TECHNISCHE UNIVERSITÄT MÜNCHEN

Department Chemie

Lehrstuhl für Biotechnologie

Principles of antibody domain integrity

Benedikt Weber

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigten Dissertation

Vorsitzender:	Prof. Dr. Matthias Feige
Prüfer der Dissertation:	1. Prof. Dr. Johannes Buchner
	2. Prof. Dr. Bernd Reif
	3. Prof. Dr. Roberto Sitia

Die Dissertation wurde am 18.09.2018 bei der Technischen Universität München eingereicht und durch die Chemie am 19.11.2018 angenommen.

Content

Summary

Antibodies and their fragments, e.g. Fab, Fv and single domains, are widely used as diagnostics and therapeutics. However, the principles governing antibody domain integrity are still not understood in sufficient detail. In this context, various factors causing amyloid fibril formation of antibody fragments, particularly the light chain (LC) and its N-terminal variable domain (V_L) still remain unclear. Thus, the underlying molecular basis of antibody stability is of major importance. This thesis describes the investigation of two different IgG-derived antibody fragments, the LC and the C_H2 domain, with a focus on their conformational stability and structural integrity.

In AL amyloidosis, the most common systemic amyloidosis, the full-length LC and its N-terminal fragments undergo conformational transitions from the natively folded state to amyloid fibrils. These fibrils are distributed by the blood stream and deposited as amyloid plaques in organs, causing toxic effects ultimately leading to organ failure and death. Several amyloidosis-associated mutations within the V_L domain have been reported in the literature. Here, the goal was to determine the impact of the LC linker and the presence of the C_L domain on the structural integrity of V_L in the context of the full-length LC. It was found that the C_L domain renders certain LC variants resistant against fibril formation. Moreover, the LC linker is of key importance to regulate the relative orientation of the two constituent LC domains to each other. Particularly, the linker residue R108 maintains important salt bridges within the antibody LC. However, some LC variants are shown to be still able to form amyloid fibrils under physiological conditions *in vitro*. Interestingly, for LC I2E and LC R61A the kinetics for the full-length LC variants were even accelerated compared to their corresponding V_L variants. Most likely this is due to an altered interaction network within the V_L domain and non-native interactions between V_L and C_L.

The C_H2 domain has been reported to be untypically instable compared to other immunoglobulin domains. Together with the C_H3 domain, it is responsible for receptor binding and effector functions such as triggering immune response. To test the impact of C-terminal residues on the conformational stability of the C_H2 domain, it was systematically extended by naturally occurring residues connecting it to C_H3. Interestingly, the presence of the C-terminal Lys101 increases the conformational stability of the domain by ~ 14 °C. This is achieved by the formation of two important secondary structure elements, a very C-terminal β-strand and an

adjacent α-helix, which are connected by a dipole interaction between Lys101 and the α-helix. Both, the important structure elements and the dipole interaction protect the hydrophobic core leading to a substantial overall increase of conformational stability. As the C-terminal lysine and its interactions are a conserved features of constant antibody domains, it is proposed to reconsider antibody domain boundaries based on biophysical and structural parameters rather than genetic organization.

Zusammenfassung

Antikörper und deren Fragmente, z. B. Fab, Fv und Einzeldomänen, werden häufig als Diagnostika und Therapeutika verwendet. Allerdings sind die exakten Grundlagen der Antikörperdomänenintegrität noch nicht zufriedenstellend aufgeklärt. In diesem Zusammenhang gibt es ungeklärte Faktoren, die zur Amyloidfibrillenbildung von Antikörperfragmenten, vor allem von der leichte Kette (LC) und ihrer N-terminalen variablen Domäne (V_L), beitragen. Folglich ist das Verständnis der molekularen Basis der Antikörperstabilität von sehr hoher Bedeutung. Diese Arbeit stellt die Untersuchung zweier unterschiedlicher IgG Antikörperfragmente, LC und C_H2 Domäne, mit Fokus auf die konformationelle Stabilität und die strukturelle Integrität, dar.

Bei AL Amyloidose, der meistverbreiteten systemischen Amyloidoseerkrankung, kommt es bei der Volllänge-LC und ihren N-terminalen Fragmenten zu einem Übergang vom nativ gefalteten Zustand zu amyloiden Fibrillen. Diese Fibrillen werden über den Blutstrom im Organismus verteilt und als amyloide Plaques in Organen abgelagert, was zu toxischen Effekten und schließlich zu Organversagen und Tod führt. In der Literatur gibt es Hinweise auf einige AL Amyloidose assoziierte Mutationen innerhalb der V_L Domäne. In dieser Arbeit lag der Fokus auf der Bestimmung des Einflusses des „LC Linkers“ und der C_L Domäne auf die strukturelle Integrität der V_L Domäne im Kontext der Volllänge-LC. Die Ergebnisse zeigen, dass die Gegenwart der C_L Domäne bestimmte LC Varianten resistent gegen Fibrillenbildung macht. Zudem hat der „LC Linker“ eine Schlüsselfunktion in der Regulierung der relativen Orientierung der beiden LC Domänen, V_L und C_L, zueinander. Besonders der „LC Linker“-Rest R108 bildet dabei wichtige Salzbrücken innerhalb der LC. Im Gegensatz dazu gibt es LC Varianten, die weiterhin *in vitro* unter physiologischen Bedingungen Fibrillen bilden. Zudem wurde gezeigt, dass die beiden Varianten LC I2E und LC R61A beschleunigte Fibrillenkinetiken aufweisen, als deren jeweilige V_L Varianten. Dieser Effekt beruht höchstwahrscheinlich auf Änderungen im Interaktionsnetzwerk innerhalb der V_L Domäne, sowie nicht-nativen Interaktionen zwischen V_L und C_L.

Die C_H2 Domäne ist im Vergleich zu anderen Antikörperdomänen bemerkenswert instabil. Zusammen mit der C_H3 Domäne ist sie verantwortlich für Rezeptorbindung und Effektorfunktionen wie das Auslösen einer Immunreaktion. Um den Einfluss von C-terminalen Resten der C_H2 Domäne zu untersuchen, wurde diese systematisch mit natürlich

vorkommenden Resten verlängert. Interessanterweise erhöht das C-terminale Lys101 die strukturelle Stabilität der C_H2 Domäne um ~ 14 °C. Das wird erreicht durch die Bildung zweier wichtiger Sekundärstrukturelemente, einem C-terminalen β-Faltblatt, sowie einer benachbarten α-Helix, ausgelöst in Gegenwart von Lys101. Beide Sekundärstrukturelemente sind verbunden durch Dipolinteraktionen. Die Sekundärstrukturelemente, sowie die Dipolinteraktion schützen den hydrophoben Proteinkern und führen daher zu einem starken Stabilitätsgewinn. Da das C-terminale Lysin und seine Interaktionen in konstanten Antikörperdomänen konserviert sind, sollten Antikörperdomänengrenzen biophysikalisch und strukturell statt genetisch definiert werden.

1. Introduction

1.1. Protein folding

Proteins are highly abundant polypeptides, which accomplish a wide variety of tasks in living organisms, e.g. cell differentiation, immunity, signaling, catalysis and countless more functions. Naturally occurring polypeptides are assembled of 20 proteinogenic amino acids by establishing covalent peptide bonds between the single residues. Genes, which are the RNA and thereby protein encoding parts of the DNA, contain all the information necessary to initiate transcription and translation of the primary protein structure, i.e. the amino acid sequence. After or already during protein biosynthesis by the ribosomes, the nascent linear polypeptide chain of most proteins rapidly adopts the native conformation (Gething and Sambrook, 1992; Kleizen and Braakman, 2004). Protein folding is of major importance since the function of a given protein is a direct consequence of its unique three-dimensional structure (Orengo et al., 1999). In contrast, intrinsically disordered proteins (IDP) like the tau protein and α-synuclein, comprise long inherently unfolded stretches (Luo et al., 2014; Uversky, 2003; Wright and Dyson, 1999). Furthermore, some proteins are prone to misfolding or aggregation, especially in the highly crowded cell environment, which is described in more detail in chapter 1.3. (van den Berg et al., 1999; Munishkina et al., 2004). Protein folding is generally dictated by intramolecular non-covalent interactions, i.e. hydrogen bonds, salt bridges and hydrophobic interactions, which are mainly van der Waals interactions (Dill, 1990; Dobson, 2003). The latter ones are mainly driven by the aqueous surrounding of a protein leading to polar interactions between amino acids and water molecules and intramolecular hydrophobic interactions between non-polar residues (Pace et al., 2011). Thus, hydrophobic patches form the core of a protein with little surface exposure. This leads to an energetically favored state, particularly by increasing the entropy of the water molecules (Privalov and Makhatadze, 1993a). Consequently, folding guides the protein to adopt the most thermodynamically stable conformation called the native state (Makhatadze and Privalov, 1993a). A combination of several structural elements such as α-helices, β-strands, loops and randomly folded coil make up the secondary structure of a protein (Kabsch and Sander, 1983). Intramolecular hydrogen bonds between the polypeptide backbone amides and carbonyls as well as interactions involving the amino acid side chains drive secondary structure formation. The ensemble of all secondary structure elements of a protein is called the tertiary structure, which describes the three-dimensional conformation (Richardson, 1981). A complex of two or

more individual polypeptides make up a proteins quaternary structure. In general, protein folding from an unfolded to the native state represents a first order reaction as the reaction rate is dependent on the protein concentration.

In vivo, protein folding is actively managed by many regulatory elements such as molecular chaperones, enzymes like peptidylprolyl *cis-trans* isomerases (PPIase) and protein disulfide isomerases (PDI), efficient intracellular protein transport and many more (Gething and Sambrook, 1992). However, a variety of proteins is able to adopt the native fold *in vitro* without assistance of the cell (Dill and MacCallum, 2012; Rudolph and Lilie, 1996). Thus, it was suggested that the amino acid sequence of the polypeptide chain represents the entire information needed for protein to fold correctly into the native state (Anfinsen, 1973). As amino acids in a polypeptide chain can randomly adopt numerous relative orientations to each other, a protein composed of 101 amino acids can theoretically explore 5×10^{47} conformations. Consequently, following a trial and error approach to find the native state would take 10^{27} years considering a conformation exploring rate of 10^{13} per second (Levinthal, 1969; Zwanzig et al., 1992). In contrast, the majority of the proteins adopt the stable native three-dimensional structure within milliseconds to a few minutes. Therefore, Levinthal's paradox describes that protein folding is a multi-step process with defined conformational constraints leading and underlying pathways, which allow to efficiently adopt an energetically favored and therefore stable native conformation (Dill and Chan, 1997a; Levinthal, 1968). However, the mechanism of how a protein natively folds depending on its primary structure still remains largely unknown. Based on the increasing availability of structural data correlated with the amino acid sequence, applications like secondary structure prediction and homology modelling have been created, which are dependent on the underlying dataset (Drozdetskiy et al., 2015; McGuffin et al., 2000; Pirovano and Heringa, 2010). Several protein folding models have been proposed. Thus, the applicability of each model depends on the protein type or class. The hydrophobic collapse model describes protein folding initiated by a so-called hydrophobic collapse as a result of repulsive interactions between patches of hydrophobic residues and the surrounding water molecules (Pace et al., 2011). This drives the formation of a hydrophobic protein core and subsequently the interactions of surface-exposed polar side chains with the aqueous buffer. Water increases its entropy while the free energy of the protein folding state is minimized, which is supported by the favorable polar interactions resulting in a low total energy state for the natively folded protein. The diffusion collision

model explains protein folding initialized by partially folded fluctuating short secondary structure elements, so called microdomains, which rapidly adopt various conformations (Karplus and Weaver, 1994). Finally, a distinct tertiary structure is formed, implemented by numerous microdomain-microdomain interactions and other possibly rate-limiting events like enzymatic isomerization. A third proposed model is the nucleation condensation theory (Fersht, 1995, 1997). In this concept, a folding nucleus initiates parallel folding of secondary and tertiary structure indicating a supporting character of the tertiary structure to maintain the secondary structure and vice versa.

