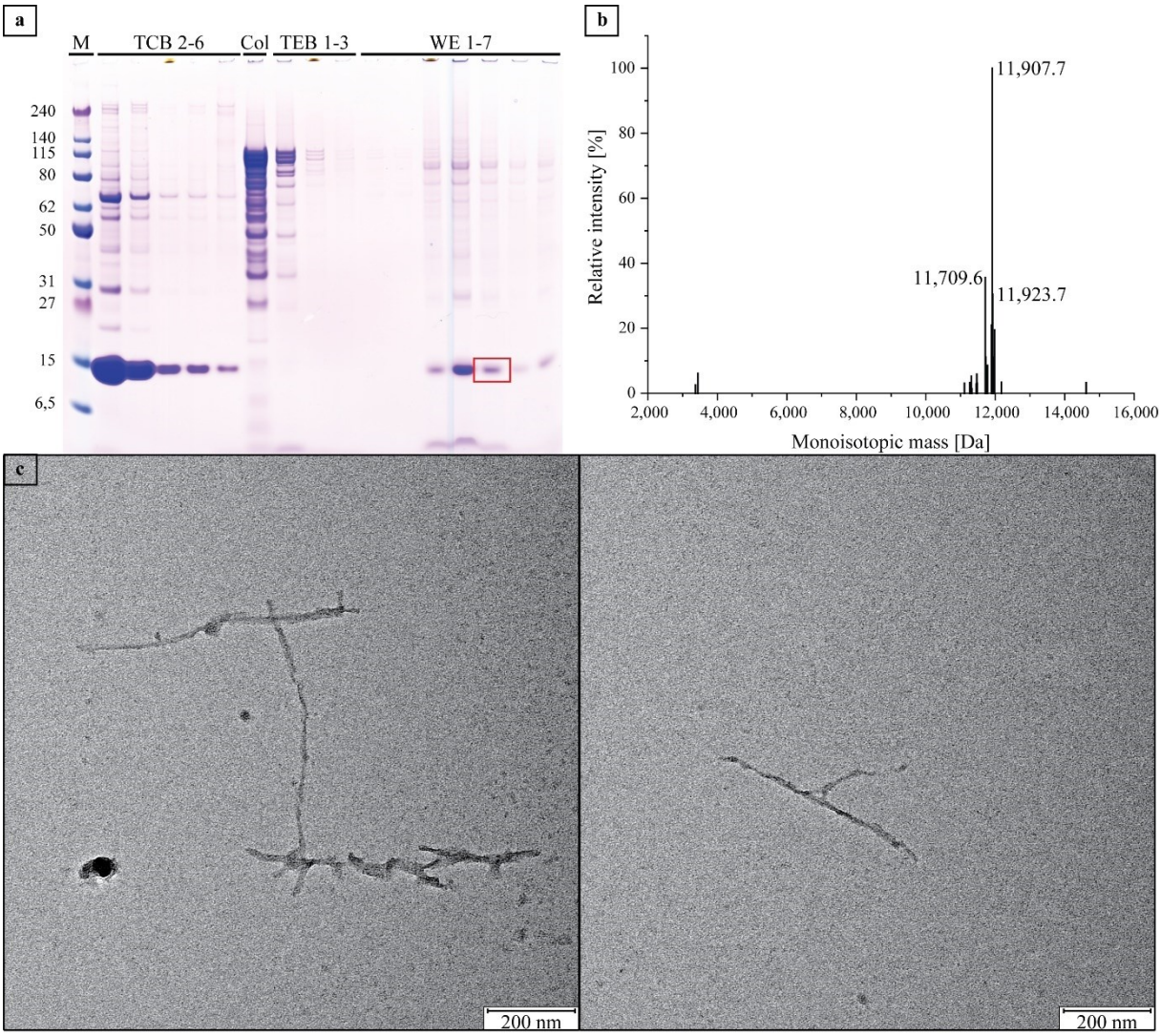


Figure 87: FOR157. a) LDS-PAGE of the water extraction. M: marker in kDa. Col: supernatant after collagenase incubation. WE: water extract. b) Deconvoluted mass spectrum of the extracted AL proteins. Y-axis: relative intensity in percent. X-axis: monoisotopic mass in Da. Number: mass assigned to a sequence. c) Negative stain TEM images (20 k magnification) of WE5 (red rectangle).

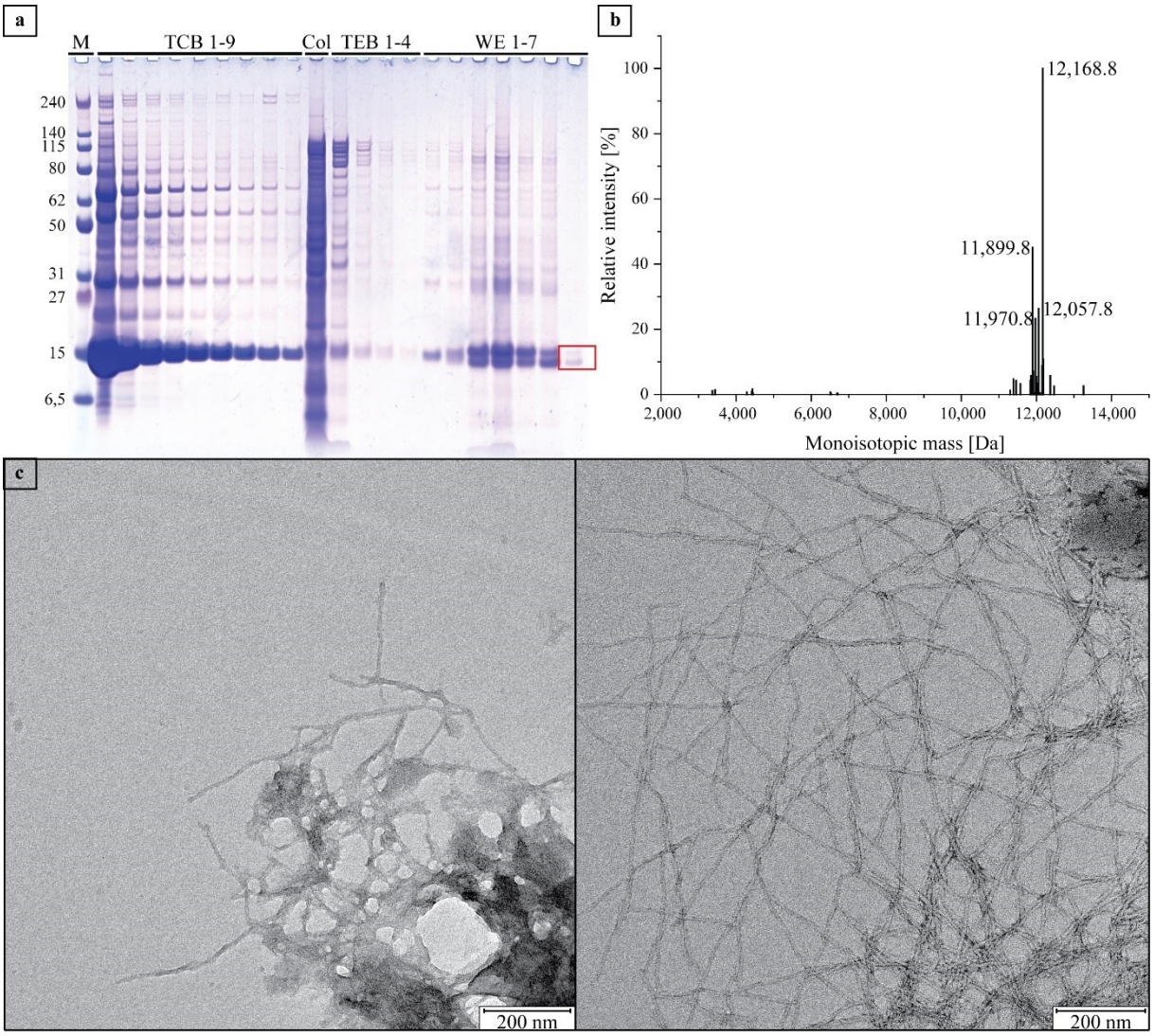


Figure 88: FOR159. a) LDS-PAGE of the water extraction. M: marker in kDa. Col: supernatant after collagenase incubation. WE: water extract. b) Deconvoluted mass spectrum of the extracted AL proteins. Y-axis: relative intensity in percent. X-axis: monoisotopic mass in Da. Number: mass assigned to a sequence. c) Negative stain TEM images (20 k magnification) of WE7 (red rectangle).

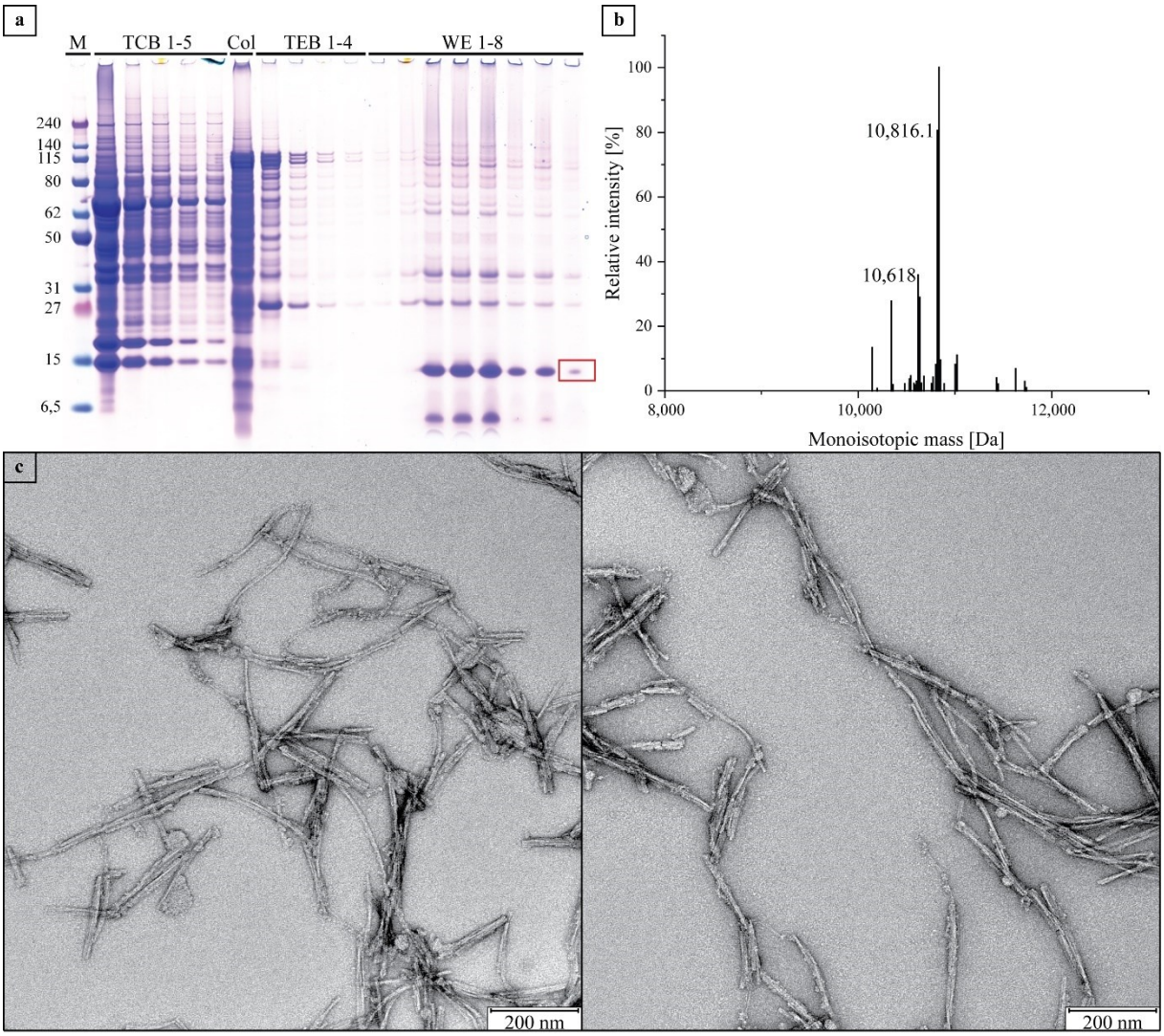


Figure 89: FOR161. a) LDS-PAGE of the water extraction. M: marker in kDa. Col: supernatant after collagenase incubation. WE: water extract. b) Deconvoluted mass spectrum of the extracted AL proteins. Y-axis: relative intensity in percent. X-axis: monoisotopic mass in Da. Number: mass assigned to a sequence. c) Negative stain TEM images (20 k magnification) of WE8 (red rectangle).

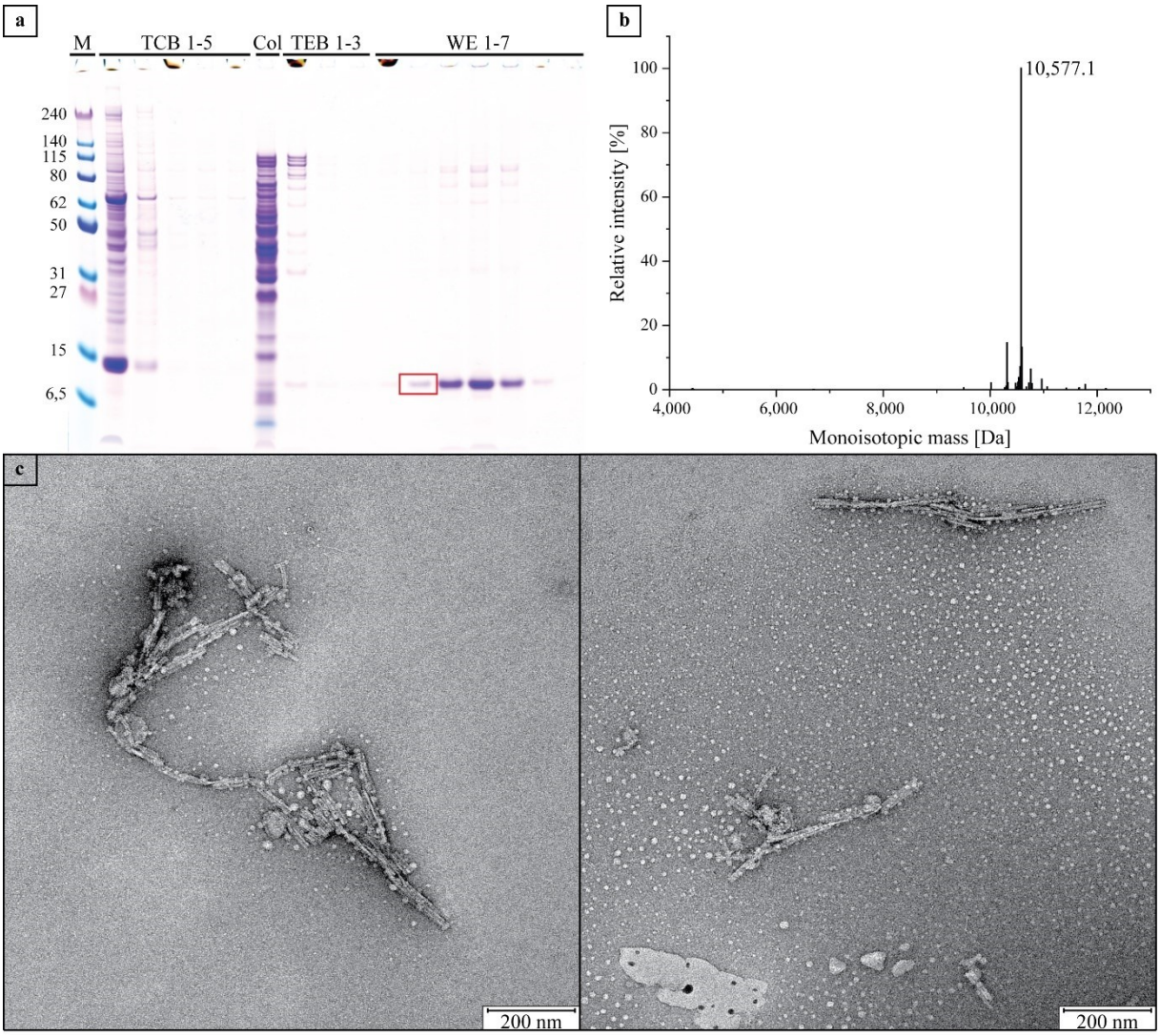


Figure 90: FOR1001. a) LDS-PAGE of the water extraction. M: marker in kDa. Col: supernatant after collagenase incubation. WE: water extract. b) Deconvoluted mass spectrum of the extracted AL proteins. Y-axis: relative intensity in percent. X-axis: monoisotopic mass in Da. Number: mass assigned to a sequence. c) Negative stain TEM images (20 k magnification) of WE2 (red rectangle).

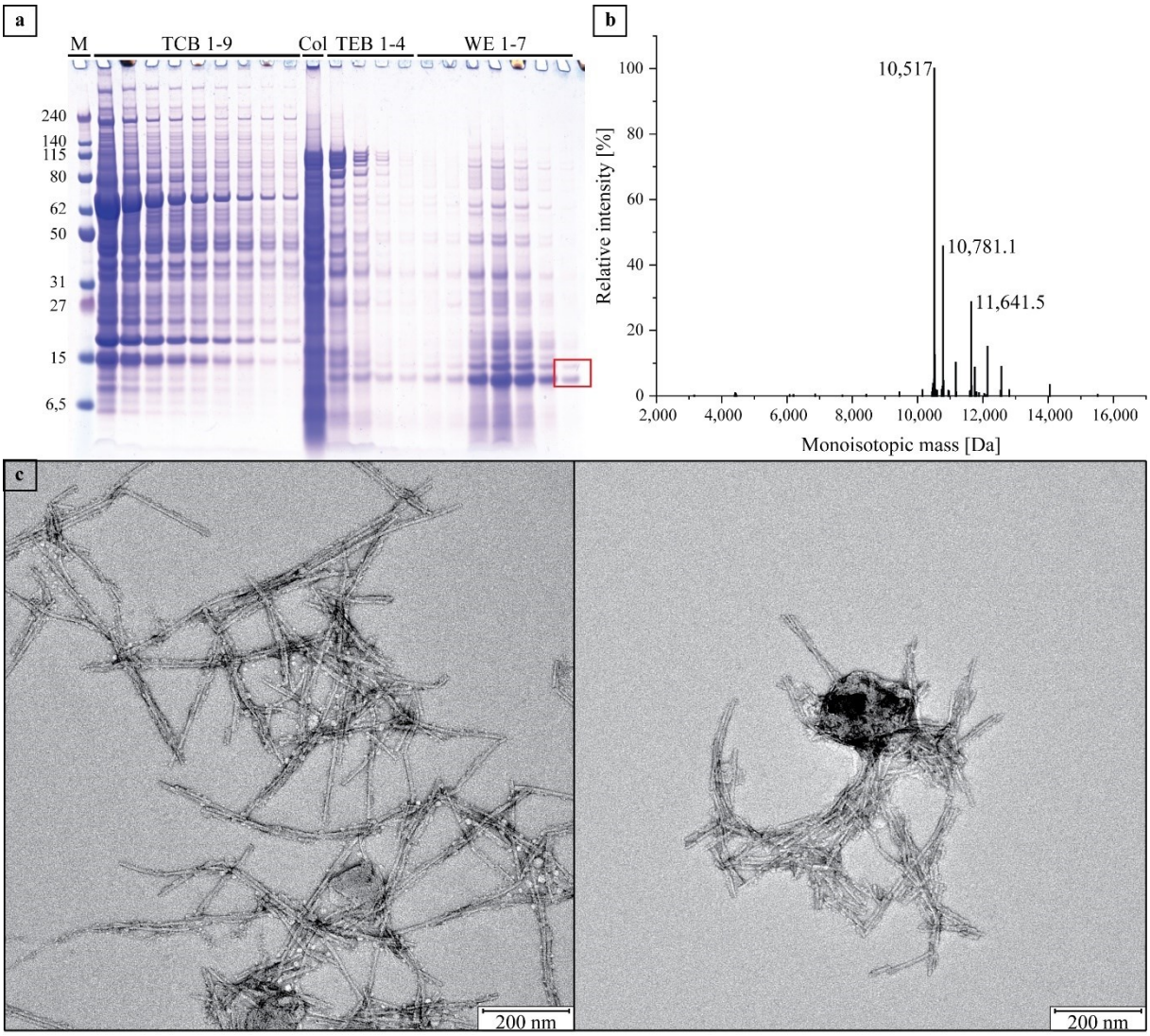


Figure 91: FOR1007. a) LDS-PAGE of the water extraction. M: marker in kDa. Col: supernatant after collagenase incubation. WE: water extract. b) Deconvoluted mass spectrum of the extracted AL proteins. Y-axis: relative intensity in percent. X-axis: monoisotopic mass in Da. Number: mass assigned to a sequence. c) Negative stain TEM images (20 k magnification) of WE7 (red rectangle).

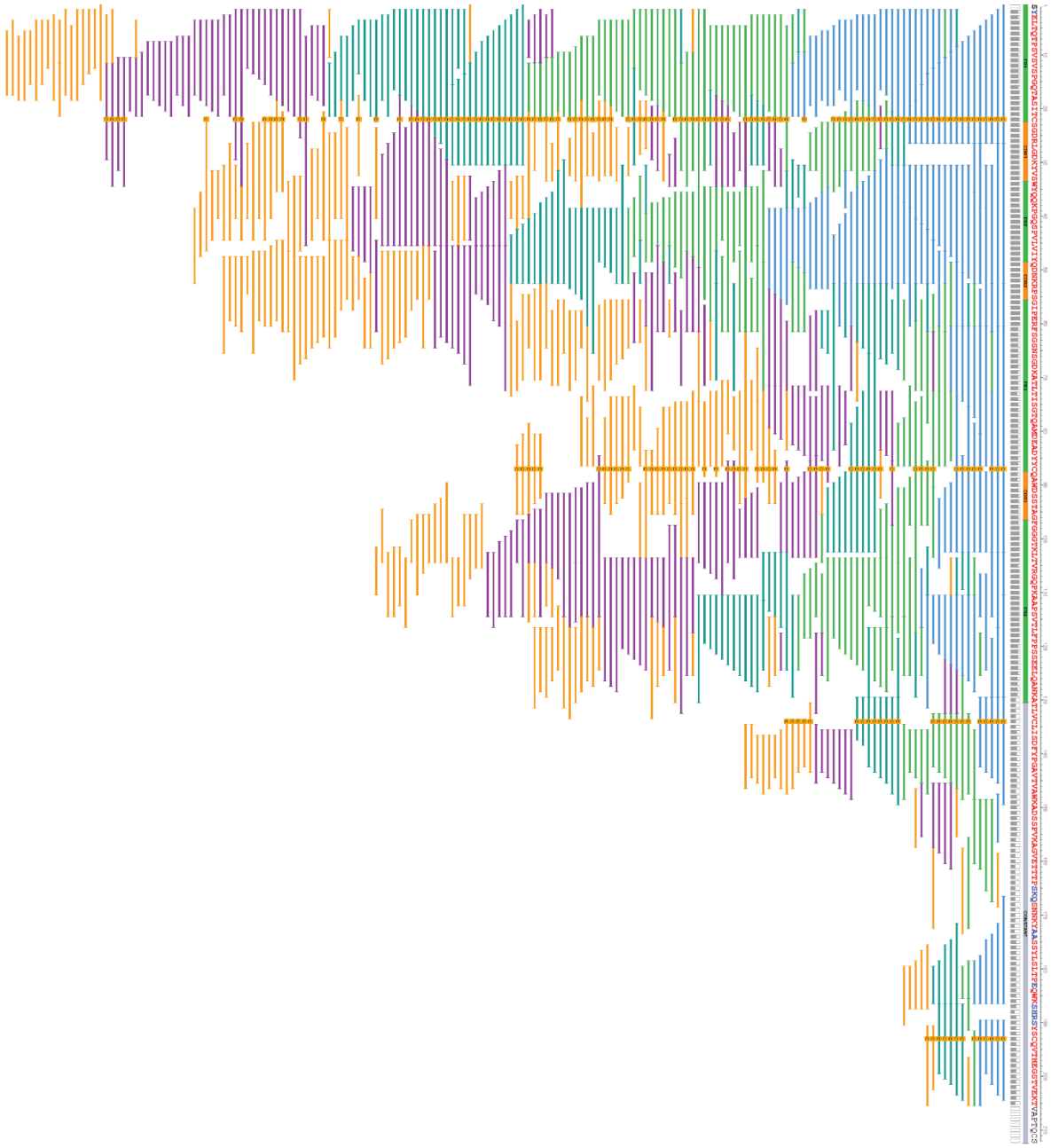


Figure 92: De novo sequencing results of FOR009 using PEAKS AB. Mass accuracy for fragmented ions: 0.5 Da. Yellow C: carbamidomethylation. Peptides: light blue (trypsin), light green (chymotrypsin), yellow (elastase), violet (pepsin), and cyan (LysC). Above the peptides (top to bottom): Amino acid position, sequence, CDRs, FRs, confidence score per amino acid. Confidence scores are also illustrated by the different colors of the amino acids: red (>95 %), dark blue (>85 %), bold black (<85 %), and black (not verified suggestions).

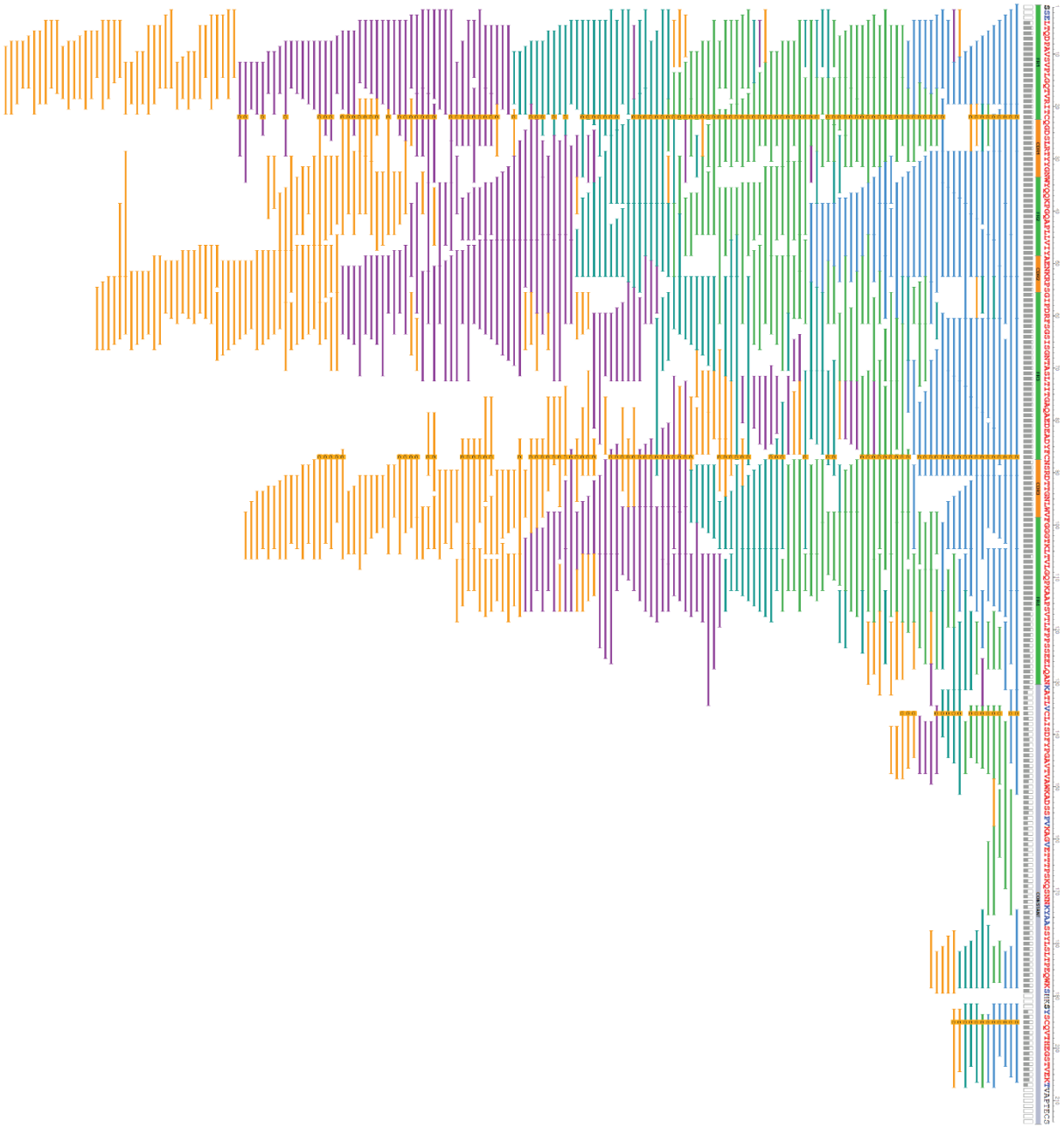


Figure 93: De novo sequencing results of FOR010 using PEAKS AB. Mass accuracy for fragmented ions: 0.5 Da. Yellow C: carbamidomethylation. Peptides: light blue (trypsin), light green (chymotrypsin), yellow (elastase), violet (pepsin), and cyan (LysC). Above the peptides (top to bottom): Amino acid position, sequence, CDRs, FRs, confidence score per amino acid. Confidence scores are also illustrated by the different colors of the amino acids: red (>95 %), dark blue (>85 %), bold black (<85 %), and black (not verified suggestions).