The underlying goal of the proposed protein folding models is to accomplish minimal Gibbs free energy, reflecting a favorable and stable conformational state (Dobson, 2003). The energy landscape of the folding funnel is a model for protein folding incorporating enthalpy and entropy as the main folding-driving forces (Figure 1) (Bartlett and Radford, 2009a; Dill and Chan, 1997b; Neira and Fersht, 1999). The model describes the folding process for proteins adopting a distinct conformation starting from the unfolded to the native state, which often occurs via one or more intermediate states. A two-state folding pathway without intermediates represents the simplest way of protein folding (Plaxco et al., 2000). The surface of the funnel depicts the lowest energy state for a given protein with fixed conformational constraints and entropy. The model shows that a protein can reach the native state on several pathways as indicated by the two dimensions. However, some pathways exhibit local energy minima leading to intermediate states, which are kinetically trapped. Depending on the conditions, which modulate entropy and free energy, protein folding is either stopped or paused at an intermediate state. In general, intermediates can be on- or off-pathway either leading to the native state or partially folded intermediates, aggregates or amyloid fibrils, respectively.

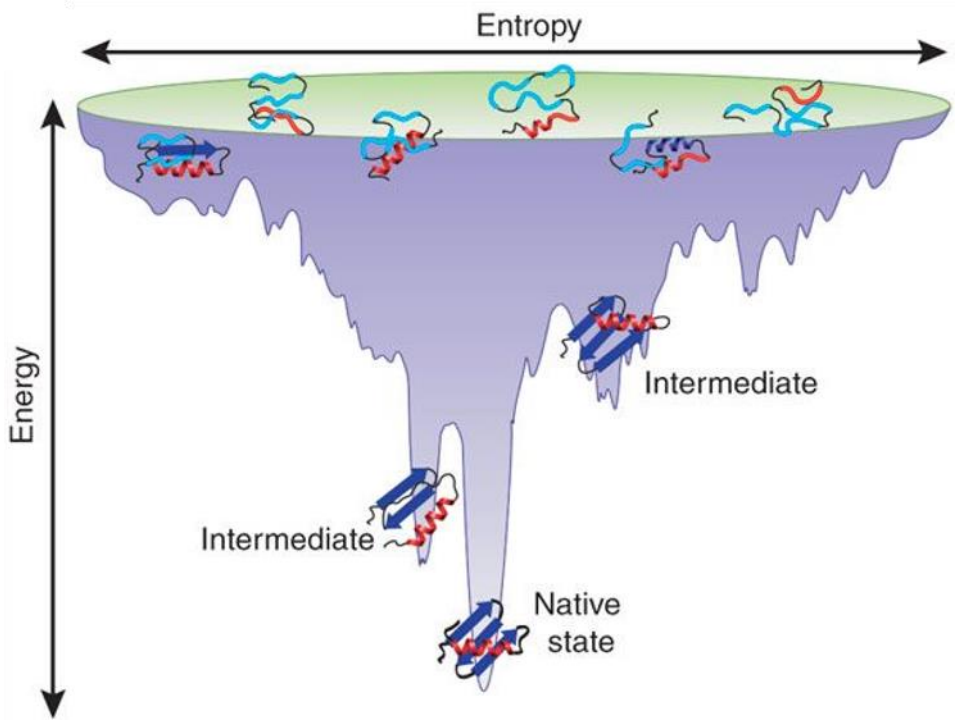

***Figure 1. Folding funnel energy landscape**. Unfolded conformations occur in the opening of the funnel characterized by high, unfavorable energy states. Fewer conformations exist at lower energy states represented by a narrower funnel. Proteins located in hollows represent intermediate folding states trapped in local energy minima. Proteins, which successfully adopted the native fold, adopt a unique low energy state. Adapted with permission from Bartlett and Radford, 2009.*

Some intermediate states have been described in detail, e.g. the alternatively folded state (AFS) and the molten globule state (Buchner et al., 1991; Kuwajima, 1989; Lilie and Buchner, 1995; Ohgushi and Wada, 1983). The AFS is adopted at low pH and exhibits a substantial secondary structure content, which is markedly different from the native state as shown for individual immunoglobulin (Ig) domains. Cytochrome c and other proteins can adopt the molten globule state, which is characterized by a maintained native-like secondary structure, but a looser tertiary structure compared to the native fold (Goto and Nishikiori, 1991). The state transition to a molten globule conformation is driven by a low pH as well as high salt concentrations. In contrast to the AFS, proteins in the molten globule state exhibit a lower conformational stability than their natively folded counterparts. Amyloid fibrils represent another important folding state, which has significant impact in various diseases as described in chapter 1.3.1.

1.2. Protein stability, unfolding and aggregation

Proteins are only marginally stable, which is needed to successfully perform their functions and maintain flexibility and protein dynamics. In this context, it has been shown that there is a trade-off between protein stability and function such as catalytic activity (Shoichet et al., 1995; Tokuriki et al., 2008). As indicated by the folding funnel, protein folding is a complex chemical reaction, based on equilibria between two or more distinct states (Dobson, 2003; Englander and Mayne, 2014). Ideally, the structural equilibrium of a protein shifts towards the native conformation during protein folding. However, depending on the protein and the conditions (free energy and entropy), the equilibrium can be highly sensitive and shift towards unfolded or intermediate states. This process is called (partial) unfolding of a protein. Thus, the native conformation is for many proteins a weighted equilibrium between at least two states (Baldwin, 1995; Bryngelson et al., 1995; Lattman and Rose, 1993). In general, small one domain proteins share the intrinsic ability to reversibly fold and unfold. To unfold a protein from its native state, energy needs to be applied to the system, represented by an energy barrier as observed for exothermic chemical reactions (Makhatadze and Privalov, 1993b; Privalov and Makhatadze, 1993b). This barrier varies from protein to protein depending on its stability, which is energetically described by the Gibbs free energy changes (Equation 1):

$$\Delta G = \Delta H - T\Delta S$$

Equation 1. Gibbs free energy*. G: Gibbs energy; H: enthalpy; T: absolute temperature; S: entropy.*

In protein folding, the entropy increase is driven by hydrophobic interactions, which result from a higher order of the surrounding water molecules upon protein folding. Enthalpy depends on the solvation of the protein, which is represented by polar interactions. The larger ΔG is between the native and the unfolded states, the more stable is the native fold of the protein. In general, there are two ways to induce protein unfolding (i) applying external energy or (ii) modulate the conditions, which determine the ΔG between the native and the unfolded states. External energy can be applied by increasing the temperature (Myers et al., 1995; Xia et al., 2013). ΔG can be minimized by changing the pH or adding a chemical denaturant like urea or guanidinium chloride (GdmCl) (Makhatadze and Privalov, 1992; Wetlaufer et al., 1964). Unfolding studies are routinely performed to assess the stability of a protein, which is of major importance for understanding the folding, unfolding and particularly misfolding of a given protein.

Protein stability is significantly linked to a high amount of favorable, stabilizing interactions making up secondary and tertiary structure elements (Goto and Hamaguchi, 1979; Pace et al., 2011; Politi and Harries, 2010). Among these are covalent and transient non-covalent interactions like disulfide bonds and electrostatic bonds (hydrogen bonds, ionic bonds), van der Waals forces and hydrophobic effects, respectively. Disulfide bonds emerge from the oxidization of two cysteine residues. This covalent link is known to significantly stabilize proteins, especially if found within the hydrophobic core, e.g. highly conserved for Ig domains (Feige et al., 2010). In general, a favorable distance between two cysteine residues in the polypeptide chain is important for the stabilizing effect (Pornillos et al., 2010; Shimaoka et al., 2002). Electrostatic bonds majorly occur as ionic interactions between charged amino acid side chains or as hydrogen bonds, which is a polar interaction between a hydrogen and a more electronegative atom like an oxygen of polypeptide backbone carbonyl (Vogt et al., 1997). Interestingly, the stabilizing effect of electrostatic forces is strongly protein-dependent, i.e. the unique polypeptide chain, fold and the given conditions (Joh et al., 2008; Takano et al., 2003). In fact, it was shown that hydrogen bonds can even destabilize the native fold (Campos et al., 2005). Van der Waals forces and the hydrophobic interactions largely maintain the hydrophobic core of a protein (Pace et al., 2011; Tych et al., 2016). However, a generally applicable theory of the impact of different interaction types on the overall protein stability is still a subject of research.

Under unfavorable conditions, e.g. destabilizing buffer, stress, extreme concentrations, non-native interactions or mutations, the mentioned interactions are the driving forces towards protein aggregation (Chi et al., 2003; Fink, 1998). Protein aggregation is typically a second order reaction as it is dependent on the interaction between the protein molecules. This occurs frequently during protein folding starting from an (partially) unfolded state. These aggregates represent off-pathway intermediates, of which some have a high intrinsic tendency to from aggregates. Besides, aggregation can take place during storage of natively folded proteins likely due to an unfavorable storage buffer or other destabilizing conditions. These induce a certain proportion of partially unfolded content, which is prone to aggregation. This phenomenon is especially important for the shelf life of biopharmaceuticals since aggregates result in immune reactions and the available amount of the active compound is decreased (Wang, 2005). The aggregation pathway is like folding and unfolding an equilibrium process, which can be manipulated by adjusting the conditions such as temperature, buffer,

UV radiation or concentration. However, in a living organism, many other factors are correlated with aggregation. Specific mutations can alter the stability and aggregation tendency of a protein. Moreover, failure in complex mechanisms like ER protein quality control, the ubiquitin-proteasome system or aberrant protein biosynthesis might lead to an accumulation of aggregated proteins in the cell (Ellgaard and Helenius, 2003; McNaught et al., 2001; Tyedmers et al., 2010).

1.3. Protein misfolding disorders and amyloidosis

Misfolding of proteins in the organism is the underlying reason of a variety of different diseases called protein misfolding disorders (Vendruscolo et al., 2011). Depending on the localization of the emerging aggregates or amyloids the diseases can be categorized in at least three distinct groups (i) neurodegenerative disorders, (ii) non-neuropathic systemic amyloidosis and (iii) non-neuropathic local amyloidosis (Knowles et al., 2014). An overview of the most common disorders of these groups is given in Table 1. Among these are prominent and wide-spread diseases like Alzheimer's disease, Parkinson's disease and type II diabetes, which have significant impact on today's social lives, healthcare systems and economies. Interestingly, most amyloidosis-associated precursor proteins are only susceptible for amyloid formation under specific conditions such as mutations, fragmentation, unfavorable environment, overexpression, stress etc. (Chiti et al., 1999; Fink, 1998). Therefore, intense research efforts are undergone during the last decades to improve early diagnosis and find curative treatments. However, the molecular mechanisms of most protein misfolding diseases remain unclear.

In vivo, protein folding is guided by many factors such as chaperones (Hartl and Hayer-Hartl, 2002). In case of failure, misfolded proteins are directed to the proteasome for degradation (Figure 2) (Bence et al., 2001; Glickman and Ciechanover, 2002). Thus, in a well-functioning cell, misfolded proteins are disarmed before any damage is caused. Generally, amyloid precursor proteins experience a transition from a (partially) unfolded state or an intrinsically disordered state to amyloid fibrils. The amyloid state is represented by generic precursor-independent, stable, insoluble, highly ordered and β-strand-rich fibrillar protein assemblies, which are primarily extracellularly deposited (Chiti and Dobson, 2017) (cf. chapter 1.3.1.). Despite numerous studies on the native, misfolded, disordered and amyloid states, the exact

molecular reasons for entering the amyloid folding pathway are enigmatic, particularly regarding the structural diversity of the precursor proteins (Table 1). Recent studies suggest that soluble prefibrillar species cause cell toxicity rather than the mature amyloid fibrils (Demuro et al., 2005; Ferreira et al., 2007).