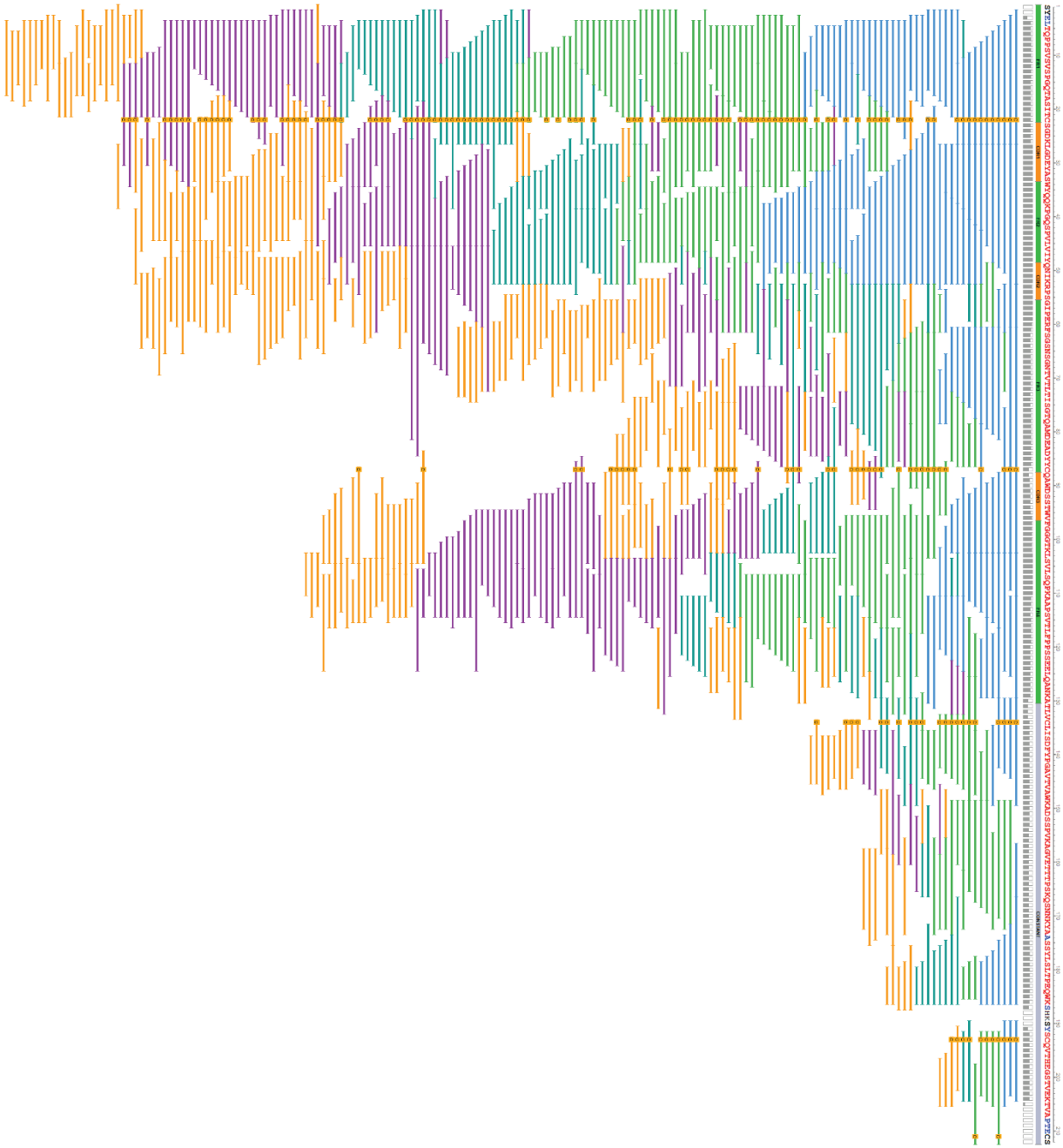


Figure 94: De novo sequencing results of FOR012 using PEAKS AB. Mass accuracy for fragmented ions: 0.5 Da. Yellow C: carbamidomethylation. Peptides: light blue (trypsin), light green (chymotrypsin), yellow (elastase), violet (pepsin), and cyan (LysC). Above the peptides (top to bottom): Amino acid position, sequence, CDRs, FRs, confidence score per amino acid. Confidence scores are also illustrated by the different colors of the amino acids: red (>95 %), dark blue (>85 %), bold black (<85 %), and black (not verified suggestions).

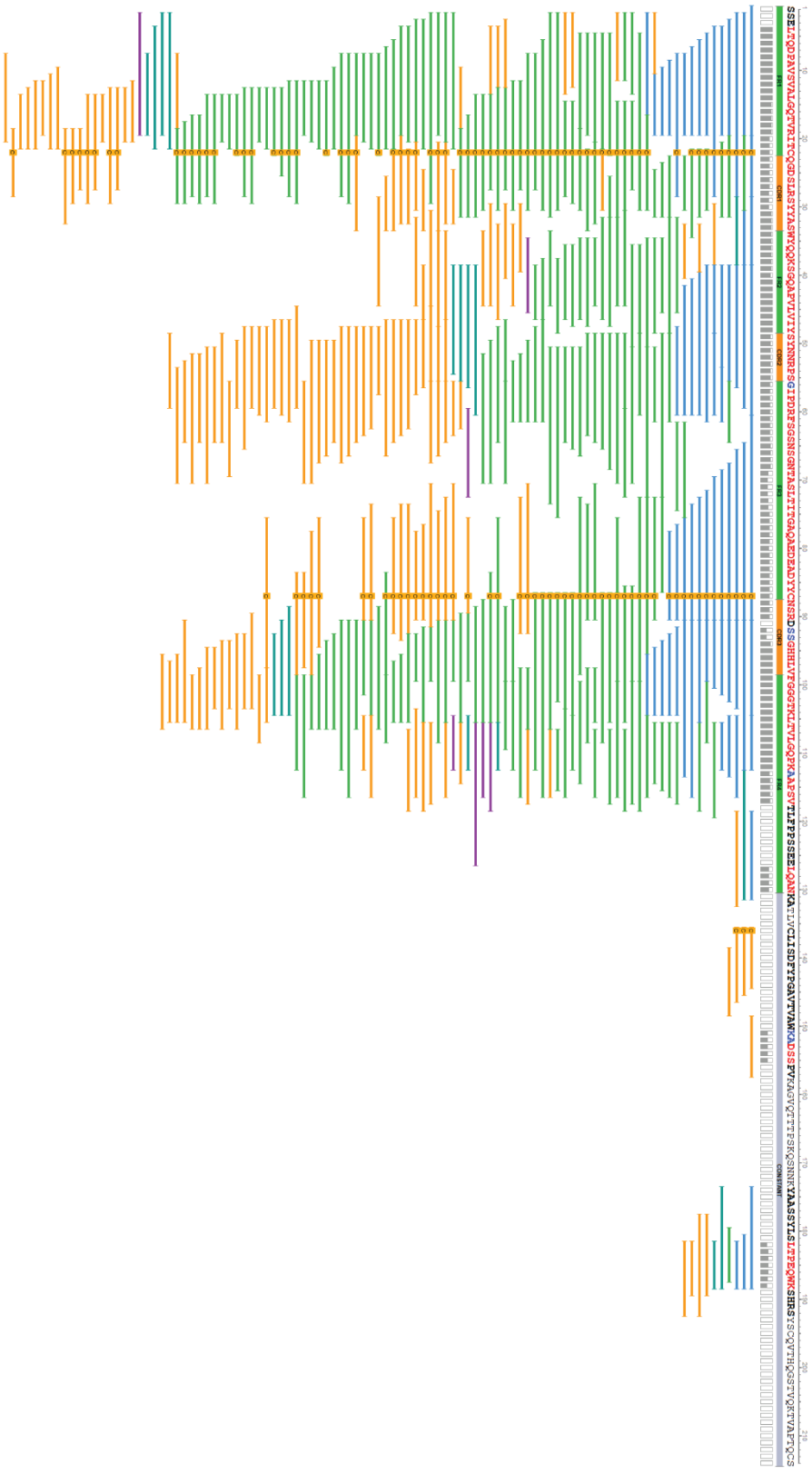


Figure 95: De novo sequencing results of FOR103 using PEAKS AB. Mass accuracy for fragmented ions: 0.05 Da. Yellow C: carbamidomethylation. Peptides: light blue (trypsin), light green (chymotrypsin), yellow (elastase), violet (pepsin), and cyan (LysC). Above the peptides (top to bottom): Amino acid position, sequence, CDRs, FRs, confidence score per amino acid. Confidence scores are also illustrated by the different colors of the amino acids: red (>95 %), dark blue (>85 %), bold black (<85 %), and black (not verified suggestions).

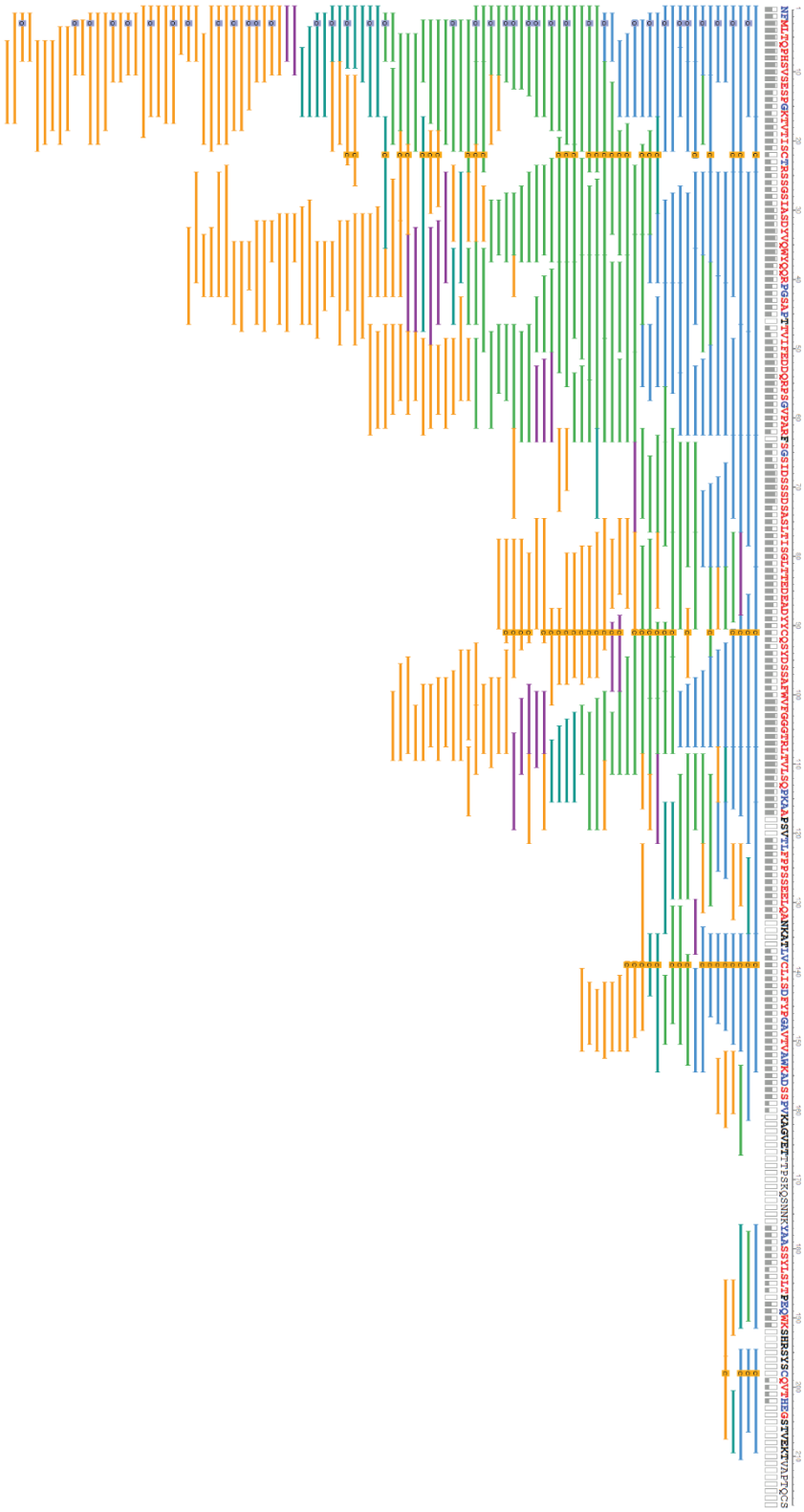


Figure 96: De novo sequencing results of FOR140 using PEAKS AB. Mass accuracy for fragmented ions: 0.05 Da. Yellow C: carbamidomethylation. Violet o: oxidation. Peptides: light blue (trypsin), light green (chymotrypsin), yellow (elastase), violet (pepsin), and cyan (LysC). Above the peptides (top to bottom): Amino acid position, sequence, CDRs, FRs, confidence score per amino acid. Confidence scores are also illustrated by the different colors of the amino acids: red (>95 %), dark blue (>85 %), bold black (<85 %), and black (not verified suggestions).

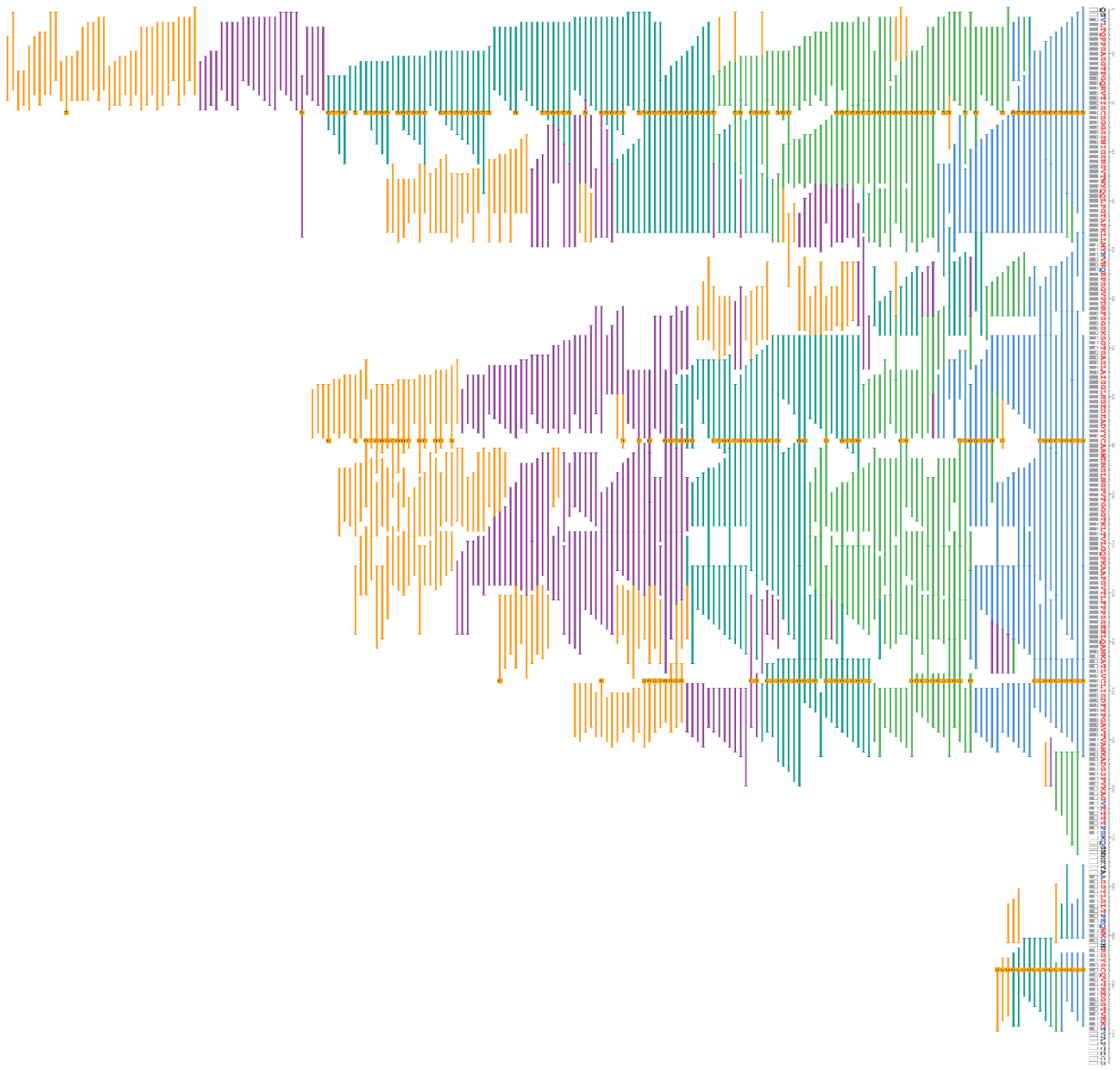


Figure 97: De novo sequencing results of FOR147 using PEAKS AB. Mass accuracy for fragmented ions: 0.5 Da. Yellow C: carbamidomethylation. Peptides: light blue (trypsin), light green (chymotrypsin), yellow (elastase), violet (pepsin), and cyan (LysC). Above the peptides (top to bottom): Amino acid position, sequence, CDRs, FRs, confidence score per amino acid. Confidence scores are also illustrated by the different colors of the amino acids: red (>95 %), dark blue (>85 %), bold black (<85 %), and black (not verified suggestions).