Disorder	Precursor protein or peptide	Affected organs	Structure of precursor
Neurodegenerative disorders			
Alzheimer's disease	Amyloid-beta (Aβ) peptide	Brain	Intrinsically disordered
Amyotrophic lateral sclerosis	Superoxide dismutase 1	Brain	β-sheet and Ig-like
Familial amyloidotic polyneuropathy	Transthyretin mutants	Brain	β-sheet
Huntington's disease	Huntingtin fragments	Brain	Mostly intrinsically disordered
Parkinson's disease	α-Synuclein	Brain	Intrinsically disordered
Non-neuropathic systemic amyloidosis			
Amyloid A amyloidosis	Serum amyloid A1 protein fragments	Spleen, kidney, adrenal glands, liver	α-helical
Antibody light chain (AL) amyloidosis	Immunoglobulin (Ig) light chains or fragments	Heart, kidney, liver, and more	β-sheet and Ig-like
Senile systemic amyloidosis	Transthyretin	Mainly heart	β-sheet
Non-neuropathic local amyloidosis			
Type II diabetes	Islet amyloid polypeptide (IAPP)	Pankreas	Intrinsically disordered
Cataract	γ-Crystallins	Eye	All-β, γ-crystallin like

Table 1. Human protein misfolding disorders. *Data summarized from Chiti and Dobson, 2017; Gillmore et al., 2001; Knowles et al., 2014; Lu et al., 2014 and Wechalekar et al., 2016.*

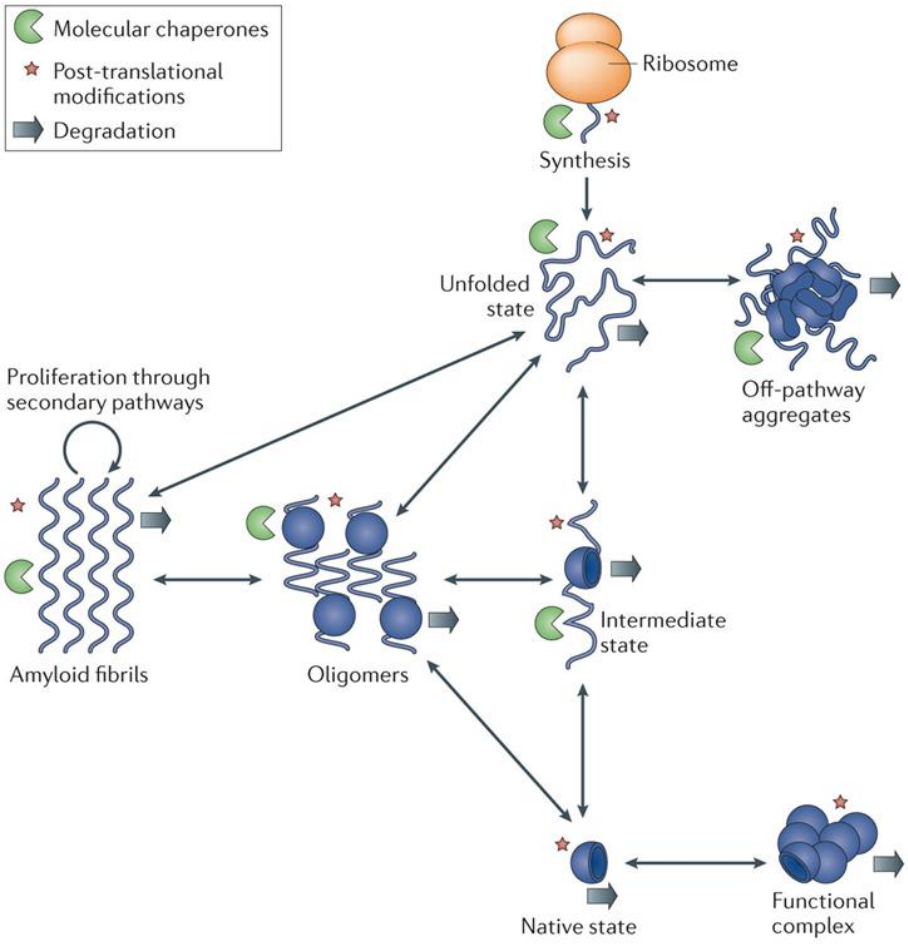

Figure 2. Scheme representing different conformational states of a protein including off-pathway aggregates and amyloid fibrils. *A variety of factors determines the equilibrium between the different states such as protein stability, synthesis, degradation, chaperone interaction and post translational modifications. In contrast to (partially) unfolded states, amyloid fibrils are highly ordered structures, which proliferate fast once a seed has formed by self-replication mechanisms. Interestingly, the molecular conformation of the amyloid state is generic, thus independent from the precursor protein. Adapted with permission from Knowles et al., 2014.*

1.3.1. Amyloid fibril formation and structure

Amyloid fibrils represent a distinct, generic, highly ordered β-strand-rich conformational state, which derives from structurally heterogeneous precursor proteins. It is characterized by a high thermodynamic and kinetic stability, insolubility and a common fibril architecture (Baldwin et al., 2011; Chiti and Dobson, 2017). Fibril formation typically initiates from unfolded or intermediate states, however in some cases also directly from the native state (Figure 2), which is for approximately half of the known amyloidogenic precursor proteins an intrinsically disordered conformation (Knowles et al., 2014). Amyloid fibril formation experiences three distinct kinetic phases, i.e. (i) the lag phase, (ii) the elongation phase and (iii) the plateau

phase, described by a sigmoidal kinetic curve (Figure 3A) (Collins et al., 2004; Morris et al., 2009). Three models fibril formation models are generally accepted, depending on the properties of the amyloidogenic precursor protein (Figure 3B) (Kumar and Udgaonkar, 2009). The most commonly applied model is the nucleated polymerization model (Chen et al., 2002). A folded or disordered protein undergoes a thermodynamically non-preferred transition to a structurally fibril-like polymerization nucleus, represented by the lag phase. Recruitment of protein monomers by the fibril nuclei lead to rapid fibril assembly (elongation phase), which is stalled (plateau phase) when no more monomers are available or the specific polymer-monomer equilibrium is reached. The nucleated conformational conversion model deviates from this by the conversion of early aggregates derived from the precursor proteins to growth-competent nuclei, representing the rate-limiting step (Lee et al., 2011). Both models require either IDPs or partially unfolded proteins prone to aggregation or nucleus formation, respectively. Compact, stable globular proteins are intrinsically more aggregation-resistant, thereby excluding a major precondition of the two described fibril formation models (Chiti and Dobson, 2009). However, it was shown that globular proteins are able to adopt native-like conformations under specific conditions, exposing a small proportion of usually buried aggregation-prone regions. This leads to native-like aggregation, followed by the assembly of polymerization nuclei, resulting in fibril formation. Once a fibril is formed, secondary processes such as secondary nucleation or fragmentation can occur (Chiti and Dobson, 2017). The lag phase of the fibril formation kinetic can be overcome by seeding with fragmented fibril material or polymerization nuclei (Jucker and Walker, 2011). Therefore, the rate-limiting steps are skipped and the elongation phase is immediately entered. Interestingly, studies exhibited that cross-seeding, i.e. fragments of mature amyloid fibrils recruit monomers of other precursor proteins, is a wide-spread phenomenon *in vivo* (Lundmark et al., 2005; Morales et al., 2013; Ono et al., 2012).

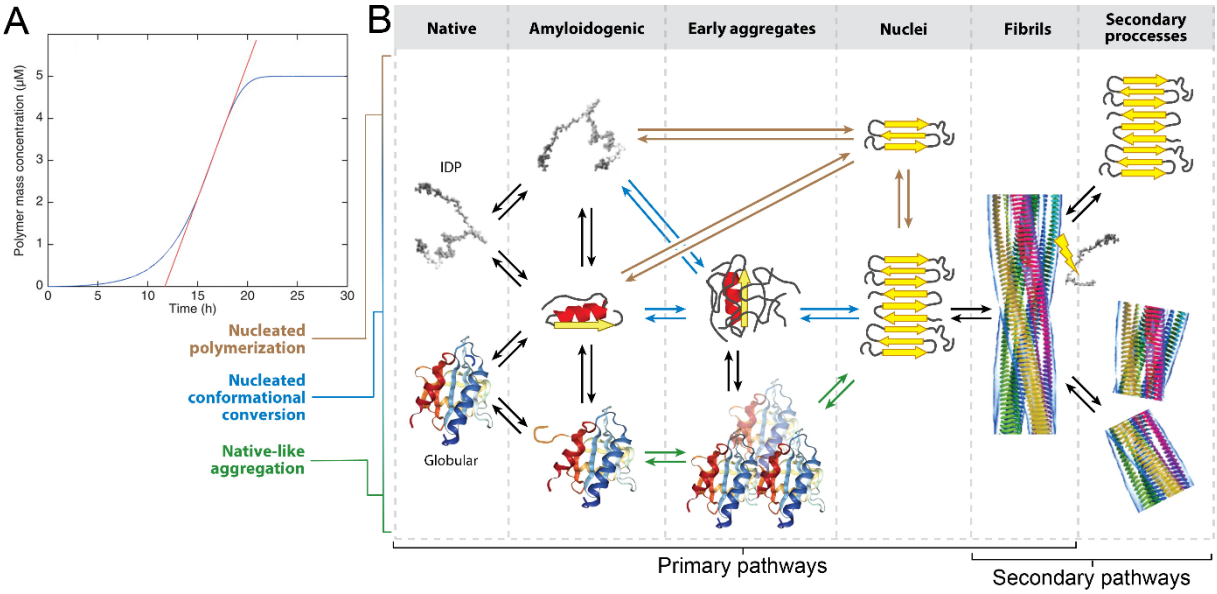

Figure 3. Amyloid fibril formation. A) *Characteristic sigmoidal fibril growth curve showing the three kinetic phases: lag phase, elongation phase and plateau phase. The red line depicts the maximal slope of the transition.* ***B)*** *Mechanisms of amyloid fibril formation including three different models: nucleated polymerization (brown), nucleated conformational conversion (blue) and native-like aggregation (green) starting with either globular or IDPs. Fibril formation processes are categorized in primary and secondary pathways. Adapted an modified with permission from Chiti and Dobson, 2017 and Knowles et al., 2009.*

Although amyloid fibrils reveal differing morphologies, the molecular structure is highly conserved among all amyloid filaments. Amyloid fibrils are composed of β-sheets assembled by hydrogen bonds to a characteristic cross-β fibril core running along the fibrils length (Figure 4) (Fitzpatrick et al., 2013, 2017; Gremer et al., 2017; Schmidt et al., 2016). The common cross-β-sheet core provides high stability and is maintained and elongated by polypeptide backbone hydrogen bonds. The specific shape of a β-sheet incorporated in the fibril core originates from side chain interactions and is thus dependent on the amino acid sequence of the precursor protein (Baldwin et al., 2011). Therefore, due to the differing β-sheets of the cross-β-sheet core, cross-seeding is assumed to be still amino acid sequence-dependent to some extent (Krebs et al., 2004).

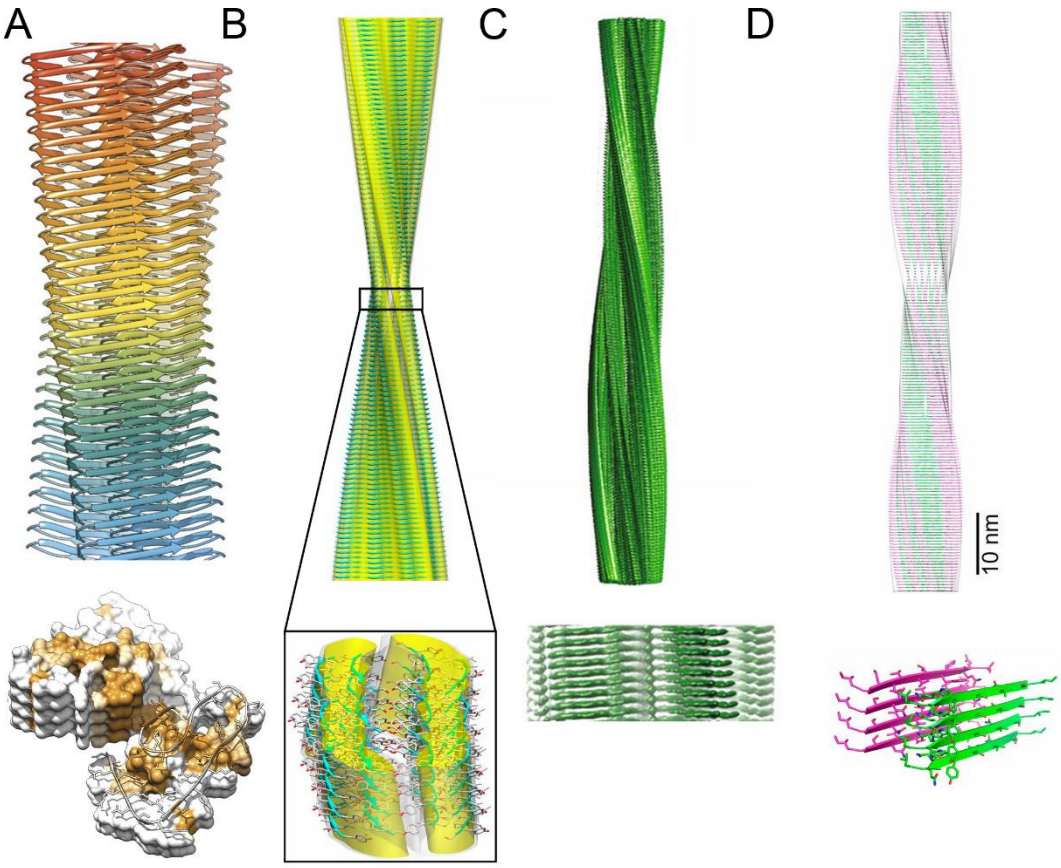

Figure 4. The generic amyloid fibril structure*. Despite diverse morphologies (upper panel), amyloid fibrils comprise a common characteristic molecular core structure (lower panel).* ***A)*** *Cryo-EM structure of amyloid-β(1-42).* ***B)*** *Combined MAS NMR and cryo-EM structure of an eleven-residue fragment of transthyretin (TTR).* ***C)*** *Cryo-EM structure of tau protein.* ***D)*** *Cryo-EM structure of an antibody λ light chain. Adapted and modified with permission from Fitzpatrick et al., 2013, 2017; Gremer et al., 2017 and Schmidt et al., 2016.*

Noteworthy, there is increasing evidence that specific functional amyloids undertake important biological tasks, particularly in bacteria. Bacterial fimbriae of some *E.coli* and *Salmonella* strains are composed of functional amyloids (Dueholm et al., 2010; Fowler et al., 2007). Moreover, some amyloid peptides like amyloid-β(1-42) exhibit antimicrobial properties even in the state of mature amyloid fibrils by targeting and dismantle the cell membrane (Last and Miranker, 2013; Soscia et al., 2010).

1.3.2. Systemic antibody light chain (AL) Amyloidosis

Antibody light chain (AL) amyloidosis is the most common non-neuropathic systemic amyloidosis with an incidence of three to fourteen cases per million human beings per year (Kyle et al., 1992; Pinney et al., 2013). Due to its systemic appearance, various organs are involved during the progress of the disease, e.g. heart, kidney, liver, most commonly leading to death by heart failure (Baden et al., 2009; Pepys, 2006). AL Amyloidosis is caused by a proliferating monoclonal plasma cell clone, which overexpresses high quantities of aberrant antibody light chains (LC). Interestingly, particularly the N-terminal fragment, i.e. the variable light chain (V_L) was reported to form amyloid fibrils. Unlike immunoglobulin production in healthy plasma cells, in AL amyloidosis LCs are secreted into the blood stream without being assembled with a heavy chain (HC). The disorder is typically secondary to a benign monoclonal gammopathy, frequently found as a comorbidity to multiple myeloma or other plasma cell dyscrasia (Merlini et al., 2011). Circulating LCs and their fragments form amyloid fibrils, which are extracellularly deposited as plaques, interfering with organ function. Recent data suggest that in amyloid disorders, low molecular weight aggregates of the precursor protein cause toxic effects after being incorporated into the cell (Kayed et al., 2003). A major drawback is often the late diagnosis due to unspecific symptoms, such as weight loss, fatigue, dyspnea (shortness of breath), edemata (accumulation of serous fluids) of extremities depending on which organs are involved (Grogan et al., 2017; Sanchorawala, 2006). Moreover, patients with a monoclonal gammopathy of undetermined significance (MGUS) diagnosis are typically solely monitored for the development of multiple myeloma or lymphoma since both diseases have a much higher incidence (Gertz, 2018). This interferes with a curative disease treatment due to already irreversible organ damage by the time of late diagnosis. Current treatments include chemotherapy using Melphalan (cytostatic which alkylates guanines, causing plasma cell death) together with Dexamethasone or Prednisone (both immunosuppressive glucocorticoids), Bortezomib (proteasome inhibitor which block metabolic pathways in cancer cells) plus Dexamethasone, Thalidomide (immunomodulatory drug inhibiting cancer cell proliferation and angiogenesis) plus Dexamethasone or Lenalidomide (similar to Thalidomide) plus Dexamethasone depending on the distinct development of the disease as well as autologous stem cell transplantation, which is similar to multiple myeloma treatment (Gertz, 2018). Daratumumab, a therapeutic monoclonal antibody directed against CD38, is currently subject of clinical trials in order to achieve approval by health authorities to be indicated for

AL amyloidosis patients (de Weers et al., 2011). The exact mechanisms and reasons why (i) the monoclonal plasma cells release aberrant LCs into the blood and why (ii) some LCs are dramatically more amyloidogenic than others as well as (iii) the impact of proteolytic cleavage of the LCs in the bloods are largely unknown and subjects of current research.

1.4. Immunoglobulins

Immunoglobulins (Igs) or antibodies are an important part of the adaptive immune system of higher vertebrates (Kindt et al., 2003; Schroeder and Cavacini, 2010). They occur as cell bound B cell receptors and as secreted macromolecules upon antigen binding and B cell differentiation to a mature plasma cell. They are part of the Ig superfamily together with other receptors like T cell receptors, MHC I and II (involved in antigen presentation), CD4 and CD8 (co-receptors), etc. (Bateman et al., 1996).

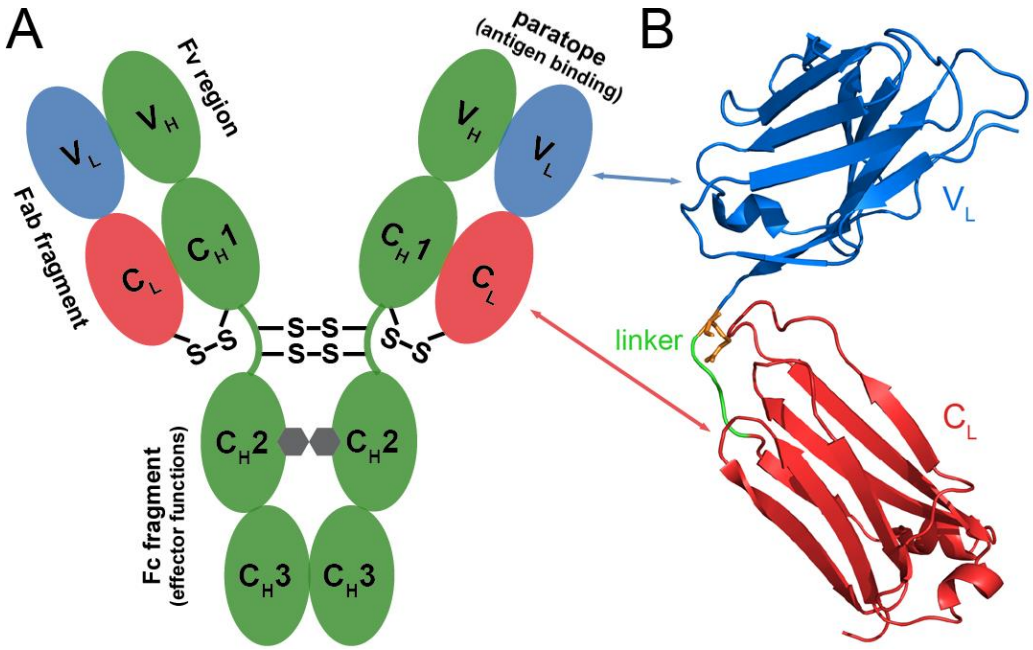

Figure 5. Structure of immunoglobulin G (IgG) and the antibody light chain (LC). A) *Schematic structure of an IgG antibody. Each oval area represents an individual Ig domain. Antibodies are composed of two heavy chains (HC) (green) and two light chains (LC) (blue and red), which are connected by disulfide bonds (S-S). The IgG C_H2 domain is glycosylated (grey hexagon).* ***B)*** *Crystal structure of the MAK33 antibody light chain (PDB: 1FH5, chain L) (Augustine et al., 2001). The variable light chain (V_L) and the constant light chain (C_L) are connected by an unstructured linker (light green) comprising Arg108 (orange). Adapted and modified with permission from Feige et al., 2010.*

Antibodies share a characteristic Y-shape and are composed of two identical heavy chains (HC) and two identical light chains (LC), which are connected by disulfide bonds (Figure 5A). The paratopes of antibodies are located in the Fv (fragment variable) region, part of the Fab

(fragment antigen binding) region, and have the inherent ability to specifically bind antigens, which are typically foreign pathogens. This is facilitated by the variable light chain (V_L) and the variable heavy chain (V_H) domain, which each comprise three hypervariable complementarity determining regions (CDRs), defined by V(D)J recombination and somatic hypermutation during B cell maturation (French et al., 1989). In general, only these segments within the variable domains exhibit a high sequence variation, whereas all other antibody domains are sequentially conserved. Higher vertebrates comprise five antibody classes, IgA, IgD, IgE, IgG and IgM, which are defined by the HC. Mediated by the Fc (fragment crystallizable) region, certain immunoglobulins can activate the complement system leading to local inflammation and lysis of the foreign cell (Ravetch and Bolland, 2001; Rus et al., 2005). Other physiological modes of action are (i) neutralization by sterically blocking distinct surface areas of pathogens, (ii) agglutination of pathogens, and (iii) precipitation of pathogens. Sophisticated non-physiological mechanisms have been developed for therapeutic antibodies, such as binding radioactive agents to the antibody (Sharkey and Goldenberg, 2005; Weiner et al., 2010). The antibody light chain is composed of the V_L domain and the constant light chain (C_L) domain, which are connected by an unstructured linker (Figure 5B). There are two types of LCs, λ and κ, which share the same structure, but exhibit differences in the sequence. Both LC types can assemble with each HC class. More information on the immunoglobulin LC is provided in chapter 1.4.2.

1.4.1. The Immunoglobulin (Ig) fold

Although Ig domains, particularly V_L and V_H, vary in their amino acid sequence, they share a highly conserved tertiary structure, called the Ig fold, which features two β-sheets forming a sandwich-like greek-key β-barrel topology (Figure 6) (Bork et al., 1994; Feige et al., 2010). The topology of variable and constant Ig domains varies in the number of defined β-strands (Williams and Barclay, 1988). Variable domains, i.e. V_L and V_H, comprise nine β-strands (ABCC'C''DEFG), whereas constant domains contain seven strands (ABCDEFG), lacking the additional C' and C''' strands.