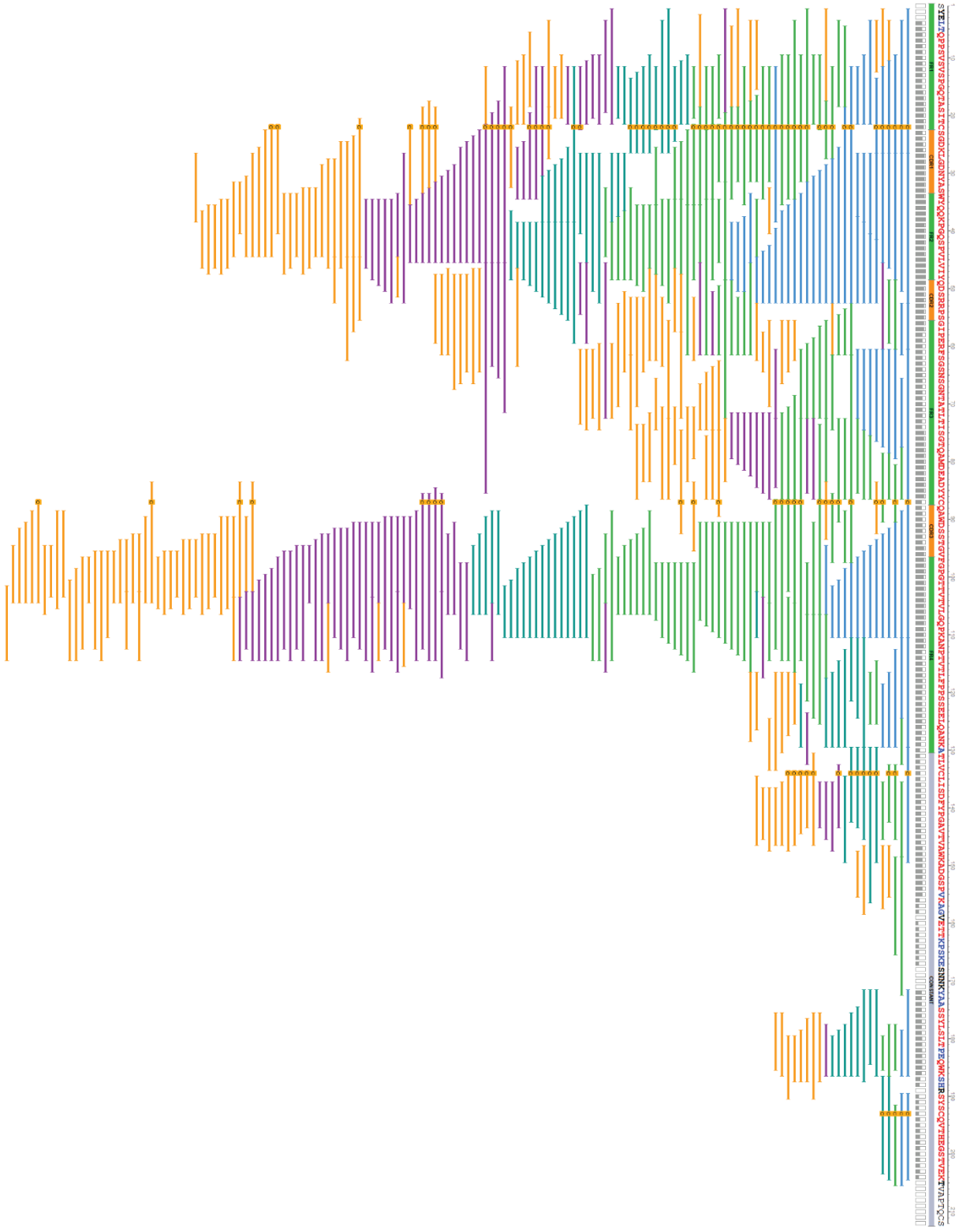


Figure 98: De novo sequencing results of FOR161 using PEAKS AB. Mass accuracy for fragmented ions: 0.5 Da. Yellow C: carbamidomethylation. Peptides: light blue (trypsin), light green (chymotrypsin), yellow (elastase), violet (pepsin), and cyan (LysC). Above the peptides (top to bottom): Amino acid position, sequence, CDRs, FRs, confidence score per amino acid. Confidence scores are also illustrated by the different colors of the amino acids: red (>95 %), dark blue (>85 %), bold black (<85 %), and black (not verified suggestions).

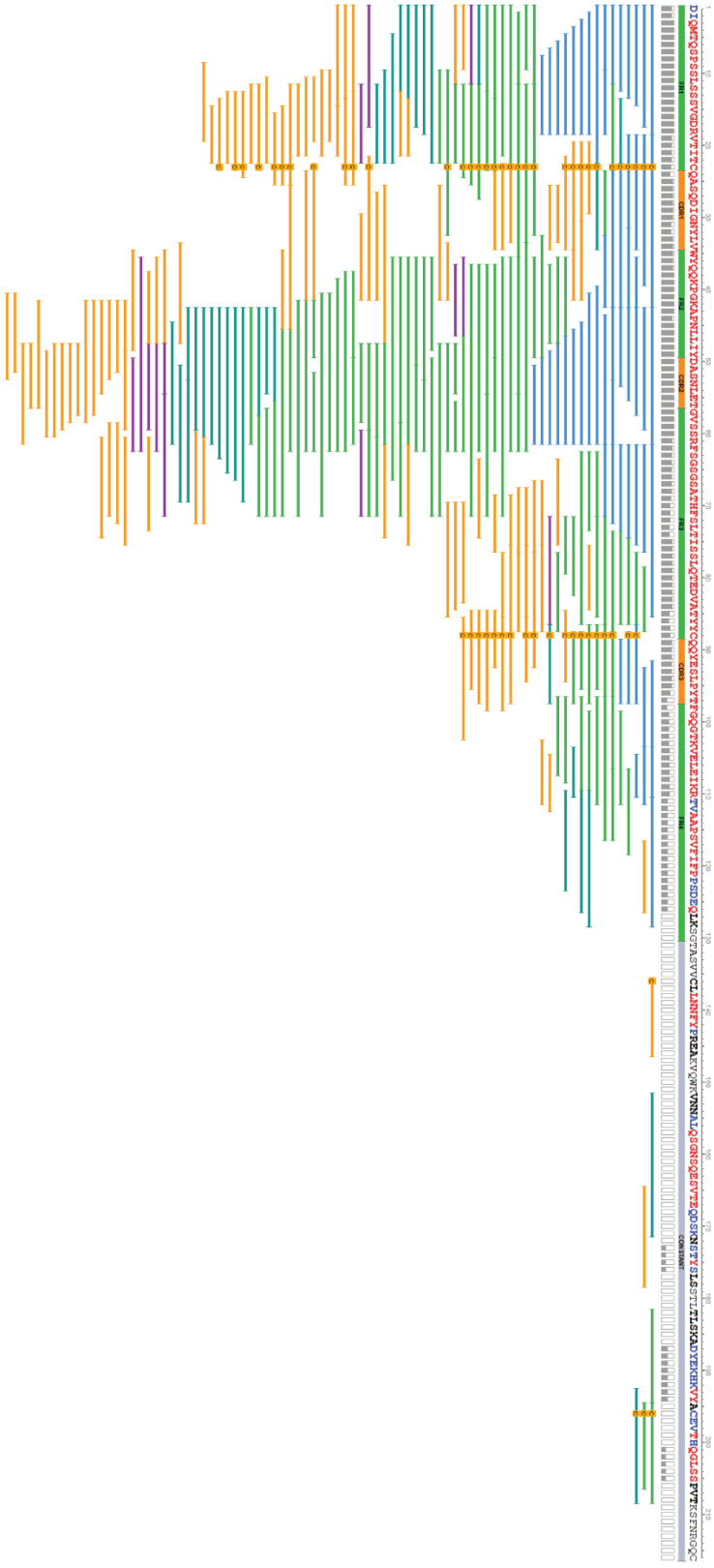


Figure 99: De novo sequencing results of FOR1001 using PEAKS AB. Mass accuracy for fragmented ions: 0.5 Da. Yellow C: carbamidomethylation. Peptides: light blue (trypsin), light green (chymotrypsin), yellow (elastase), violet (pepsin), and cyan (LysC). Above the peptides (top to bottom): Amino acid position, sequence, CDRs, FRs, confidence score per amino acid. Confidence scores are also illustrated by the different colors of the amino acids: red (>95 %), dark blue (>85 %), bold black (<85 %), and black (not verified suggestions).

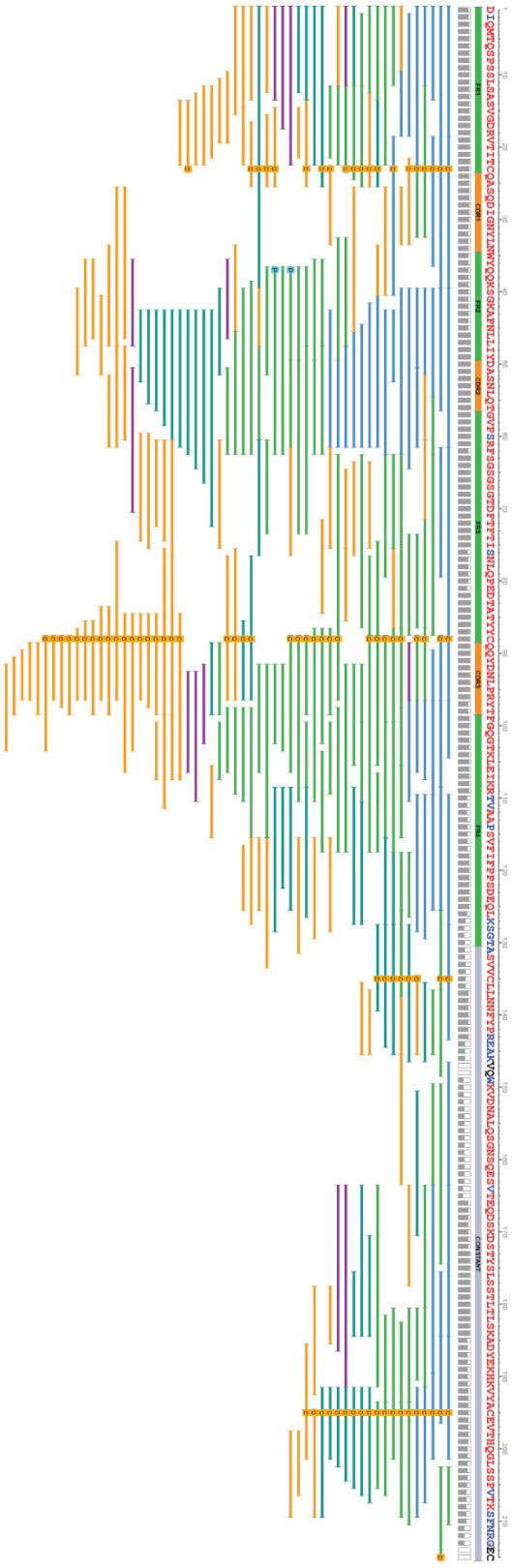


Figure 100: De novo sequencing results of FOR1007 using PEAKS AB. Mass accuracy for fragmented ions: 0.5 Da. Yellow C: carbamidomethylation. Peptides: light blue (trypsin), light green (chymotrypsin), yellow (elastase), violet (pepsin), and cyan (LysC). Above the peptides (top to bottom): Amino acid position, sequence, CDRs, FRs, confidence score per amino acid. Confidence scores are also illustrated by the different colors of the amino acids: red (>95 %), dark blue (>85 %), bold black (<85 %), and black (not verified suggestions).