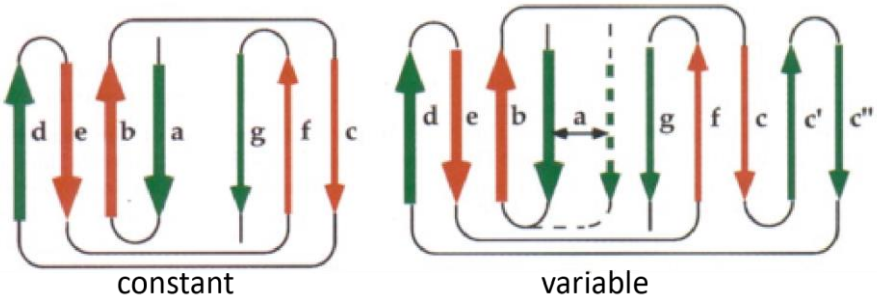

Figure 6. Topology of the Ig fold. *Constant (left) and variable (right) antibody domains contain seven and nine β-strands, respectively. β-strand a is known to have two alternative locations (dotted line). The red strands define the hydrophobic core. Thin and thick arrows represent β-strands at separate locations on the Z-axis, i.e. either in the front or in the back. Note that the length of the strands does not match reality. Adapted and modified with permission from Bork et al., 1994.*

For both types, the same strands make up the hydrophobic core. Within the core, antibody domains comprise a conserved disulfide bond in close proximity to a tryptophan residue (Lesk and Chothia, 1982). The disulfide bond does significantly contribute to the overall stability of an Ig domain and is orientated perpendicular to the β-sheets (Glockshuber and Schmidt, 1992). Moreover, it is able to quench the intrinsic fluorescence of the nearby, highly conserved tryptophan residue. Upon unfolding of the domain, the disulfide bond and the tryptophan move apart from each other, leading to an increased fluorescence intensity (Feige et al., 2004; Goto and Hamaguchi, 1979). Therefore, intrinsic tryptophan fluorescence is commonly used to probe protein unfolding of antibodies (Ghisaidoobe and Chung, 2014; Phillips et al., 1986).

1.4.2. The antibody light chain (LC)

The antibody light chain consists of a variable (V_L) and a constant (C_L) domain, connected by a short unstructured linker. The V_L domain comprises the CDRs which are responsible for antigen binding. The sequence of the CDRs is hypervariable due to V(D)J rearrangement followed by subsequent somatic hypermutation (French et al., 1989; Schroeder and Cavacini, 2010). To buffer adverse effects of the sequence variability, the LC must exhibit an intrinsic high tolerance against misfolding by destabilizing residues as well as a sophisticated quality control mechanism (Ellgaard and Helenius, 2003). However, in some cases, compensation and quality control fail and in consequence mutations in the light chain sequence results in systemic antibody light chain (AL) amyloidosis (cf. chapter 1.3.2.). In the past, primarily the V_L fragment was found in fibril samples extracted from human tissues (Buxbaum, 1986; Buxbaum et al., 1990; Glenner et al., 1970, 1971). However, more recent data provide evidence for the

presence of both, the full-length LC and the V_L in amyloid deposits, along with other proteins like apolipoprotein E (Dasari et al., 2015; Klimtchuk et al., 2010; Lavatelli et al., 2008). In this context, a key role of proteolytic cleavage accompanied with protein dynamics and transient unfolding of LCs triggering fibril formation and a severe pathogenesis of AL patients has been suggested (Buxbaum, 1986; Linke et al., 1973; Morgan and Kelly, 2016; Morgan et al., 2017; Oberti et al., 2017). The folding of the individual V_L domain occurs relative slowly compared to other Ig domains, e.g. C_L and C_H2. There are two distinct folding pathways for the V_L domain to adopt the native conformation (Simpson et al., 2009). Both pathways contain a common first intermediate state characterized by a widely undefined tertiary structure similar to the molten globule state (cf. chapter 1.1.). Starting from this first intermediate, the folding pathway exhibits two branches either directly adopting the native state or forming a second native-like intermediate. It was suggested that the first intermediate is prone to initiate an off-pathway transition to the fibrillary state due to its exposed hydrophobic patches. In contrast to constant antibody domains, it is known for the V_L domain that the loops comprising the CDRs impact the folding pathway, the maintenance of the native conformation and its stability (Helms and Wetzel, 1995; Jung and Pluckthun, 1997; Kurchan et al., 2005). In comparison to V_L, the C_L domain folds relatively fast, while exhibiting an intermediate state. In detail, the first transition from the unfolded to the intermediate state is slower than the second transition from the intermediate to the native state (Goto and Hamaguchi, 1981, 1982). The slow-folding first step of the folding pathway is prolyl isomerization-dependent, which is a common feature of the folding pathway of most Ig domains (Feige et al., 2010). It was shown that the C_L domain can adopt a native-like conformation without the oxidization of the intramolecular disulfide bond, which is however, more aggregation-prone and less stable than the fully oxidized C_L (Feige et al., 2007; Goto and Hamaguchi, 1979). Recent results indicate that the two α-helices, which are featured by constant antibody domains, are responsible for rapid folding initiation by properly orientating the two characteristic β-sheets resulting in favorable interactions making up the hydrophobic core (Feige et al., 2008). As a result, less amyloidogenic intermediates occur during protein folding reducing the overall amyloidogenicity of C_L. Besides aggregation-inducing and destabilizing mutations, previous studies identified the LC linker residue Arg108 to be important for the integrity and stability of both, the individual V_L and C_L, domains (Nokwe et al., 2014, 2015, 2016). However, little is known about how the LC modulates the properties of its individual domains and how the LC linker is involved in the

development of AL amyloidosis. In this context, Klimtchuk et al. found a destabilizing effect of an amyloidogenic V_L domain on a $C_L\lambda$ domain, whereas Blancas-Meija et al. showed that a CLλ and a $C_L\kappa$ domain do not impact the stability of a full-length LC (Blancas-Mejía et al., 2015; Klimtchuk et al., 2010). Thus, the role of the LC linker and the C_L for amyloid fibril formation remains still enigmatic.

1.4.3. The constant heavy domain 2 (C_H2)

The Fc region of immunoglobulins is a dimer each part consisting of two linked constant heavy chain domains, namely C_H2 and C_H3 (Bengtén et al., 2000). Both domains have been shown to play a key role in mediating effector functions and preserving antibody stability (M H Tao, S M Canfield, 1991; Morgan et al., 1995). Understanding the elements that govern the stability, integrity and effector function of the Fc domains is of key interest. The C_H2 domain is a typical antibody domain, which exhibits the highly conserved Ig fold (Feige et al., 2009a). In mature full-length antibodies, the C_H2 domain is glycosylated at the conserved Asn297 residue (Arnold et al., 2007). This glycosylation provides stability for the domain by protecting the inner hydrophobic core from the aqueous solvent. The unglycosylated murine MAK33 C_H2 domain is an untypically instable and aggregation-prone monomer, the unfolding of which already starts at physiological temperatures (Feige et al., 2004). In the full-length antibody context, a non-glycosylated C_H2 domain lowers the stability and integrity of the entire protein (Mimura et al., 2001). The isolated C_H2 domain can adopt the native state via three distinct folding pathways (Feige et al., 2004). It was shown that the human IgG C_H2 domain, which shares 64 % homology with the MAK33 IgG C_H2, is significantly more stable when seven residues at the N-terminus were deleted (Gong et al., 2013a). It was further stabilized by a second engineered intramolecular disulfide bond (Gong et al., 2009). However, the factors governing the stability of the C_H2 domain and other Ig domains are still incompletely understood. The human C_H2 domain has been suggested to serve as a scaffold, which easily penetrates into tissues and mediates effector functions in parallel (Gehlsen et al., 2012). Antigen binding can be achieved by engineering target recognizing surface-exposed loops (Li et al., 2017).

1.5. Objectives

This thesis aims to address two different research objectives, which focus on the integrity of antibody domains or their assembly to amyloid fibrils.

The first goal was to continue work on amyloidogenic VL domains (Nokwe et al., 2014, 2015, 2016) and expand it to the full-length LC. The impact of the LC linker was of special interest since it has been suggested that the LC linker is particularly susceptible for proteolytic cleavage leading to LC fragments (Nokwe et al., 2015), which represent the major proportion of amyloid plaques (cf. chapter 1.4.2.). In this context, another important aim was to determine the role of the C_L domain in the full-length LC context, as there is increasing evidence that the LC, not only the V_L-containing fragments, plays a significant role in amyloid fibril formation (cf. chapter1.4.2.). Thus, MAK33 LCκ and amyloidogenic variants were investigated regarding stability, fold and fibril formation propensity.

The second objective was to investigate the impact of the C-terminus of the MAK33 IgG1 C_H2 domain on the domain integrity. In the past, the stability of antibody domains was primarily assigned to the intramolecular disulfide bonds, the hydrophobic core and the presence or absence of elements interfering with the Ig fold, e.g. mutations, which disrupt an important β-strand. Surprisingly, Gong et al. found that shortening the N-terminus of a human IgG1 C_H2 domain by seven unstructured residues leads to a substantial stability increase (Gong et al., 2013a) (cf. chapter 1.4.3.). As the IgG1 C_H2 domain represents a typical and well-characterized antibody domain scaffold, which is N- and C-terminally extended by the C_H1 and C_H3 domains, respectively, it serves as an excellent framework to investigate the influence of the C-terminal region on the Ig framework.

2. Material and Methods

2.1. Materials

2.1.1. Devices

Device	Supplier
Autoclave Varioclav EP-Z	EP-Z
Cell Disruption Apparatus Basic Z	Constant Systems
Centrifuges	
Avanti J25 and J26 XP	Beckman Coulter
Rotina 420R	Hettich
Rotina 46R	Hettich
Tabletop centrifuge 5418	Eppendorf
Tabletop centrifuge Mikro R200	Hettich
Universal 320R	Hettich
Chromatography systems	
ÄKTA Prime	GE Healthcare
ÄKTA Purifier	GE Healthcare
Dawn Heleos II MALS detector	Wyatt
Frac-900/950 fraction collectors	GE Healthcare
HPLC System	Shimdazu
Superloops (150 mL, 10 mL)	GE Healthcare
Circular dichroism spectropolarimeters	
J710 (with PFD-350S Peltier device)	Jasco
J715 (with PTC 348 WI Peltier device)	Jasco
Fluorescence spectrophotometers	
Jasco FP-6500	Jasco
FluoroMax-4	Horiba Jobin Yvon
Gel electrophoresis devices	Hoefer
Hydrogen/deuterium exchange instruments	
ACQUITY M-Class UPLC	Waters
PAL RTC	LEAP
Synapt G2-S ESI-TOF mass spectrometer	Waters
Ice maker	Zieger
Image Scanner III	GE Healthcare
Incubator	New Brunswick Scientific
JEM-1400 Plus	JOEL
Magnetic stirrer	
MR2000	Heidolph
MR3001	Heidolph

MR80	Heidolph
MARKII refractometer	Leica
Membrane vacuum pump	Sartorius
Microplate readers	
GENios	Tecan
Infinite M Nano Microplate Reader	Tecan
Mixer Homogenizer SilentCrusher M	Heidolph
NMR spectrometer	
Bruker AVIII 500 MHZ spectrometer	Bruker
Bruker AVIII 600 MHZ spectrometer	Bruker
pH meter	WTW
Power amplifiers EPS 3500, 3501, 1001	GE Healthcare
Scales	
BP 121 S	Sartorius
BL 310	Sartorius
Thermal cycler MJ Mini 48 well	Biorad
Thermal cycler Primus 25	MWG
Thermoblocks	
Digital heat block	VWR
Eppendorf-Thermomixer	Eppendorf
TB1 Thermoblock	Biometra
Ultraflex II MALDI ToF/ToF	Bruker Daltonics
UV-Vis spectrophotometers	
UltroSpec 1100 pro	Amersham Biosciences
Nanodrop	Peqlab
Vortex MS2	IKA
Water bath F6-K	Haake

2.1.2. Chemicals

Chemical	Supplier
^{13}C-Glucose	Cambridge Isotope Laboratories
$^{15}NH_4Cl$ (ammonium chloride, nitrogen 15 isotope)	Cambridge Isotope Laboratories
2-Mercaptoethanol	Merck
5,5'-Dithiobis(2-nitrobenzoic acid) (DTNB)	Merck
Acetic acid	Roth
Acetonitrile	Merck

Acrylamid/Bisacrylamide solution 38:2 (40% w:v)	Serva
Agar Agar	Serva
Agarose	Serva
Ammonium chloride (NH_4Cl)	Merck
Ammonium persulfate (APS)	Roth
Ampicillin sodium salt	Roth
Biotin	Merck
Boric acid (H_3BO_3)	Merck
Bromphenol blue	Serva
Calcium chloride ($CaCl_2$)	Merck
Cobalt (II) chloride ($CoCl_2$)	Merck
Coomassie Blue R	Serva
Coomassie Brilliant Blue R-250	Serva
Copper (II) chloride ($CuCl_2$)	Merck
Deoxynucleoside triphosphates (dNTPs)	NEB
Deuterium oxide (D_2O)	Merck
Dimethyl sulfoxide (DMSO)	NEB
Dithiothreitol (DTT)	Roth
EDTA	Merck
Ethanol	Merck
Formic acid	Merck
Glucose	Merck
Glutathione, oxidized (GSSG)	Merck
Glutathione, reduced (GSH)	Merck
Glycerol	Roth
Glycine	Roth
Guanidinium chloride (GdmCl)	Merck
HEPES	Merck
Hydrochloric acid (HCl) 32%	Merck
Iron (III) chloride ($FeCl_3$)	Merck
Isopropanol	Merck
Isopropyl β-d-1-thiogalaktopyranoside (IPTG)	Serva
Kanamycin sulfate	Roth
L-Arginine	Merck
LB medium	Serva
Magnesium sulfate ($MgSO_4$)	Merck
Manganese (II) chloride	Merck
Potassium chloride (KCl)	Roth
Potassium phosphate, monobasic (KH2PO4)	Merck
Protease inhibitor Mix G, HP	Serva
Sodium acetate anhydrous (NaOAc)	Merck
Sodium azide (NaN_3)	Merck
Sodium chloride (NaCl)	Merck
Sodium dodecylsulfate (SDS)	Serva
Sodium hydroxide (NaOH)	Merck
Sodium phosphate, dibasic (Na2HPO4 * 2 H2O)	Merck
Sodium phosphate, monobasic (NaH_2PO_4 * 2 H_2O)	Merck
Stain G	Merck

Tetraethylethylendiamin (TEMED)	Roth
Thiamin-HCl	Merck
Thioflavin T (ThT)	Merck
Tris	Roth
Tris(2-carboxyethyl)phosphine hydrochloride (TCEP)	Merck
Triton X-100	Merck
Uranyl acetate	Science Services
Urea	Merck
Zinc chloride ($ZnCl_2$)	Merck

Deionized water produced by Millipore Direct Q5-System

2.1.3. Consumables

Consumable	Supplier
200-mesh copper grids	Merck
Amicon Ultra-0.5 Centrifugal Filter Units	Millipore
Amicon Ultra-15 Centrifugal Filter Units	Millipore
Amicon Ultra-4 Centrifugal Filter Units	Millipore
Biacore Sensor Chip CM5	GE Healthcare
Crystal Clear PP sealing foil	HJ-Bioanalytik GmbH
Cuvettes, plastic, 1 mL	Brand
Dialysis membranes Spectra/Por (6-8 kDa)	Spectrum Laboratories
Immobilon-P membrane (PVDF)	Roth
Membrane discs	Sartorius
Membrane filter 0.22 µm	Millipore
Microplates 96 well, black, COC, #655809	Greiner Bio-One
Microplates 96 well, black, PS, #437112	Nunc
Millex - GS 0.22 µm syringe filter units	Millipore
Nitrocellulose membrane, pore size 0.2 µl	Amersham Biosciences
PCR tubes	BioRad Laboratories
PE tubes, 15 and 50 mL	Greiner & Söhne
Petri dishes, PS, 94 mm	Greiner & Söhne
pH indicator	Merck
Reaction tubes, various volumes	Sarstedt
Sterile filter 0.2 µm	Zefa

2.1.4. Enzymes, Standards and Kits

Product	Supplier
DNAseI	AppliChem Pancreac
FastLoad 1 kb DNA ladder	Serva
Gel loading dye, purple	NEB
GoTaq DNA Polymerase	Promega
Low-Range SDS-PAGE Standard	Biorad
NEBuffer 2.1	NEB
Pf1 phage	ASLA biotech
Phusion HF DNA polymerase	NEB

Precision Plus Protein Dual Color Standard	Biorad
Q5 HF DNA Polymerase	NEB
Q5 Site-Directed Mutagenesis Kit	NEB
Restriction enzymes	NEB and Promega
T4 DNA Polymerase	NEB
Wizard Plus SV Minipreps DNA Purification System	Promega
Wizard SV Gel and PCR clean-up system	Promega

2.1.5. Buffers and solutions

2.1.5.1. Buffers for protein purification

Type	Composition
5x Inclusion body preparation buffer	250 mM Tris / HCl pH 7.5 50 mM EDTA 50 mM NaCl add protease inhibitor mix before use
Q-Sepharose buffer A, low salt	50 mM Tris / HCl pH 8.0 10 mM EDTA 5 M Urea
Q-Sepharose buffer B, high salt	50 mM Tris / HCl pH 8.0 10 mM EDTA 5 M Urea 1 M NaCl
Inclusion body dissolving buffer	50 mM Tris / HCl pH 8.0 10 mM EDTA 8 M Urea 0.1 % (v/v) 2-Mercaptoethanol, before use
Refolding buffer	250 mM Tris / HCl pH 8.0 10 mM EDTA 100 mM L-arginine 0.5 mM GSH, add freshly before use 1.0 mM GSSG, add freshly before use
10x Phosphate buffered saline (PBS)	100 mM Na_2HPO_4 x 2 H_2O 18 mM KH_2PO_4 / pH 7.4 27 mM KCl 1.37 M NaCl

2.1.5.2. Buffers for SDS-PAGE

Fairbanks A (staining)	25 % (v/v) Isopropanol 10 % (v/v) Acetic acid 0.05 % (w/v) Coomassie Blue R
Fairbanks D (destaining)	10 % (v/v) Acetic acid
5x Laemmli buffer, reducing	0.3 M Tris / HCl pH 6.8 10 % (w/v) SDS 50 % (v/v) Glycerol 5 % (v/v) 2-Mercaptoethanol 0.05 % (w/v) Bromophenol blue
5x Laemmli buffer, non-reducing	0.3 M Tris / HCl pH 6.8 10 % (w/v) SDS 50 % (v/v) Glycerol 0.05 % (w/v) Bromophenol blue
10 x SDS running buffer	0.25 M Tris / HCl pH 8.0 2 M Glycine 1 % (w/v) SDS
4x Separation gel buffer	1.5 M Tris / HCl pH 8.8 0.8 % (w/v) SDS
2x Stacking gel buffer	0.25 M Tris / HCl pH 6.8 0.4 % (w/v) SDS

2.1.5.3. Buffers for Molecular Biology

50x TAE	2 M Tris / Acetate pH 8.0 50 mM EDTA

2.1.5.4. Buffers for fibril formation assay

Assay buffer	1x PBS pH 7.4 0.05 % (w/v) NaN_3 10 µM ThT, final concentration
500 µM ThT stock solution	1x PBS pH 7.4 500 µM ThT

2.1.5.5. Buffers for chemical competent cells

Solution A	13 ml	3 M NaOAc, pH 5.5
	100 ml	1 M CaCl2
	25 ml	2.8 M MnCl2
	862 ml	H_2O
Solution B	69ml	Glycerol (87%)
	331ml	Solution A

2.1.5.6. Buffers for hydrogen/deuterium exchange experiments

Quenching buffer	200 mM	Na_2HPO_4, pH 2.2
	200 mM	NaH_2PO_4
	250 mM	Tris(2-carboxyethyl)phosphine
	3 M	GdmCl

2.1.6. Chromatography materials and columns

Type	Supplier
ACQUITY UPLC BEH C18	Waters
Enzymate BEH Pepsin	Waters
Q Sepharose Fast Flow	GE Healthcare
Superdex 200 10/300 GL	GE Healthcare
Superdex 200 16/60	GE Healthcare
Superdex 75 10/300 GL	GE Healthcare
Superdex 75 16/60	GE Healthcare

2.1.7. Bacteria strains and plasmids

Strain	Genotype	Origin
E.coli BL21 (DE3)	codon plus F– *ompT hsdS*(rB– mB–) *dcm+* Tetr *gal endA* Hte (*argU proL* Camr)	Stratagene
E.coli Mach1	ΔrecA1398 endA1 tonA Φ80ΔlacM15 ΔlacX74 hsdR (rk - mk +)	Invitrogen

2.1.8. Media for Bacteria

Medium	Composition
Luria Bertani (LB_0), Serva	20 g/l for medium 15 g/l agar agar to LB medium for plates
Terrific broth (TB) for competent cells	10 mM HEPES / HCl pH 6.7 15 mM $CaCl_2$ 250 mM KCl 55 mM $MnCl_2$

1x M9 medium	100 ml	10x M9 solution
	10 ml	100x trace elements
	1 ml	1 M $MgSO_4$
	0.3 ml	1 M $CaCl_2$
	1 ml	1 mg/ml thiamin-HCl
	1 ml	1 mg/ml biotin
	20 ml	20 % (w/v) glucose
	867 ml	H_2O, sterile
10 x M9 solution	33.7 mM Na_2HPO_4 / NaOH pH 7.2	
	22.0 mM KH_2PO_4	
	8.55 mM NaCl	
	9.35 mM NH_4Cl	
100x trace elements	13.4 mM EDTA	
	3.1 mM $FeCl_3$	
	0.62 mM $ZnCl_2$	
	76 µM $CuCl_2$	
	42 µM $CoCl_2$	
	162 µM H_3BO_3	
	8.1 µM $MnCl_2$	
1000x Kanamycin	35 mg/ml (in H_2O)	

Media were autoclaved at 121 °C for 20 min. When autoclaving was not possible (antibiotics, thiamin-HCl, biotin, labeled ^{13}C-glucose, $^{15}NH_4Cl$, IPTG) due to heat-instable chemicals, sterile filters (0.22 µm) were utilized. Antibiotic and IPTG stocks were stored at -20°C.

2.1.9. Plasmids

Plasmid	Vector	Origin	Reference
MAK V_L wt	pET28b	Cardine Nokwe	(Nokwe et al., 2014)
MAK V_L I2E	pET28b	Cardine Nokwe	(Nokwe et al., 2014)
MAK V_L S20N	pET28b	Cardine Nokwe	(Nokwe et al., 2016)
MAK V_L R61A	pET28b	Cardine Nokwe	(Nokwe et al., 2016)
MAK V_L ΔR108	pET28b	Cardine Nokwe	(Nokwe et al., 2015)
MAK C_L wt	pET28b	Cardine Nokwe	(Nokwe et al., 2015)
MAK C_L ΔR108	pET28b	Cardine Nokwe	(Nokwe et al., 2015)
MAK LC wt	pET28b	Eva Seedig	N/A
MAK C_H2 wt	pET28b	Christine John	(John, 2017)