Table 16: GL sequences of all cases created by combining the V, J and C segments of the IMGT/GENE database **[116]** *to a gap-free sequence.*

Case	Protein sequence
FOR009	SYELTQPPSVSVSPGQTASITCSGDKLGDKYACWYQQKPGQSPVLVIYQDSKRPSGIPERFSGS NSGNTATLTISGTQAMDEADYYCQAWDSSTAVVFGGGTKLTVLGQPKAAPSVTLFPPSSEEL QANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQWKS HRSYSCQVTHEGSTVEKTVAPTECS
FOR010	SSELTQDPAVSVALGQTVRITCQGDSLRSYYASWYQQKPGQAPVLVIYGKNNRPSGIPDRFSG SSSGNTASLTITGAQAEDEADYYCNSRDSSGNHWVFGGGTKLTVLGQPKAAPSVTLFPPSSEE LQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQWK SHKSYSCQVTHEGSTVEKTVAPTECS
FOR012	SYELTQPPSVSVSPGQTASITCSGDKLGDKYACWYQQKPGQSPVLVIYQDSKRPSGIPERFSGS NSGNTATLTISGTQAMDEADYYCQAWDSSTAWVFGGGTKLTVLGQPKAAPSVTLFPPSSEEL QANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQWKS HKSYSCQVTHEGSTVEKTVAPTECS
FOR101	QSALTQPASVSGSPGQSITISCTGTSSDVGGYNYVSWYQQHPGKAPKLMIYDVSNRPSGVSNR FSGSKSGNTASLTISGLQAEDEADYYCSSYTSSSTLYVFGTGTKVTVLGQPKANPTVTLFPPSS EELQANKATLVCLISDFYPGAVTVAWKADGSPVKAGVETTKPSKQSNNKYAASSYLSLTPEQ WKSHRSYSCQVTHEGSTVEKTVAPTECS
FOR102	QSVLTQPPSASGTPGQRVTISCSGSSSNIGSNTVNWYQQLPGTAPKLLIYSNNQRPSGVPDRFS GSKSGTSASLAISGLQSEDEADYYCAAWDDSLNGYVFGTGTKVTVLGQPKANPTVTLFPPSSE ELQANKATLVCLISDFYPGAVTVAWKADGSPVKAGVETTKPSKQSNNKYAASSYLSLTPEQW KSHRSYSCQVTHEGSTVEKTVAPTECS
FOR103	SSELTQDPAVSVALGQTVRITCQGDSLRSYYASWYQQKPGQAPVLVIYGKNNRPSGIPDRFSG SSSGNTASLTITGAQAEDEADYYCNSRDSSGNHVVFGGGTKLTVLGQPKAAPSVTLFPPSSEEL QANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQWKS HRSYSCQVTHEGSTVEKTVAPTECS
FOR107	QSVLTQPPSASGTPGQRVTISCSGSSSNIGSNTVNWYQQLPGTAPKLLIYSNNQRPSGVPDRFS GSKSGTSASLAISGLQSEDEADYYCAAWDDSLNGVVFGGGTKLTVLGQPKAAPSVTLFPPSSE ELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQW KSHRSYSCQVTHEGSTVEKTVAPTECS
FOR124	QSALTQPASVSGSPGQSITISCTGTSSDVGGYNYVSWYQQHPGKAPKLMIYDVSNRPSGVSNR FSGSKSGNTASLTISGLQAEDEADYYCSSYTSSSTLWVFGGGTKLTVLGQPKAAPSVTLFPPSS EELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQ WKSHKSYSCQVTHEGSTVEKTVAPTECS
FOR128	QSVLTQPPSASGTPGQRVTISCSGSSSNIGSNTVNWYQQLPGTAPKLLIYSNNQRPSGVPDRFS GSKSGTSASLAISGLQSEDEADYYCAAWDDSLNGWVFGGGTKLTVLGQPKAAPSVTLFPPSSE ELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQW KSHRSYSCQVTHEGSTVEKTVAPTECS
FOR137	QSVLTQPPSASGTPGQRVTISCSGSSSNIGSNTVNWYQQLPGTAPKLLIYSNNQRPSGVPDRFS GSKSGTSASLAISGLQSEDEADYYCAAWDDSLNGVVFGGGTKLTVLGQPKAAPSVTLFPPSSE ELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQW KSHRSYSCQVTHEGSTVEKTVAPTECS
FOR138	QSVLTQPPSVSGAPGQRVTISCTGSSSNIGAGYDVHWYQQLPGTAPKLLIYGNSNRPSGVPDRF SGSKSGTSASLAITGLQAEDEADYYCQSYDSSLSGYVFGTGTKVTVLGQPKANPTVTLFPPSSE

ELQANKATLVCLISDFYPGAVTVAWKADGSPVKAGVETTKPSKQSNNKYAASSYLSLTPEQW
KSHRSYSCQVTHEGSTVEKTVAPTECS

FOR140 NFMLTQPHSVSESPGKTVTISCTRSSGSIASNYVQWYQQRPGSSPTTVIYEDNQRPSGVPDRFS
GSIDSSSNSASLTISGLKTEDEADYYCQSYDSSNWVFGGGTKLTVLGQPKAAPSVTLFPPSSEE
LQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQWK
SHRSYSCQVTHEGSTVEKTVAPTECS

FOR142 QSVLTQPPSASGTPGQRVTISCSGSSSNIGSNTVNWYQQLPGTAPKLLIYSNNQRPSGVPDRFS
GSKSGTSASLAISGLQSEDEADYYCAAWDDSLNGVVFGGGTKLTVLGQPKAAPSVTLFPPSSE
ELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQW
KSHRSYSCQVTHEGSTVEKTVAPTECS

FOR147 QSVLTQPPSASGTPGQRVTISCSGSSSNIGSNTVNWYQQLPGTAPKLLIYSNNQRPSGVPDRFS
GSKSGTSASLAISGLQSEDEADYYCAAWDDSLNGVVFGGGTKLTVLGQPKAAPSVTLFPPSSE
ELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQW
KSHRSYSCQVTHEGSTVEKTVAPTECS

FOR157 QSALTQPASVSGSPGQSITISCTGTSSDVGGYNYVSWYQQHPGKAPKLMIYDVSNRPSGVSNR
FSGSKSGNTASLTISGLQAEDEADYYCSSYTSSSTLYVFGTGTKVTVLGQPKANPTVTLFPPSS
EELQANKATLVCLISDFYPGAVTVAWKADGSPVKAGVETTKPSKQSNNKYAASSYLSLTPEQ
WKSHRSYSCQVTHEGSTVEKTVAPTECS

FOR159 QSALTQPASVSGSPGQSITISCTGTSSDVGGYNYVSWYQQHPGKAPKLMIYDVSNRPSGVSNR
FSGSKSGNTASLTISGLQAEDEADYYCSSYTSSSTLWVFGGGTKLTVLGQPKAAPSVTLFPPSS
EELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQ
WKSHKSYSCQVTHEGSTVEKTVAPTECS

FOR161 SYELTQPPSVSVSPGQTASITCSGDKLGDKYACWYQQKPGQSPVLVIYQDSKRPSGIPERFSGS
NSGNTATLTISGTQAMDEADYYCQAWDSSTAYVFGTGTKVTVLGQPKANPTVTLFPPSSEEL
QANKATLVCLISDFYPGAVTVAWKADGSPVKAGVETTKPSKQSNNKYAASSYLSLTPEQWKS
HRSYSCQVTHEGSTVEKTVAPTECS

FOR1001 DIQMTQSPSSLSASVGDRVTITCQASQDISNYLNWYQQKPGKAPKLLIYDASNLETGVPSRFSG
SGSGTDFTFTISSLQPEDIATYYCQQYDNLPYTFGQGTKLEIKRTVAAPSVFIFPPSDEQLKSGT
ASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKV
YACEVTHQGLSSPVTKSFNRGEC

FOR1007 DIQMTQSPSSLSASVGDRVTITCQASQDISNYLNWYQQKPGKAPKLLIYDASNLETGVPSRFSG
SGSGTDFTFTISSLQPEDIATYYCQQYDNLPYTFGQGTKLEIKRTVAAPSVFIFPPSDEQLKSGT
ASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKV
YACEVTHQGLSSPVTKSFNRGEC

Table 17: Precursor LC sequences derived from cDNA sequencing results from Heidelberg University Hospital. Yellow background: missing N-terminal amino acids. Red background: missing C-terminal amino acids. Violet X: Uncertainties. The N-terminal gaps are filled with amino acids provided by MS, and the C-terminal gaps are filled with the corresponding GL sequence. No cDNA was provided for FOR009, FOR010, and FOR012.

Case	Protein sequence
FOR009	-
FOR010	-
FOR012	-
FOR101	QSALTQPDSVSGSPGQSITISCTGSSSDVGGYNYVSWYQQHPGKAPRLIIYHVNNRPSGVSNRF SGSKSGNTASLTISGLQAEDEAEYYCGSFISSNIPYVFATGTXVTVLGQPKANPTVTLFPPSSEE LQANKATLVCLISDFYPGAVTVAWKADGSPVKAGVETTKPSKQSNNKYAASSYLSLTPEQW KSHRSYSCQVTHEGSTVEKTVAPTECS
FOR102	QSVLTQPPSASGTPGQRVIISCSGSSSNIGSNTVSWYLHLPGTAPKLLIYTNTQRPSGVGDRFSG SKSGTSASLAISGLQSGDEADYYCAAWDDSLRGPVFGSGTKVTVLGQPKANPTVTLFPPSSEE LQANKATLVCLISDFYPGAVTVAWKADGSPVKAGVETTKPSKQSNNKYAASSYLSLTPEQW KSHRSYSCQVTHEGSTVEKTVAPTECS
FOR103	SSELTQDPAVSVALGQTVRITCQGDSLRSYYASWYQQKSGQAPVLVIYSYNNRPSGIPDRFSG SNSGNTASLTITGAQAEDEADYYCNSRDSSGHHLVFGGGTKLTVLGQPKAAPSVTLFPPSSEE LQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQWK SHRSYSCQVTHEGSTVEKTVAPTECS
FOR107	QSVLTQPPSASGTPGQRVTISCSGSGSNIGGNSVNWYQHLPGAAPKFLIYSNNQRPSGVPDRFS GSKSGTSASLAISGLQSEDEADYYCATWDDSLNGVIFGGGTKLTVLGQPKAAPSVTLFPPSSEE LQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQWK SHRSYSCQVTHEGSTVEKTVAPTECS
FOR124	QSALTQPASVSGSPGQSITISCTGTSSDIGGHNFVSWYQQNPGKAPKLIIYDVTNRPSGVSNRFS GSKSGNTASLSISGLQAEDEADYYCSSYTSSSTPPWVFGGGTKVTVLGQPKAAPSVTLFPPSSE ELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQW KSHKSYSCQVTHEGSTVEKTVAPTECS
FOR128	QSVLTQPPSASGTPGRGVTISCSGGSSNIGINTVTWYQQVPGGAPKLLIYSNNQRSSGVPDRFS GSKSGTSASLAISGLQSEDEADYYCAAWEDTLNGWVFGGGTKLTVLGQPKAAPSVTLFPPSSE ELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQW KSHRSYSCQVTHEGSTVEKTVAPTECS
FOR137	QSELTQPPSASGTPGQRVTISCSGGSSNIGSNFVNWYQHLPGTAPKLLIYTINQRPSGVPDRFSG SKSGTSASLAISGLRSEDEADYYCSAWDDSLNGPVFGGGTKLTVLGQPKAAPSVTLFPPSSEEL QANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQWKS HRSYSCQVTHEGSTVEKTVAPTECS
FOR138	QSVLTQPPSVSGAPGQRVIISCTGTSSNIGAGYDVHWYQQLPGTAPKIVIYGNNNRPSGVPDRF SGSKSGTSASLAITGLQAEDEADYYCHSYERGLSGEGVFGTGTKVTVLSQPKANPTVTLFPPSS EELQANKATLVCLISDFYPGAVTVAWKADGSPVKAGVETTKPSKQSNNKYAASSYLSLTPEQ WKSHRSYSCQVTHEGSTVEKTVAPTECS
FOR140	NFMLTQPHSVSESPGKTVTISCTRSSGSIASDYVQWYQQRPGSAPTTVIFENDQRPSGVPARFS GSIDSSSDSASLTISGLTTEDEADYYCQSYDSSAFWVFGGGTRLTVLSQPKAAPSVTLFPPSSEE LQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQWK SHRSYSCQVTHEGSTVEKTVAPTECS

FOR142 QSVLTQPPSASGTPGQRVTISCSGSSSNLGSNTVNWYQHLPGXAPKLLIYSNNQRPSGVSDRFS
GSKSGTSASLAIGGLQSDDESDYYCAAWDDSLNGPVFGGGTKLTVLGQPKAAPSVTLFPPSSE
ELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQW
KSHRSYSCQVTHEGSTVEKTVAPTECS

FOR147 QSVLTQPPSASGTPGQRVTISCSGGSSNIGSNSVTWYQQLPGTAPKLLMYITNQRPSGVPDRFS
GSKSGTSASLAISGLHSEDEADYYCAAWDNSLNGPVFGGGTKLTVLGQPKAAPSVTLFPPSSE
ELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQW
KSHRSYSCQVTHEGSTVEKTVAPTECS

FOR157 QSALTQPASVSGSPGQSITISCTGTSSDVGGYYYVSWYQQHPGKVPKLMIFDVSNRPSGVSDR
FSGSKSGNTASLTISGLQAEDEADYYCSSYTSSGTGVFGXGTKVTVLGQPKANPTVTLFPPSSE
ELQANKATLVCLISDFYPGAVTVAWKADGSPVKAGVETTKPSKQSNNKYAASSYLSLTPEQW
KSHRSYSCQVTHEGSTVEKTVAPTECS

FOR159 QSALTQPASVSGSPGQSITISCTGSSSDVGSYNYVSWYQQXPGKAPKLMIYDVTNRPSGVSNR
FSGSKSGNTASLTISGLQAEDEADYYCSSYTATSALGVFGGGTKLTVLGQPKAAPSVTLFPPSS
EELQXNKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQ
WKSHKSYSCQVTHEGSTVEKTVAPTECS

FOR161 SYELTQPPSVSVSPGQTASITCSGDKLGDNYASWYQQKPGQSPVLVIYQDSRRPSGIPERFSGS
XXGNTATLTISGTQAMDEADYYCQAWDSSTGVFGPGTTVTVLGQPKANPTVTLFPPSSEELQ
ANKATLVCLISDFYPGAVTVAWKADGSPVKAGVETTKPSKQSNNKYAASSYLSLTPEQWKSH
RSYSCQVTHEGSTVEKTVAPTECS

FOR1001 DIQMTQSPSSLSSSVGDRVTITCQASQDIGNYLVWYQQKPGKAPNLLIYDASNLETGVSSRFSG
SGSATHFSLTISSLQTEDVATYYCQQYESLPYTFGQGTKVELEIKRTVAAPSVFIFPPSDEQLKS
GTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKH
KVYACEVTHQGLSSPVTKSFNRGEC

FOR1007 DIQMTQSPSSLSASVGDRVTITCQASQDIGNYLNWYQQKSGKAPNLLIYDASNLQTGVPSRFS
GSGSGTDFTFTISNLQPEDTATYYCQQYDNLPRYTFGQGTKLEIKRTVAAPSVFIFPPSDEQLKS
GTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKH
KVYACEVTHQGLSSPVTKSFNRGEC

Table 18: Precursor LC sequences of all cases resulting from MS in combination with cDNA information from Heidelberg University Hospital. Those sequences have been used for all biochemical studies of this thesis.