2.1.10. Oligonucleotides

Template	Name	sequence (5’ to 3’)
MAK V_L (QuikChange)		
	R108A fw	caagctggagctgaaagcataatagaagcttgcgg
	R108A rv	ccgcaagcttctattatgctttcagctccagcttg
	R108E fw	ccaagctggagctgaaagaataatagaagcttgcgg
	R108E rv	ccgcaagcttctattattctttcagctccagcttgg
MAK C_L wt (QuikChange)		
	R108A fw	gggtgctgcgtctgctgccatggtatatctcctt
	R108A rv	aaggagatataccatggcagcagacgcagcaccc
	R108E fw	cagtggtgctgcgtctgcctccatggtatatctccttctt
	R108E rv	aagaaggagatataccatggaggcagacgcagcacccactg
MAK LC (QuikChange)		
	I2E fw	gctggagactgagttagcacctcatccatggtatatctccttcttaaag
	I2E rv	ctttaagaaggagatataccatggatgaggtgctaactcagtctccagc
	S20N fw	ccctgcaagaaagattgacgctatctcctggag
	S20N rv	ctccaggagatagcgtcaatctttcttgcaggg
	R61A fw	cactgccactgaacgcggaggggatcccag
	R61A rv	ctgggatcccctccgcgttcagtggcagtg
	R108A fw	cagcatcagccgctttcagctccagcttggtccc
	R108A rv	gggaccaagctggagctgaaagcggctgatgctg
	R108E fw	cagcatcagcctctttcagctccagcttggtccc
	R108E rv	gggaccaagctggagctgaaagaggctgatgctg
MAK LC (Q5 mutagenesis)		
	link+G fw	ggggctgatgctgcaccaact
	link+G rv	ccgtttcagctccagctt
	link+GAGAG fw	agcaggggctgatgctgcaccaact
	link+GAGAG fw	ccagccccccgtttcagctccagctt
	Δlink fw	actgtatccatcttcccacc
	Δlink rv	cagctccagcttggtccc
	R61A S40P fw	tcaacaaaaaccacatgagtctc
	R61A S40P rv	taccagtgtaggttgttg
	R61A T80A fw	aatacattccaaaatcttcagcctccacactgttgatactgag
	R61A T80A rv	ctcagtatcaacagtgtggaggctgaagattttggaatgtatt
	R61A S168A fw	ctgtaggtgctgtctttggcgtcctgatcagtccaact
	R61A S168A rv	agttggactgatcaggacgccaaagacagcacctacag
	R61A K169A fw	tcaggacagcgcagacagcacct
	R61A K169A rv	tcagtccaactgttcagg
MAK C_H2 (QuikChange)		
MAK C_H2 (Q5 mutagenesis)		
	C_H2-S fw	tcctaatagcccaagcttgcg
	C_H2-S rv	gatggttttctcgatggg
	C_H2-SK fw	tccaaataatagcccaagcttgcg
	C_H2-SK rv	gatggttttctcgatggg
	C_H2-SKTK fw	acaaaataatagcccaagcttgcg
	C_H2-SKTK rv	tttggagatggttttctcgatggg

2.2. Software, databases and web-based tools

Software

Name	Supplier
Adobe Illustrator CS5	Adobe Inc.
Adobe Photoshop CS5	Adobe Inc.
ASTRA 5.3.4	Wyatt
CcpNmr	CCPN
DynamX 3.0	Waters
Epson TT8 Launch Silver Fast	Epson
GIMP 2	GIMP team
GROMACS 2016	(Abraham et al., 2015)
ISDB PLUMED	(Bonomi and Camilloni, 2017)
Mendeley Desktop	Mendeley Ltd.
Microsoft Office 2010 and 2016	Microsoft
Origin 7	OriginLab Corp.
OriginPro 2015	OriginLab Corp.
PLUMED 2	(Tribello et al., 2014)
Protein Lynx Global Server PLGS 3.0.3	Waters
PyMOL v1.3	Schrödinger
Serial Cloner 2.5	Franck Perez
TopSpin	Bruker

Databases

Name	Website
Abysis	http://www.bioinf.org.uk/abysis2.7
ALBase	http://albase.bumc.bu.edu/aldb
Google Scholar	http://scholar.google.de/
IMGT	http://www.imgt.org
PDB	http://www.rcsb.org/
PubMed	http://www.ncbi.nlm.nih.gov/pubmed
UniProt	http://www.uniprot.org

Web-based tools

Name	Website
Abysis	http://www.bioinf.org.uk/abysis2.7
Clustal Omega alignment	http://www.ebi.ac.uk/Tools/msa/clustalo/
Denaturant Calculator	http://sosnick.uchicago.edu/gdmcl.html
ExPASy ProtParam	http://web.expasy.org/protparam/
NCBI Blast	http://blast.ncbi.nlm.nih.gov/
NCBI IgBlast	http://www.ncbi.nlm.nih.gov/igblast/
NEBaseChanger	http://nebasechanger.neb.com/
NEBioCalculator	http://nebiocalculator.neb.com/#!/main
NEBuilder	http://nebuilder.neb.com/
OligoAnalyzer 3.1	http://eu.idtdna.com/analyzer/applications/oligoanalyz

Phyre2	http://www.sbg.bio.ic.ac.uk/phyre2
PrimerX	http://bioinformatics.org/primerx/
QuikChange Primer Design	http://www.genomics.agilent.com/primerDesignProgram.jsp
T-COFFEE alignment	http://tcoffee.crg.cat/

2.3. Molecular biology methods

All solutions applied were autoclaved or sterile filtered. If not otherwise stated, all experiments were performed at room temperature.

2.3.1. Preparation of chemical competent *E. coli* cells

Chemical competent *E. coli* cells were prepared according to the protocol by Sambrook et al., 1989. In brief, an aliquot of frozen *E. coli* cells was thawed, transferred to 100 ml LB_0 medium and incubated at 37°C under continuous shaking. Bacterial growth was monitored by measuring the optical density at 600 nm (OD_{600}) in plastic cuvettes with 1 mm path length using a UV-Vis spectrophotometer. 2 ml 1 M $MgCl_2$ were added when an OD_{600} of 0.6 to 1.0 was reached. The cells were subsequently incubated at 37°C under continuous shaking. After 10 min cells were put on ice for 1 h. Cells were harvested by centrifugation at 5000 rpm for 5 min and resuspended in solution A. The cell suspension incubated on ice for 1 h. Then, the cells were again harvested by centrifugation and resuspended in 2 ml of Solution B. Aliquots were prepared, snap-frozen and stored at -80°C.

2.3.2. Transformation of *E.coli* cells

100 µl chemical competent *E.coli* were thawed and kept on ice for ~ 10 minutes. Next, 1 µl (50 - 100 ng) of plasmid DNA was added to the cells, mixed and incubated on ice for 5 minutes. The mixture was heat-shocked at 42 °C for 45 sec, followed by a 2 min incubation on ice. 900 µl LB_0 medium were added and the cells were incubated at 37°C for 1 h under agitation. The cells were harvested by centrifugation at 14000 rpm for 1 min, resuspended in ~ 200 µl LB_0 medium and plated on the LB_{Kan} plates (35 µg/ml final concentration). Plates were incubated at 37 °C overnight.

2.3.3. DNA isolation and storage

DNA plasmids were isolated from 4 ml *E.coli* cultures the Wizard® Plus SV Mini-Prep Kit according to the manufacturer's protocol (Promega). PCR products and restriction-digested DNA were purified using the Wizard® SV Gel and PCR Clean-Up System according to the manufacturer's protocol (Promega). The isolated DNA was stored in sterilized nuclease free H_2O at -20 °C. For DNA sequencing, plasmid DNA samples with a concentration of 50 to 100 ng/µl were sent to GATC Biotech AG or Eurofins Genomics GmbH for sequencing using the T7 forward or T7 reverse primer.

2.3.4. Cloning strategies

2.3.4.1. QuikChange site-directed mutagenesis

Single point substitutions were generated using QuikChange site-directed mutagenesis (Agilent). Both, forward and reverse primers, carry the desired mutation. Primers were designed using the web-based QuikChange Primer Design tool and the annealing temperature was adjusted to the properties of the primers. After the PCR reaction, DpnI digest was performed for 2 h. 10 µl of the digested product were transformed with 100 µl *E.coli* Mach1 cells, plated on LB_{Kan} plates and incubated at 37 °C overnight. DNA isolation and sequencing were performed as described above.

2.3.4.2. Q5 site-directed mutagenesis

Double and triple point substitutions, insertions and deletions were generated using Q5 site-directed mutagenesis (NEB). In this approach, only the forward primer, carries the desired mutation. Primers were designed using the web-based NEBaseChanger tool and the annealing temperature was adjusted to the properties of the primers. After the PCR reaction, 1 µl PCR product was transferred in a reaction tube and assembled for kinase, ligase and DpnI (KLD) treatment according to the manufacturer's protocol. 10 µl of the digested product were transformed with 100 µl *E.coli* Mach1 cells, plated on LB_{Kan} plates and incubated at 37 °C overnight. DNA isolation and sequencing were performed as described above.

2.3.4.3. Sequence- and ligation-independent cloning (SLIC)

Sequence- and ligation-independent cloning (SLIC) (Li and Elledge, 2012) utilizes the exonuclease function of the T4 DNA polymerase to generate single-stranded 5' overhangs for vector and insert DNA. These fragments assemble in a site-directed manner by homologous recombination after transformation into *E.coli* Mach1 cells. To enable homologous recombination, insert and vector share at least 15 bp homology at the regions which should be assembled. Insert and vector were amplified by standard PCR. The vector was linearized by NcoI and HindIII restriction enzymes. Insert and digested vector were purified from an agarose gel. 100 ng vector, 2-fold excess of insert and 0.4 µl T4 DNA polymerase were incubated in 1 µl 10x NEBuffer 2.1 (NEB), filled up with nuclease free H_2O to 10 µl volume (= assembly mix), for 2.5 min and subsequently stored on ice for 10 min. The whole assembly mix was transformed with *E.coli* Mach1 cells, plated on an LB_{Kan} agar plate and incubated overnight at 37 °C. DNA isolation and sequencing were performed as described above.

2.3.4.4. Standard PCR

Standard PCR DNA amplification was used to subclone a gene of interest (GOI) into the SUMOpro Kan vector (LifeSensors). The SUMOpro vector enables SUMO fusion constructs, which keep their native N-terminus after cleavage of the SUMO-tag using Sumo protease by not modifying the amino acid sequence of the N-terminus. Subcloning of the GOI into the SUMOpro vector was performed according to the manufacturer's protocol.

2.3.5. Polymerase Chain Reaction (PCR)

DNA amplification and site-directed mutagenesis, i.e. QuikChange (Agilent) and Q5 site-directed mutagenesis (NEB) were performed by PCR. Since QuikChange and Q5 mutagensis require different DNA polymerases, two different protocols were used as described below. QuikChange and Q5 mutagensis were performed using Phusion DNA Polymerase (NEB) and Q5 DNA Polymerase (NEB), respectively.

QuikChange site-directed mutagenesis:

The following pipetting scheme was used for setting up QuikChange PCR reactions:

0.5 µl	Phusion DNA Polymerase
10 µl	5x Phusion buffer
2 µl	10 mM dNTPs
2 µl	10 µM forward primer
2 µl	10 µM reverse primer
1.5 µl	DMSO
2 µl	Template DNA (~ 5 ng)
29 µl	nuclease free H_2O

The following thermocycler program was used for QuikChange PCR reactions:

Step	Temperature	Duration	Repeats
Initial denaturation	95 °C	2 min	-
Denaturation	95 °C	30 sec	25x
Annealing	54 – 70 °C	1 min	
Elongation	72 °C	10 min	
Storage	4 °C	hold	-

Q5 site-directed mutagenesis:

The following pipetting scheme was used for setting up Q5 mutagenesis PCR reactions:

12.5 µl	Q5 Hot Start HF 2x Master Mix
1.25 µl	10 µM forward primer
1.25 µl	10 µM reverse primer
1 µl	Template DNA (~ 10 ng)
9 µl	nuclease free H_2O

The following thermocycler program was used for Q5 mutagenesis PCR reactions:

Step	Temperature	Duration	Repeats
Initial denaturation	98 °C	30 sec	-
Denaturation	98 °C	10 sec	25x
Annealing	54 – 70 °C	30 sec	
Elongation	72 °C	1min/kb	
Final extension	72 °C	2 min	-
Storage	4 °C	hold	-

Standard and SLIC PCR:

The following pipetting scheme was used for setting up standard and SLIC PCR reactions:

20 µl 5x GoTaq Polymerase buffer
2 µl 10 mM dNTPs
1 µl 10 µM forward primer
1 µl 10 µM reverse primer
2 µl Template DNA (100 - 200 ng)
73 µl nuclease free H_2O

The following thermocycler program was used for standard and SLIC PCR reactions:

Step	Temperature	Duration	Repeats
Initial denaturation	95 °C	2 min	-
Denaturation	95 °C	30 sec	35x
Annealing	54 – 70 °C	1 min	
Elongation	72 °C	1min	
Final extension	72 °C	10 min	-
Storage	4 °C	hold	-

2.3.6. Agarose gel electrophoresis

For isolation and size determination of DNA fragments and PCR products, agarose gel electrophoresis was utilized. By standard, 1 % agarose gels were prepared in TAE buffer with Stain G (Serva) according to the manufacturer's protocol. Stain G enables DNA visualization under UV radiation. The FastLoad 1 kb DNA ladder (Serva) was applied to the gels to determine the size of DNA fragments of interest. Agarose gels were run at 120 mV for 20 min.