Case	Protein sequence
FOR009	SYELTQTPSVSVSPGQTASITCSGDRLGDKYVSWYQQKPGQSPVLVIYQDNKRPSGIPERFSGS NSGDKATLTISGTQAMDEADYYCQAWDSSTAGFGGGTKLTVRGQPKAAPSVTLFPPSSEELQ ANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQWKSH RSYSCQVTHEGSTVEKTVAPTECS
FOR010	SSELTQDPAVSVPLGQTVRITCQGDSLRTYYGNWYQQKPGQAPLLVIYAENKRPSGIPDRFSG SISGNTASLTITGAQAEDEADYFCNSRDTTGNLWVFGGGTKLTVLGQPKAAPSVTLFPPSSEEL QANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQWKS HKSYSCQVTHEGSTVEKTVAPTECS
FOR012	SYELTQPPSVSVSPGQTASITCSGDKLGDEYASWYQQKPGQSPVLVIYQNIKRPSGIPERFSGSN SGNTVTLTISGTQAMDEADYYCQAWDSSTWVFGGGTKLSVLSQPKAAPSVTLFPPSSEELQA NKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQWKSHK SYSCQVTHEGSTVEKTVAPTECS
FOR101	QSALTQPDSVSGSPGQSITISCTGSSSDVGGYNYVSWYQQHPGKAPRLIIYHVNNRPSGVSNRF SGSKSGNTASLTISGLQAEDEAEYYCGSFISSNIPYVFATGTKVTVLGQPKANPTVTLFPPSSEE LQANKATLVCLISDFYPGAVTVAWKADGSPVKAGVETTKPSKQSNNKYAASSYLSLTPEQW KSHRSYSCQVTHEGSTVEKTVAPTECS
FOR102	QSVLTQPPSASGTPGQRVIISCSGSSSNIGSNTVSWYLHLPGTAPKLLIYTNTQRPSGVGDRFSG SKSGTSASLAISGLQSGDEADYYCAAWDDSLRGPVFGSGTKVTVLGQPKANPTVTLFPPSSEE LQANKATLVCLISDFYPGAVTVAWKADGSPVKAGVETTKPSKQSNNKYAASSYLSLTPEQW KSHRSYSCQVTHEGSTVEKTVAPTECS
FOR103	SSELTQDPAVSVALGQTVRITCQGDSLRSYYASWYQQKSGQAPVLVIYSYNNRPSGIPDRFSG SNSGNTASLTITGAQAEDEADYYCNSRDSSGHHLVFGGGTKLTVLGQPKAAPSVTLFPPSSEE LQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQWK SHRSYSCQVTHEGSTVEKTVAPTECS
FOR107	QSVLTQPPSASGTPGQRVTISCSGSGSNIGGNSVNWYQHLPGAAPKFLIYSNNQRPSGVPDRFS GSKSGTSASLAISGLQSEDEADYYCATWDDSLNGVIFGGGTKLTVLGQPKAAPSVTLFPPSSEE LQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQWK SHRSYSCQVTHEGSTVEKTVAPTECS
FOR124	QSALTQPASVSGSPGQSITISCTGTSSDIGGHNFVSWYQQNPGKAPKLIIYDVTNRPSGVSNRFS GSKSGNTASLSISGLQAEDEADYYCSSYTSSSTPPWVFGGGTKVTVLGQPKAAPSVTLFPPSSE ELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQW KSHKSYSCQVTHEGSTVEKTVAPTECS
FOR128	QSVLTQPPSASGTPGRGVTISCSGGSSNIGINTVTWYQQVPGGAPKLLIYSNNQRSSGVPDRFS GSKSGTSASLAISGLQSEDEADYYCAAWEDTLNGWVFGGGTKLTVLGQPKAAPSVTLFPPSSE ELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQW KSHRSYSCQVTHEGSTVEKTVAPTECS
FOR137	QSELTQPPSASGTPGQRVTISCSGGSSNIGSNFVNWYQHLPGTAPKLLIYTINQRPSGVPDRFSG SKSGTSASLAISGLRSEDEADYYCSAWDDSLNGPVFGGGTKLTVLGQPKAAPSVTLFPPSSEEL QANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQWKS HRSYSCQVTHEGSTVEKTVAPTECS
FOR138	QSVLTQPPSVSGAPGQRVIISCTGTSSNIGAGYDVHWYQQLPGTAPKIVIYGNNNRPSGVPDRF SGSKSGTSASLAITGLQAEDEADYYCHSYERGLSGEGVFGTGTKVTVLSQPKANPTVTLFPPSS

EELQANKATLVCLISDFYPGAVTVAWKADGSPVKAGVETTKPSKQSNNKYAASSYLSLTPEQ
WKSHRSYSCQVTHEGSTVEKTVAPTECS

FOR140 NFMLTQPHSVSESPGKTVTISCTRSSGSIASDYVQWYQQRPGSAPTTVIFEDDQRPSGVPARFS
GSIDSSSDSASLTISGLTTEDEADYYCQSYDSSAFWVFGGGTRLTVLSQPKAAPSVTLFPPSSEE
LQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQWK
SHRSYSCQVTHEGSTVEKTVAPTECS

FOR142 QSVLTQPPSASGTPGQRVTISCSGSSSNLGSNTVNWYQHLPGTAPKLLIYSNNQRPSGVSDRFS
GSKSGTSASLAIGGLQSDDESDYYCAAWDDSLNGPVFGGGTKLTVLGQPKAAPSVTLFPPSSE
ELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQW
KSHRSYSCQVTHEGSTVEKTVAPTECS

FOR147 QSVLTQPPSASGTPGQRVTISCSGGSSNIGSNSVTWYQQLPGTAPKLLMYNVNQRPSGVPDRF
SGSKSGTSASLAISGLHSEDEADYYCAAWDNSLNGPVFGGGTKLTVLGQPKAAPSVTLFPPSS
EELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQ
WKSHRSYSCQVTHEGSTVEKTVAPTECS

FOR157 QSALTQPASVSGSPGQSITISCTGTSSDVGGYYYVSWYQQHPGKVPKLMIFDVSNRPSGVSDR
FSGSKSGNTASLTISGLQAEDEADYYCSSYTSSGTGVFGTGTKVTVLGQPKANPTVTLFPPSSE
ELQANKATLVCLISDFYPGAVTVAWKADGSPVKAGVETTKPSKQSNNKYAASSYLSLTPEQW
KSHRSYSCQVTHEGSTVEKTVAPTECS

FOR159 QSALTQPASVSGSPGQSITISCTGSSSDVGSYNYVSWYQQHPGKAPKLMIYDVTNRPSGVSNR
FSGSKSGNTASLTISGLQAEDEADYYCSSYTATSALGVFGGGTKLTVLGQPKAAPSVTLFPPSS
EELQXNKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQ
WKSHKSYSCQVTHEGSTVEKTVAPTECS

FOR161 SYELTQPPSVSVSPGQTASITCSGDKLGDNYASWYQQKPGQSPVLVIYQDSRRPSGIPERFSGS
NSGNTATLTISGTQAMDEADYYCQAWDSSTGVFGPGTTVTVLGQPKANPTVTLFPPSSEELQA
NKATLVCLISDFYPGAVTVAWKADGSPVKAGVETTKPSKQSNNKYAASSYLSLTPEQWKSHR
SYSCQVTHEGSTVEKTVAPTECS

FOR1001 DIQMTQSPSSLSSSVGDRVTITCQASQDIGNYLVWYQQKPGKAPNLLIYDASNLETGVSSRFSG
SGSATHFSLTISSLQTEDVATYYCQQYESLPYTFGQGTKVELEIKRTVAAPSVFIFPPSDEQLKS
GTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKH
KVYACEVTHQGLSSPVTKSFNRGEC

FOR1007 DIQMTQSPSSLSASVGDRVTITCQASQDIGNYLNWYQQKSGKAPNLLIYDASNLQTGVPSRFS
GSGSGTDFTFTISNLQPEDTATYYCQQYDNLPRYTFGQGTKLEIKRTVAAPSVFIFPPSDEQLKS
GTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKH
KVYACEVTHQGLSSPVTKSFNRGEC

Table 19: pI calculations of the full-length LCs, V_L domains, C_L domains, and the corresponding GLs. P values of the two-tailed t-test and Cohen's d are shown below the table.

Case	pI						ΔpI (LC-GL)		
	GL			LC					
	Full-length	V_L	C_L	Full-length	V_L	C_L	Full-length	V_L	C_L
FOR138	7.0	5.6	7.9	7.1	6.1	7.7	0.1	0.5	-0.2
FOR102	6.9	5.0	7.9	8.3	8.0	7.9	1.4	3.0	0.0
FOR107	6.1	5.0	6.9	6.3	5.6	6.9	0.2	0.6	0.0
FOR128	6.1	5.0	6.9	6.2	5.1	6.9	0.1	0.1	0.0
FOR137	6.1	5.0	6.9	6.3	5.6	6.9	0.2	0.6	0.0
FOR142	6.1	5.0	6.9	6.3	5.5	6.9	0.2	0.5	0.0
FOR147	6.1	5.0	6.9	7.0	6.8	6.9	0.9	1.8	0.0
FOR101	7.8	6.7	7.9	7.8	6.9	7.9	0.0	0.2	0.0
FOR124	7.0	6.7	6.9	7.0	6.7	6.9	0.0	0.0	0.0
FOR157	7.8	6.7	7.9	7.0	5.6	7.9	-0.8	-1.1	0.0
FOR159	7.0	6.7	6.9	7.0	6.7	6.9	0.0	0.0	0.0
FOR009	5.7	4.8	6.9	6.1	5.2	6.9	0.4	0.4	0.0
FOR012	5.7	4.8	6.9	5.4	4.6	6.7	-0.3	-0.2	-0.2
FOR161	6.1	4.8	7.9	5.4	4.3	7.9	-0.7	-0.5	0.0
FOR010	6.3	5.6	6.9	5.7	4.9	6.9	-0.6	-0.7	0.0
FOR103	6.3	5.6	6.9	6.1	5.5	6.9	-0.2	-0.1	0.0
FOR140	5.6	4.8	6.9	5.0	4.4	6.7	-0.6	-0.4	-0.2
FOR1001	5.2	4.6	6.1	5.2	4.6	6.1	0.0	0.0	0.0
FOR1007	5.2	4.6	6.1	5.4	4.7	6.1	0.2	0.1	0.0
Mean ± SD	6.3 ± 0.8	5.4 ± 0.8	7.1 ± 0.6	6.3 ± 0.9	5.6 ± 1.0	7.0 ± 0.6	0.0 ± 0.5	0.3 ± 0.9	0.0 ± 0.1

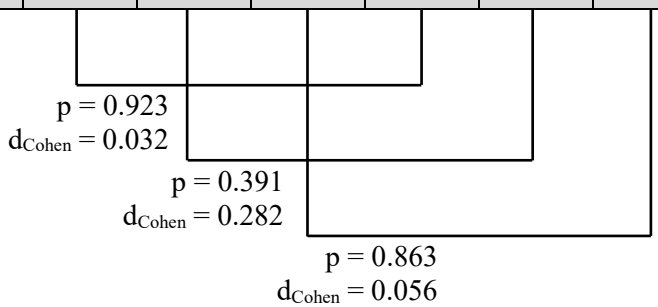


Table 20: pI calculations of the AL proteins (LC fragments) and the corresponding GL. Mean pI: sum of pI values of all fragments from one molecular mass divided by the number of fragments. Mean pI per case: mean of all mean pIs per case. The P value of the two-tailed t-test and Cohen's d are shown below the table.

Case	possible assignable fibril protein			pI (GL)			mean pI (GL)	mean pI (GL) per case	pI (LC fragments)			mean pI (LC frag.)	mean pI (LC frag.) per case	ΔpI (LC–GL)
FOR138	1–123	3–125	4–126	6.8	6.7	6.8	6.8	6.3	6.9	6.9	6.9	6.9	6.6	0.3
	3–104	4–105		5.0	5.0		5.0		5.5	5.5		5.5		
	3–121			6.7			6.7		6.9			6.9		
	3–123			6.7			6.7		6.9			6.9		
FOR102	3–125	4–126	6–128	6.3	6.3	6.3	6.3	6.3	8.7	8.7	8.7	8.7	8.7	2.4
FOR107	2–124	3–125	4–126	6.0	6.3	6.3	6.2	6.3	6.5	6.7	6.8	6.7	6.8	0.5
	4–125			6.3			6.3		6.8			6.8		
	8–123	9–124	10–125	6.6	6.0	6.3	6.3		7.1	6.5	6.8	6.8		
FOR128	1–125			6.3			6.3	6.3	6.3			6.3	6.3	0.0
	2–124	3–125	4–126	6.0	6.3	6.3	6.2		6.0	6.3	6.3	6.2		
			6–128			6.3					6.3			
FOR137	4–121			6.3			6.3	6.3	8.0			8.0	8.0	1.7
	4–125			6.3			6.3		8.0			8.0		
	16–121			6.3			6.3		8.0			8.0		
	17–121			6.3			6.3		8.0			8.0		
FOR142	2–124	3–125	4–126	6.0	6.3	6.3	6.2	6.2	6.5	6.7	6.7	6.7	6.7	0.5
FOR147	2–124	3–125	4–126	6.0	6.3	6.3	6.2	6.3	7.8	8.0	8.0	8.0	8.0	1.7
			6–128			6.3					8.0			
	4–125			6.3			6.3		8.0			8.0		
	8–123	9–124	10–125	6.6	6.0	6.3	6.3		8.3	7.8	8.0	8.0		
FOR101	4–123			8.0			8.0	8.0	8.0			8.0	8.0	0.0
	4–125			8.0			8.0		8.0			8.0		

FOR124	4–103			5.6			5.6	5.6	5.6			5.6	5.6	0.0
	4–104			5.6			5.6		5.6			5.6		
	4–105			5.6			5.6		5.6			5.6		
FOR157	4–123	5–124	6–125	8.0	7.7	8.0	7.9	8.0	6.8	6.4	6.8	6.7	6.8	-1.2
	4–125			8.0			8.0		6.8			6.8		
FOR159	1–125			8.0			8.0	7.9	8.0			8.0	7.9	0.0
	2–125			7.7			7.7		7.7			7.7		
	2–124	3–125		7.7	8.0		7.9		7.7	8.0		7.9		
	4–125			8.0			8.0		8.0			8.0		
FOR009	1–124	2–125		5.1	5.1		5.1	5.1	6.1	6.4		6.3	6.4	1.3
	2–127			5.1			5.1		6.4			6.4		
FOR012	4–125			6.3			6.3	6.3	5.1			5.1	5.1	-1.2
FOR161	6–117	12–123		5.0	6.3		5.7	6.0	4.4	4.7		4.6	4.7	-1.3
	12–125	13–126		6.3	6.0		6.2		4.7	4.7		4.7		
FOR010	4–121			8.0			8.0	8.3	6.3			6.3	7.0	-1.3
	4–125			8.0			8.0		6.3			6.3		
	8–125			8.8			8.8		8.3			8.3		
FOR103	1–124	2–125	7–130	6.5	6.5	8.0	7.0	7.0	6.0	6.0	6.9	6.3	6.3	-0.7
FOR140	1–125			5.2			5.2	5.2	4.5			4.5	4.5	-0.7
FOR1001	1–105			4.3			4.3	4.3	4.5			4.5	4.5	0.2
FOR1007	1–103			4.3			4.3	4.3	4.4			4.4	4.4	0.1
	1–105			4.3			4.3		4.4			4.4		
	1–115			4.4			4.4		4.5			4.5		
Mean ± SD								6.3 ± 1.2					6.4 ± 1.3	0.1 ± 1.1

p = 0,768
d_{Cohen} = 0,097

Table 21: Solubility calculations of the full-length LCs, V_L domains, C_L domains, and the corresponding GLs. The P values of the two-tailed t-test and Cohen's d are shown below the table. The values represent CamSol solubility scores **[120].**

Case	Solubility						ΔSolubility (LC-GL)		
	GL			LC					
	Full-length	V_L	C_L	Full-length	V_L	C_L	Full-length	V_L	C_L
FOR138	0.5	0.6	1.0	0.6	0.8	1.0	0.1	0.2	0.0
FOR102	0.7	0.9	1.0	0.6	0.8	1.0	-0.1	-0.1	0.0
FOR107	0.7	1.0	1.0	0.7	0.9	1.0	0.0	-0.1	0.0
FOR128	0.7	1.0	1.0	0.5	0.7	1.0	-0.2	-0.3	0.0
FOR137	0.7	1.0	1.0	0.7	1.0	1.0	0.0	0.0	0.0
FOR142	0.7	1.0	1.0	0.8	1.2	1.0	0.1	0.2	0.0
FOR147	0.7	1.0	1.0	0.7	1.0	1.0	0.0	0.0	0.0
FOR101	0.2	0.2	1.0	0.1	0.1	1.0	-0.1	-0.1	0.0
FOR124	0.2	0.3	1.0	0.4	0.5	1.0	0.2	0.2	0.0
FOR157	0.2	0.2	1.0	0.2	0.2	1.0	0.0	0.0	0.0
FOR159	0.2	0.3	1.0	0.3	0.5	1.0	0.1	0.2	0.0
FOR009	0.4	0.6	1.0	0.7	1.1	1.0	0.3	0.5	0.0
FOR012	0.4	0.6	1.0	0.2	0.4	0.9	-0.2	-0.2	-0.1
FOR161	0.4	0.5	1.0	0.5	0.6	1.0	0.1	0.1	0.0
FOR010	0.5	0.7	1.0	0.5	0.7	1.0	0.0	0.0	0.0
FOR103	0.5	0.7	1.0	0.3	0.5	1.0	-0.2	-0.2	0.0
FOR140	0.6	0.8	1.0	0.4	0.5	1.0	-0.2	-0.3	0.0
FOR1001	0.2	0.5	0.7	-0.1	0.2	0.7	-0.3	-0.3	0.0
FOR1007	0.2	0.5	0.7	0.2	0.6	0.7	0.0	0.1	0.0
Mean ± SD	0.5 ± 0.2	0.6 ± 0.3	0.9 ± 0.1	0.4 ± 0.2	0.6 ± 0.3	1.0 ± 0.1	0.0 ± 0.2	0.0 ± 0.2	0.0 ± 0.0

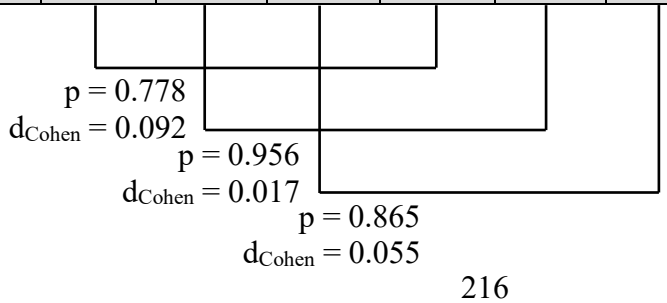


***Table 22: Solubility calculations of the AL proteins (LC fragments) and the corresponding GL. Mean Sol.: sum of solubility values of all fragments from one molecular mass divided by the number of fragments. Mean Sol. per case: mean of all mean solubilities per case. The P value of the two-tailed t-test and Cohen's d are shown below the table. Values are CamSol solubility* scores [120].**

Case	possible assignable fibril protein			Solubility (Sol.) (GL)				mean Sol. (GL)	mean Sol. (GL) per case	Sol. (LC fragments)				mean Sol. (LC frag.)	mean Sol. (LC frag.) per case	ΔSol. (LC–GL)
FOR138	1–123	3–125	4–126	6–128	0.8	0.8	0.9	0.8	0.9	1.0	1.0	1.1	0.9	1.0	1.0	
	3–104	4–105			1.1	1.1		1.1		1.1	1.1			1.1		
	3–121				0.8			0.8		1.0				1.0		
	3–123				0.8			0.8		1.0				1.0		
FOR102	3–125	4–126	6–128		1.1	1.2	1.0	1.1	1.1	1.0	1.1	0.9		1.0	1.0	-0.1
FOR107	2–124	3–125	4–126	6–128	1.2	1.2	1.1	1.2	1.2	1.1	1.1	1.2	1.1	1.1	1.2	0.0
	4–125				1.3			1.3		1.2				1.2		
	8–123	9–124	10–125		1.2	1.2	1.2	1.2		1.2	1.2	1.2		1.2		
FOR128	1–125				1.2			1.2	1.2	0.9				0.9	0.9	-0.3
	2–124	3–125	4–126	6–128	1.2	1.2	1.1	1.2		0.9	0.9	1.0	0.9	0.9		
FOR137	4–121				1.2			1.2	1.2	1.2				1.2	1.1	-0.1
	4–125				1.3			1.3		1.2				1.2		
	16–121				1.2			1.2		1.1				1.1		
	17–121				1.1			1.1		1.0				1.0		
FOR142	2–124	3–125	4–126	6–128	1.2	1.2	1.1	1.2	1.2	1.4	1.4	1.5	1.3	1.4	1.4	0.2
FOR147	2–124	3–125	4–126	6–128	1.2	1.2	1.1	1.2	1.2	1.2	1.2	1.3	1.2	1.2	1.3	0.1
	4–125				1.3			1.3		1.3				1.3		
	8–123	9–124	10–125		1.2	1.2	1.2	1.2		1.2	1.3	1.3		1.3		
FOR101	4–123				0.4			0.4	0.4	0.3				0.3	0.3	-0.1
	4–125				0.4			0.4		0.3				0.3		

FOR124	4–103			0.8			0.8	0.6	0.9			0.9	0.9	0.3
	4–104			0.7			0.7		0.9			0.9		
	4–105			0.4			0.4		0.8			0.8		
FOR157	4–123	5–124	6–125	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.0
	4–125			0.4			0.4		0.4			0.4		
FOR159	1–125			0.5			0.5	0.5	0.7			0.7	0.6	0.1
	2–125			0.5			0.5		0.6			0.6		
	2–124	3–125		0.5	0.5		0.5		0.6	0.6		0.6		
	4–125			0.5			0.5		0.6			0.6		
FOR009	1–124	2–125		0.8	0.8		0.8	0.8	1.2	1.3		1.3	1.3	0.5
	2–127			0.8			0.8		1.2			1.2		
FOR012	4–125			0.8			0.8	0.8	0.6			0.6	0.6	-0.2
FOR161	6–117	12–123		0.5	0.7		0.6	0.7	0.6	0.9		0.8	0.9	0.2
	12–125	13–126		0.7	0.7		0.7		0.9	0.9		0.9		
FOR010	4–121			0.8			0.8	0.8	0.8			0.8	0.8	0.0
	4–125			0.8			0.8		0.8			0.8		
	8–125			0.8			0.8		0.8			0.8		
FOR103	1–124	2–125	7–130	0.9	0.9	0.6	0.8	0.8	0.7	0.7	0.4	0.6	0.6	-0.2
FOR140	1–125			1.0			1.0	1.0	0.7			0.7	0.7	-0.3
FOR1001	1–105			0.5			0.5	0.5	0.1			0.1	0.1	-0.4
FOR1007	1–103			0.5			0.5	0.5	0.6			0.6	0.6	0.1
	1–105			0.5			0.5		0.6			0.6		
	1–115			0.5			0.5		0.6			0.6		
Mean ± SD								0.8 ± 0.3					0.8 ± 0.4	0.0 ± 0.2

p = 0.961
d_{Cohen} = 0.016

Table 23: GRAVY calculations of the full-length LCs, V_L domains, C_L domains, and the corresponding GLs. The P values of the two-tailed t-test and Cohen's d are shown below the table.

Case	GRAVY						ΔGRAVY (LC-GL)		
	GL			LC					
	Full-length	V_L	C_L	Full-length	V_L	C_L	Full-length	V_L	C_L
FOR138	-0.4	-0.3	-0.5	-0.4	-0.3	-0.5	0.0	0.0	0.0
FOR102	-0.4	-0.4	-0.5	-0.3	-0.2	-0.5	0.1	0.2	0.0
FOR107	-0.4	-0.3	-0.4	-0.4	-0.3	-0.4	0.0	0.0	0.0
FOR128	-0.4	-0.4	-0.4	-0.3	-0.2	-0.4	0.1	0.2	0.0
FOR137	-0.4	-0.3	-0.4	-0.4	-0.4	-0.4	0.0	-0.1	0.0
FOR142	-0.4	-0.3	-0.4	-0.4	-0.4	-0.4	0.0	-0.1	0.0
FOR147	-0.4	-0.3	-0.4	-0.4	-0.3	-0.4	0.0	0.0	0.0
FOR101	-0.4	-0.3	-0.5	-0.4	-0.3	-0.5	0.0	0.0	0.0
FOR124	-0.4	-0.3	-0.4	-0.4	-0.3	-0.4	0.0	0.0	0.0
FOR157	-0.4	-0.3	-0.5	-0.4	-0.3	-0.5	0.0	0.0	0.0
FOR159	-0.4	-0.3	-0.4	-0.4	-0.3	-0.4	0.0	0.0	0.0
FOR009	-0.4	-0.4	-0.4	-0.5	-0.6	-0.4	-0.1	-0.2	0.0
FOR012	-0.4	-0.4	-0.4	-0.4	-0.4	-0.4	0.0	0.0	0.0
FOR161	-0.5	-0.4	-0.5	-0.5	-0.5	-0.5	0.0	-0.1	0.0
FOR010	-0.5	-0.5	-0.4	-0.4	-0.4	-0.4	0.1	0.1	0.0
FOR103	-0.5	-0.5	-0.4	-0.4	-0.4	-0.4	0.1	0.1	0.0
FOR140	-0.5	-0.6	-0.4	-0.4	-0.4	-0.4	0.1	0.2	0.0
FOR1001	-0.5	-0.5	-0.5	-0.4	-0.3	-0.5	0.1	0.2	0.0
FOR1007	-0.5	-0.5	-0.5	-0.6	-0.6	-0.5	-0.1	-0.1	0.0
Mean ± SD	-0.4 ± 0.0	-0.4 ± 0.1	-0.4 ± 0.0	-0.4 ± 0.1	-0.4 ± 0.1	-0.4 ± 0.0	0.0 ± 0.1	0.0 ± 0.1	0.0 ± 0.0

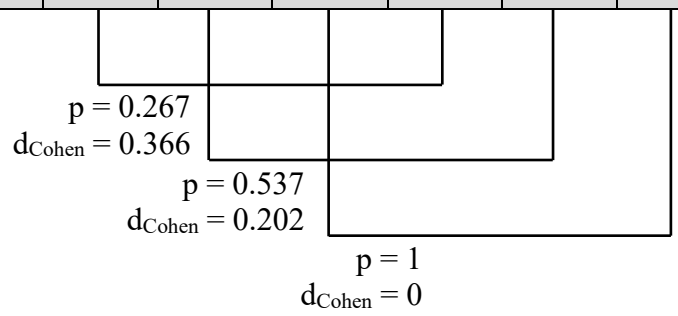


Table 24: GRAVY calculations of the AL proteins (LC fragments) and the corresponding GL. Mean GRAVY: sum of GRAVY values of all fragments from one molecular mass divided by the number of fragments. Mean GRAVY per case: mean of all mean GRAVYs per case. The P value of the two-tailed t-test and Cohen's d are shown below the table.

Case	possible assignable fibril protein			GRAVY (GL)			mean GRAVY (GL)	mean GRAVY (GL) per case	GRAVY (LC fragments)			mean GRAVY (LC frag.)	mean GRAVY (LC frag.) per case	ΔGRAVY (LC–GL)
FOR138	1–123	3–125	6–128	-0.4	-0.3	-0.3	-0.3	-0.3	-0.4	-0.4	-0.4	-0.4	-0.4	-0.1
	3–104	4–105		-0.4	-0.4		-0.4		-0.4	-0.4		-0.4		
	3–121			-0.3			-0.3		-0.3			-0.3		
	3–123			-0.3			-0.3		-0.3			-0.3		
FOR102	3–125	4–126	6–128	-0.4	-0.4	-0.4	-0.4	-0.4	-0.2	-0.2	-0.2	-0.2	-0.2	0.2
FOR107	2–124	3–125	4–126	-0.3	-0.3	-0.3	-0.3	-0.4	-0.3	-0.3	-0.3	-0.3	-0.3	0.1
	4–125			-0.4			-0.4		-0.4			-0.4		
	8–123	9–124	10–125	-0.4	-0.4	-0.4	-0.4		-0.3	-0.3	-0.3	-0.3		
FOR128	1–125			-0.4			-0.4	-0.4	-0.3			-0.3	-0.3	0.1
	2–124	3–125	4–126	-0.4	-0.4	-0.4	-0.4		-0.3	-0.3	-0.3	-0.3		
FOR137	4–121			-0.4			-0.4	-0.4	-0.4			-0.4	-0.3	0.1
	4–125			-0.4			-0.4		-0.3			-0.3		
	16–121			-0.4			-0.4		-0.3			-0.3		
	17–121			-0.4			-0.4		-0.3			-0.3		
FOR142	2–124	3–125	4–126	-0.3	-0.3	-0.3	-0.3	-0.3	-0.4	-0.4	-0.4	-0.4	-0.4	-0.1
FOR147	2–124	3–125	4–126	-0.3	-0.3	-0.3	-0.3	-0.4	-0.3	-0.3	-0.3	-0.3	-0.4	0.0
	4–125			-0.4			-0.4		-0.4			-0.4		
	8–123	9–124	10–125	-0.4	-0.4		-0.4		-0.4	-0.4	-0.4	-0.4		
FOR101	4–123			-0.4			-0.4	-0.4	-0.4			-0.4	-0.4	0.0
	4–125			-0.4			-0.4		-0.4			-0.4		

FOR124	4–103		-0.5		-0.5	-0.5	-0.5		-0.5	-0.5	0.0
	4–104		-0.5		-0.5		-0.5		-0.5		
	4–105		-0.4		-0.4		-0.4		-0.4		
FOR157	4–123	5–124 6–125	-0.4	-0.4 -0.4	-0.4	-0.4	-0.3	-0.4 -0.4	-0.4	-0.4	0.0
	4–125		-0.4		-0.4		-0.4		-0.4		
FOR159	1–125		-0.4		-0.4	-0.4	-0.3		-0.3	-0.3	0.1
	2–125		-0.3		-0.3		-0.3		-0.3		
	2–124	3–125	-0.3	-0.3	-0.3		-0.3	-0.3	-0.3		
	4–125		-0.4		-0.4		-0.3		-0.3		
FOR009	1–124	2–125	-0.4	-0.4	-0.4	-0.4	-0.6	-0.6	-0.6	-0.6	-0.2
	2–127		-0.4		-0.4		-0.6		-0.6		
FOR012	4–125		-0.4		-0.4	-0.4	-0.4		-0.4	-0.4	0.0
FOR161	6–117	12–123	-0.4	-0.5	-0.5	-0.5	-0.5	-0.5	-0.5	-0.5	0.0
	12–125	13–126	-0.5	-0.5	-0.5		-0.5	-0.5	-0.5		
FOR010	4–121		-0.5		-0.5	-0.5	-0.4		-0.4	-0.4	0.1
	4–125		-0.5		-0.5		-0.4		-0.4		
	8–125		-0.5		-0.5		-0.4		-0.4		
FOR103	1–124	2–125 7–130	-0.5	-0.5 -0.4	-0.5	-0.5	-0.5	-0.5 -0.4	-0.5	-0.5	0.0
FOR140	1–125		-0.6		-0.6	-0.6	-0.4		-0.4	-0.4	0.2
FOR1001	1–105		-0.5		-0.5	-0.5	-0.3		-0.3	-0.3	0.2
FOR1007	1–103		-0.5		-0.5	-0.5	-0.6		-0.6	-0.6	-0.1
	1–105		-0.5		-0.5		-0.6		-0.6		
	1–115		-0.5		-0.5		-0.6		-0.6		
Mean ± SD						-0.4 ± 0.1				-0.4 ± 0.1	0.0 ± 0.1

$p = 0.295$
$d_{Cohen} = 0.345$

Table 25: AGG calculations of the full-length LC, V_L domains, C_L domains, and the corresponding GLs. The P values of the two-tailed t-test and Cohen's d are shown below the table.

Case	AGG						ΔAGG (LC-GL)		
	GL			LC					
	Full-length	V_L	C_L	Full-length	V_L	C_L	Full-length	V_L	C_L
FOR138	2.9	0.5	5.5	3.0	0.7	5.5	0.1	0.2	0.0
FOR102	2.9	0.4	5.5	3.7	2.1	5.5	0.8	1.7	0.0
FOR107	2.9	0.4	5.5	2.9	0,5	5.5	0.0	0.1	0.0
FOR128	2.8	0.3	5.5	3.4	1.4	5.5	0.6	1.1	0.0
FOR137	2.9	0.4	5.5	5.0	4.5	5.5	2.1	4.1	0.0
FOR142	2.9	0.4	5.5	2.8	0.2	5.5	-0.1	-0.2	0.0
FOR147	2.9	0.4	5.5	2.8	0.3	5.5	-0.1	-0.1	0.0
FOR101	5.0	4.6	5.5	3.8	2.3	5.5	-1.2	-2.3	0.0
FOR124	5.0	4.6	5.5	3.5	1.7	5.5	-1.5	-2.9	0.0
FOR157	5.0	4.6	5.5	5.5	5.5	5.5	0.5	0.9	0.0
FOR159	5.0	4.6	5.5	3.5	1.7	5.5	-1.5	-2.9	0.0
FOR009	6.3	7.1	5.5	6.1	6.7	5.5	-0.2	-0.4	0.0
FOR012	5.7	6.0	5.5	5.8	6.1	5.5	0.1	0.1	0.0
FOR161	5.8	6.0	5.5	5.7	4.5	5.5	-0.1	-1.5	0.0
FOR010	5.2	5.0	5.5	5.6	5.8	5.5	0.4	0.8	0.0
FOR103	5.2	5.0	5.5	5.7	5.9	5.5	0.5	0.9	0.0
FOR140	3.0	0.5	5.5	4.9	4.4	5.5	1.9	3.9	0.0
FOR1001	2.1	1.2	3.1	3.9	4.8	3.1	1.8	3.6	0.0
FOR1007	2.1	1.2	3.1	2.1	1.1	3.1	0.0	-0.1	0.0
Mean ± SD	4.0 ± 1.4	2.8 ± 2.5	5.2 ± 0.8	4.2 ± 1.3	3.2 ± 2.3	5.2 ± 0.8	0.2 ± 1.0	0.4 ± 2.0	0.0 ± 0.0

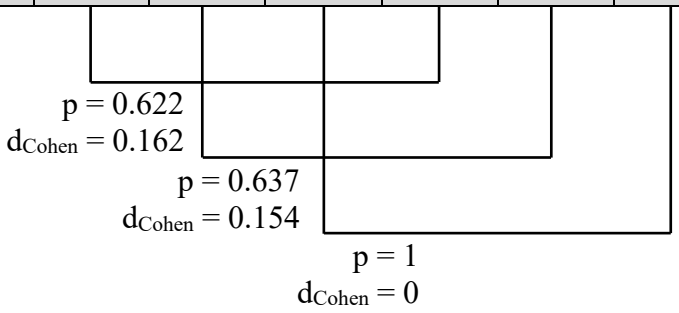


Table 26: AGG calculations of the AL proteins (LC fragments) and the corresponding GL. Mean AGG: sum of AGG values of all fragments from one molecular mass divided by the number of fragments. Mean AGG per case: mean of all mean AGGs per case. The P value of the two-tailed t-test and Cohen's d are shown below the table.

Case	possible assignable fibril protein				AGG (GL)				mean AGG (GL)	mean AGG (GL) per case	AGG (LC fragments)				mean AGG (LC frag.)	mean AGG (LC frag.) per case	ΔAGG (LC–GL)
FOR138	1–123	3–125	4–126	6–128	0.5	0.5	0.5	0.5	0.5	0.5	0.7	0.7	0.7	0.7	0.7	0.7	0.2
	3–104	4–105			0.3	0.3			0.3		0.8	0.8			0.8		
	3–121				0.6				0.6		0.7				0.7		
	3–123				0.5				0.5		0.7				0.7		
FOR102	3–125	4–126	6–128		0.4	0.4	0.4	0.4	0.4	0.4	2.0	2.0	2.0	2.0	2.0	2.0	1.6
FOR107	2–124	3–125	4–126	6–128	0.4	0.4	0.4	0.4	0.4	0.4	0.5	0.5	0.5	0.5	0.5	0.5	0.1
	4–125				0.4				0.4		0.5				0.5		
	8–123	9–124	10–125		0.4	0.4	0.4		0.4		0.6	0.6	0.6		0.6		
FOR128	1–125				0.3				0.3	0.3	1.3				1.3	1.3	1.0
	2–124	3–125	4–126	6–128	0.3	0.3	0.3	0.3	0.3		1.3	1.3	1.3	1.3	1.3		
FOR137	4–121				0.4				0.4	0.4	4.4				4.4	4.7	4.3
	4–125				0.4				0.4		4.3				4.3		
	16–121				0.4				0.4		5.0				5.0		
	17–121				0.4				0.4		5.0				5.0		
FOR142	2–124	3–125	4–126	6–128	0.4	0.4	0.4	0.4	0.4	0.4	0.2	0.2	0.2	0.2	0.2	0.2	-0.2
FOR147	2–124	3–125	4–126	6–128	0.4	0.4	0.4	0.4	0.4		0.3	0.3	0.3	0.3	0.3	0.3	-0.1
	4–125				0.4				0.4		0.3				0.3		
	8–123	9–124	10–125		0.4	0.4	0.4		0.4		0.3	0.3	0.3		0.3		
FOR101	4–123				4.5				4.5	4.5	2.2				2.2	2.2	-2.3
	4–125				4.4				4.4		2.2				2.2		

FOR124	4–103			1.5			1.5	1.5	1.9			1.9	1.9	0.4
	4–104			1.5			1.5		1.9			1.9		
	4–105			1.5			1.5		1.9			1.9		
FOR157	4–123	5–124	6–125	4.5	4.5	4.5	4.5	4.5	5.4	5.4	5.4	5.4	5.4	0.9
	4–125			4.4			4.4		5.3			5.3		
FOR159	1–125			4.3			4.3	4.3	1.6			1.6	1.6	-2.7
	2–125			4.3			4.3		1.6			1.6		
	2–124	3–125		4.3	4.3		4.3		1.6	1.6		1.6		
	4–125			4.4			4.4		1.7			1.7		
FOR009	1–124	2–125		6.7	6.7		6.7	6.7	6.3	6.3		6.3	6.3	-0.4
	2–127			6.6			6.6		6.2			6.2		
FOR012	4–125			5.7			5.7	5.7	5.9			5.9	5.9	0.2
FOR161	6–117	12–123		6.3	6.4		6.4	6.4	4.7	6.2		5.5	5.8	-0.6
	12–125	13–126		6.3	6.2		6.3		6.1	6.0		6.1		
FOR010	4–121			4.9			4.9	4.9	5.8			5.8	5.7	0.8
	4–125			4.8			4.8		5.6			5.6		
	8–125			5.0			5.0		5.8			5.8		
FOR103	1–124	2–125	7–130	4.7	4.7	5.2	4.9	4.9	5.5	5.5	6.1	5.7	5.7	0.8
FOR140	1–125			0.5			0.5	0.5	4.1			4.1	4.1	3.6
FOR1001	1–105			1.3			1.3	1.3	5.4			5.4	5.4	4.1
FOR1007	1–103			1.3			1.3	1.3	1.2			1.2	1.2	-0.1
	1–105			1.3			1.3		1.2			1.2		
	1–115			1.2			1.2		1.1			1.1		
Mean ± SD								2.6 ± 2.4					3.2 ± 2.3	0.6 ± 1.8

$p = 0{,}428$

$d_{Cohen} = 0{,}260$

9. Statutory declaration of independence

I hereby declare, in accordance with the Doctoral Regulations of the Faculty of Natural Sciences dated June 24, 2021, that this thesis with the title “Primary structural variations in systemic AL amyloidosis” is my work and written by myself. I only used the material and sources cited in this thesis.

I also hereby declare that I performed all scientific experiments according to the principles of good scientific practice, in accordance with the current “Satzung der Universität Ulm zur Sicherung guter wissenschaftlicher Praxis” [Rules of the University of Ulm for Assuring Good Scientific Practice].

I am familiar with the current Doctoral Regulations of Ulm University for the Faculty of Natural Sciences.

No one has received from me, directly or indirectly, any pecuniary benefits for work concerning the present dissertation.

This thesis has not been submitted in the same or similar form to any other examination authority or university.

Ulm, 18.01.2023 Julian Baur

10. Curriculum vitae

Julian Baur

Master of Science

Date of birth: [redacted] 1995
Marital status: [redacted]
Nationality: German
Contact: [redacted]
[redacted]
Tel.: [redacted]
E-Mail [redacted]

Education

2019 – 2023 PhD thesis: "Primary structural variations in systemic AL amyloidosis" (degree sought: Dr. rer. nat.)
University of Ulm

2017 – 2019 Master of science in Biochemistry (final grade: 1.3)
University of Ulm

2014 – 2017 Bachelor of science in Biology (final grade: 1.5)
University of Ulm

2006 – 2014 "Allgemeine Hochschulreife" [general matriculation standard] (final grade 2.2)
Schubart-Gymnasium Ulm

2002 – 2006 Elementary School
Meinloh-Grundschule Ulm-Söflingen

Research experience

09. 2019 – 03.2023 Primary structural variations in systemic AL amyloidosis
- ✓ Analysis of proteins using mass spectrometry
- ✓ Purification of proteins using the ÄKTA FPLC system

- ✓ Transmission electron microscopy
- ✓ Extraction of amyloid fibrils from human tissue
- ✓ Characterization of amyloid deposits using SDS-PAGE, Western Blot, Fluorescence spectroscopy
- ✓ Cryo-electron microscopy to reconstruct the three-dimensional structure of amyloid fibrils
- ✓ THT fibril kinetics

03.2019 – 08.2019

Investigation of the septin assembly into higher ordered structures in *S. cerevisiae* by biochemical methods and *in vivo* approaches

- ✓ Molecular cloning and site-directed mutagenesis for recombinant expression of proteins
- ✓ Protein expression in *E. coli*
- ✓ Purification of proteins using the ÄKTA FPLC system
- ✓ Characterization of proteins using SDS-PAGE
- ✓ Fluorescence microscopy with yeast cells

05.2017 – 08.2017

Biochemische Charakterisierung der Nukleotidbindung von Cdc11 [Biochemical characterization of the nucleotide-binding properties of Cdc11]

- ✓ Molecular cloning and site-directed mutagenesis for recombinant expression of proteins
- ✓ Generation of sequencing libraries for next-generation sequencing
- ✓ Protein expression in *E. coli*
- ✓ Purification of proteins using the ÄKTA FPLC system
- ✓ Working with radioisotopes

Other skills

German (native)

English (Excellent, both on speech and writing)

Computer skills: MS Word, MS Excel, MS PowerPoint, Adobe Illustrator, Adobe Photoshop, Peaks AB, Peaks X, Origin, UCSF Chimera, RELION

Publications

First author

Baur J, Berghaus N, Schreiner S, Hegenbart U, Schönland SO, Wiese S, Huhn S, Haupt C. Identification of AL proteins from 10 λ-AL amyloidosis patients by mass spectrometry extracted from abdominal fat and heart tissue. Amyloid. 2022 Jul 6:1-11. doi: 10.1080/13506129.2022.2095618. Epub ahead of print. PMID: 35792725.

First author

Baur JD, Rösler R, Wiese S, Johnsson N, Gronemeyer T. Dissecting the nucleotide binding properties of the septins from S. cerevisiae. Cytoskeleton (Hoboken). 2019 Jan;76(1):45-54. doi: 10.1002/cm.21484. Epub 2018 Nov 15. PMID: 30091242.

Second author

Banerjee S, **Baur J**, Daniel C, Pfeiffer PB, Hitzenberger M, Kuhn L, Wiese S, Bijzet J, Haupt C, Amann KU, Zacharias M, Hazenberg BPC, Westermark GT, Schmidt M, Fändrich M. Amyloid fibril structure from the vascular variant of systemic AA amyloidosis. Nat Commun. 2022 Nov 25;13(1):7261. doi: 10.1038/s41467-022-34636-4. PMID: 36433936.

Second author

Radamaker L, **Baur J**, Huhn S, Haupt C, Hegenbart U, Schönland S, Bansal A, Schmidt M, Fändrich M. Cryo-EM reveals structural breaks in a patient-derived amyloid fibril from systemic AL amyloidosis. Nat Commun. 2021 Feb 8;12(1):875. doi: 10.1038/s41467-021-21126-2. PMID: 33558536; PMCID: PMC7870857.

Third author

Anker S, Hinderhofer K, **Baur J**, Haupt C, Röcken C, Beimler J, Zeier M, Weiler M, Wühl E, Kimmich C, Schönland S, Hegenbart U. Lysozyme amyloidosis-a report on a large German cohort and the characterisation of a novel amyloidogenic lysozyme gene variant. Amyloid. 2022 May 9:1-10. doi: 10.1080/13506129.2022.2072198. Epub ahead of print. PMID: 35533055.

Third author

Steinebrei M, Gottwald J, **Baur J**, Röcken C, Hegenbart U, Schönland S, Schmidt M. Cryo-EM structure of an ATTRwt amyloid fibril from systemic non-hereditary transthyretin amyloidosis. Nat Commun. 2022 Oct 27;13(1):6398. doi: 10.1038/s41467-022-33591-4. PMID: 36302762; PMCID: PMC9613903.

Co-author	Radamaker L, Karimi-Farsijani S, Andreotti G, **Baur J**, Neumann M, Schreiner S, Berghaus N, Motika R, Haupt C, Walther P, Schmidt V, Huhn S, Hegenbart U, Schönland SO, Wiese S, Read C, Schmidt M, Fändrich M. Role of mutations and post-translational modifications in systemic AL amyloidosis studied by cryo-EM. Nat Commun. 2021 Nov 5;12(1):6434. doi: 10.1038/s41467-021-26553-9. PMID: 34741031; PMCID: PMC8571268.
Co-author	Heerde T, Rennegarbe M, Biedermann A, Savran D, Pfeiffer PB, Hitzenberger M, **Baur J**, Puscalau-Girtu I, Zacharias M, Schwierz N, Haupt C, Schmidt M, Fändrich M. Cryo-EM demonstrates the in vitro proliferation of an ex vivo amyloid fibril morphology by seeding. Nat Commun. 2022 Jan 10;13(1):85. doi: 10.1038/s41467-021-27688-5. PMID: 35013242; PMCID: PMC8748726.