2.4. Protein chemistry methods

2.4.1. Protein expression and purification

All V_L, C_L, LC, C_H2, IgM domain and IgA domain variants were expressed and purified as previously described for antibody V_L domains (Nokwe et al., 2014; Simpson et al., 2009). Briefly, the plasmid was transformed into *E.coli* BL21 (DE3) cells and grew in a 50 ml LB_{Kan} medium starter culture at 37 °C overnight. Protein overexpression was performed by inoculating 2 l LB_{Kan} medium with the starter culture and grow the cells to an OD_{600} of 0.6 – 1.0 at 37 °C. Overnight protein expression was induced by the addition of 1 mM IPTG. For NMR experiments, expression of ^{15}N-labelled or ^{13}C-^{15}N-labelled proteins was performed in M9 minimal medium containing $^{15}NH_4Cl$ and glucose or ^{13}C-glucose, respectively. Cells were harvested the next day and inclusion bodies were prepared as previously described (Thies and

Pirkl, 2000). Briefly, the cell pellet was resuspended in ~ 150 ml IB preparation buffer containing Protease Inhibitor Mix G or HP and DNAseI. Cells were disrupted using a mechanical cell cracker and inclusion bodies were harvested by centrifugation. Finally, inclusion bodies were washed twice with IB preparation buffer and stored until purification at -20 °C. For protein purification, inclusion bodies were solubilized and denatured in IB dissolving buffer and subsequently purified using a Q-Sepharose column, followed by a Superdex 75 16/60 gel filtration column. Between of the two purification steps the proteins were refolded overnight by dialysis against the refolding buffer (ratio protein solution to refolding buffer 1:20) at 4 °C. Fractions containing the protein of interest were determined by SDS-PAGE. Finally, protein purity and identity were verified by SDS-PAGE.

2.4.2. SDS polyacrylamide gel electrophoresis

SDS polyacrylamide gel electrophoresis (SDS-PAGE) gels were generally freshly prepared before use. The separation gels contained either 15 or 18 % (w/v) acrylamide/bisacrylamide (19:1). The stacking gel had a fixed concentration of 5 % (w/v) acrylamide/bisacrylamide. SDS-PAGE gels were prepared according to the following protocols:

Stacking gel 5 %	0.625 ml	acrylamide/bisacrylamide (40 % w/v)
	2.5 ml	2x stacking gel buffer
	65 µl	10 % (w/v) APS
	3.3 µl	TEMED
	1.875 ml	H_2O
Separation gel 15 %	3.75 ml	acrylamide/bisacrylamide (40 % w/v)
	2.5 ml	4x separation gel buffer
	65 µl	10 % (w/v) APS
	3.3 µl	TEMED
	3.75 ml	H_2O
Separation gel 18 %	4.5 ml	acrylamide/bisacrylamide (40 % w/v)
	2.5 ml	4x separation gel buffer
	65 µl	10 % (w/v) APS
	3.3 µl	TEMED
	4.25 ml	H_2O

Protein samples were mixed with 5x Laemmli buffer, heated for 2 min at 95 °C and subsequently loaded on the gel. SDS running buffer was poured in the gel chamber and the gel electrophoresis was performed at a constant current of 35 mA per gel for 40 – 60 min depending on the concentration of the separation gel. Gels were stained using Fairbanks A

solution (Coomassie stain) by heat incubation. Heat-activated destaining was performed using Fairbanks D solution (Coomassie destain). A protein marker was applied to determine the molecular mass of the proteins of interest.

2.5. 2.5 Spectroscopy

2.5.1. UV-Vis spectroscopy

UV-Vis spectroscopy was mainly utilized to determine the concentration of a protein. As some amino acids comprise chromophores most proteins are able to absorb light of specific wavelengths depending on the nature of the chromophores (Pace et al., 1995; Stoscheck, 1990). Chromophores are conjugated pi-bond systems with excitable electrons. Amino acids which contain chromophores are tryptophan, tyrosine and phenylalanine, which absorb light at 280, 275 and 257 nm wavelength (UV range), respectively. Disulfide bonds between two covalently connected cysteine residues contribute to the total UV light absorption of a protein. Particularly the indole group of a tryptophan is known to absorb high quantities of UV light at 280 nm. This enables the determination of the protein concentration based on the absorbance at 280 nm according to the Lambert-Beer law (Equation 2):

$$A_\lambda = \varepsilon * c * d$$

Equation 2. Lambert-Beer equation. *A_λ: absorbance at wavelength λ; ε: molar extinction coefficient (M^{-1} cm^{-1}); c: protein concentration (M); d: path length (cm).*

UV absorbance at 280 nm was measured using an Ultrospec 1100 UV/Vis spectrophotometer (Fisher scientific) or a NanoDrop spectrophotometer (Thermo scientific) at room temperature and baseline corrected. For the Ultrospec 1100 UV a quartz cuvette was used. Theoretical extinction coefficients were calculated on the basis of the amino acids sequence by the web-based ExPASy ProtParam tool.

2.5.2. Circular Dichroism spectroscopy

Circular dichroism (CD) measurements were performed to verify the correct folding, i.e. secondary and tertiary structure of proteins. Moreover, this technique was utilized to monitor protein unfolding upon increasing heat stress (thermal unfolding). CD spectroscopy monitors the differential absorption of right and left polarized light (Greenfield, 2006; Kelly et al., 2005). Proteins are chiral biomolecules due to the chirality of the α-carbon atom and the presence

of the asymmetric residue side chains, thus resulting in the optical ability to polarize light. The periodically repeated peptide bonds with their characteristic binding angles in secondary structure elements (e.g. α-helix, β-strand) absorb polarized far-UV light (180 – 260 nm) in different quantities. Hence, secondary structure elements lead to characteristic far-UV CD spectra, particularly for α-helices, β-strands and random coil. A CD spectrum, however, averages all secondary structure element, which a given proteins comprises. For the near-UV region (260 – 320 nm) aromatic side chains and disulfide bonds contribute for specific wavelength to the CD signal in the spectrum. Thus, near-UV CD spectra result in individual tertiary structure finger print of a polypeptide. The lambert-beer law is applicable for CD spectroscopy (Equation 3) (Kelly et al., 2005):

$$\Delta A(\lambda) = A_L(\lambda) - A_R(\lambda) = [\varepsilon_L(\lambda) - \varepsilon_R(\lambda)]dc = \Delta\varepsilon dc$$

Equation 3. Absorption of left- and right-handed polarized light. *A: absorption at wavelength λ for either left L or right R handed polarized light; ε: molar extinction coefficient (M^{-1} cm^{-1}); c: protein concentration (M); d: path length (cm).*

CD spectropolarimeter measure the ellipticity of a given protein depending on the path length, concentration, molecular mass and the number of amino acids. Therefore, the CD signal is normalized to the mean residue molar ellipticity (Equation 4) (Kelly et al., 2005):

$$\theta_{\mathrm{MRW}} = \frac{\theta * 100 * \mathrm{M}}{\mathrm{d} * \mathrm{c} * \mathrm{N_{aa}}}$$

Equation 4. Mean residue molar ellipticity. *ϑ_{MRW}: mean residue molar ellipticity (deg cm^2 $dmol^{-1}$), ϑ: ellipticity (deg), M: Molecular mass (g/mol), d: path length (cm), c: concentration (M), N_{aa}: number amino acids.*

CD measurements were performed using a J-720 and J-715 spectropolarimeter (Jasco) equipped with a Peltier element for temperature control. Far-UV CD spectra were measured in a 1 mm path length quartz cuvette using 10 µM protein between 260 and 200 nm wavelengths at 20 °C. Near-UV CD spectra were measured in a 2 mm path length quartz cuvette using 50 µM protein between 320 and 260 nm wavelengths at 20 °C. Spectra were accumulated 16 times, averaged, buffer corrected and normalized for mean residue molar ellipticity as described above. Thermal transitions were recorded using 10 µM protein in a 1 mm path length quartz cuvette at 205 (for C_L, V_L, LC and IgM/IgA variants) or 215 nm (for C_H2 variants) wavelength with a heating rate of 30 °C/h. Transitions were fitted using a Boltzmann function.

The ellipticity for the arithmetically (^{cal}) determined CD thermal unfolding transitions were calculated based on the experimental (^{exp}) CD unfolding transition data depending on the scaffold (LC, V_L or C_L) as (Equations 5 - 7):

$$\theta^{cal}(LC) = \theta^{exp}(V_L) + \theta^{exp}(C_L)$$

$$\theta^{cal}(V_L) = \theta^{exp}(LC) - \theta^{exp}(C_L)$$

$$\theta^{cal}(C_L) = \theta^{exp}(LC) - \theta^{exp}(V_L)$$

Equations 5-7. Ellipticity for arithmetically derived CD thermal unfolding transitions. *ϑ: ellipticity (deg).*

The analysis of the arithmetical CD thermal unfolding transitions was performed in the same way as described for the experimental data described above.

2.5.3. Fluorescence spectroscopy

Fluorescence emission is produced by the excitation of a fluorophore, i.e. an aromatic system, which is able to absorb photons and thereby exciting its ground state (S_0) to a higher energy state (S_1 and S_2) (Schmid, 2005). Photons in the excited S_1 and S_2 state can rapidly return to lowest accessible vibrational S_0 state, resulting in the emission of light with longer wavelengths (less energy) than the excitation light. This emission of light is called fluorescence, which is widely applied in the biochemistry field since some amino acids contain natural fluorophores. These amino acids are phenylalanine, tyrosine and tryptophan, of which tryptophan reveals the highest quantum yields at the emission wavelength, when excited at the maximal excitation wavelength, 280 nm.

2.5.3.1. Intrinsic tryptophan fluorescence spectroscopy

Since many proteins contain one or more tryptophanes, this residue can be used to probe conformational changes by exciting the intrinsic tryptophan fluorescence and monitor the emission signal (Schmid, 2005). This is particularly applied for antibody domains due to their highly conserved tryptophan residue in close proximity to the disulfide bond in the hydrophobic core (Feige et al., 2010). Thus, intrinsic tryptophan fluorescence was used to determine local structural changes and to assess the compactness of the investigated proteins. Native intrinsic tryptophan fluorescence spectra were measured with 7 µM protein

in a 1 cm quartz cuvette at 22°C using a FP-6500 spectrofluorometer (Jasco). Spectra were recorded between 300 and 400 nm with an excitation wavelength of 280 nm. Slits were set to 5 nm for both excitation and emission. All spectra were recorded two times, averaged and normalized.

2.5.3.2. Equilibrium chemical unfolding transitions

Chemical equilibrium unfolding transitions were executed as a complementary approach to assess and compare the conformational of the protein samples. For chemical equilibrium unfolding transitions, 5 µM protein samples were incubated with increasing concentrations of guanidium chloride (GdmCl) for at least 12 h in black COC 96 well microplates (#655809, Greiner Bio-One). A 6 M GdmCl stock solution was prepared in PBS pH 7.4 and the exact concentration refractrometrically determined. Intrinsic tryptophan fluorescence spectra of protein samples incubated with either 0 M GdmCl or 4 M GdmCl were recorded the next day as described above. By subtracting the two spectra, the maximal differential emission wavelength was determined, which was set constant for measuring the fluorescence emission intensity of all samples at 22 °C. Fixed wavelength measurements were performed using an Infinite M Nano microplate reader (Tecan). Equilibrium unfolding transitions were analyzed and fit assuming a two-state model as previously described (Equation 8) (Pace, 1986; Santoro and Bolen, 1988):

$$y(D) = y_N^0 + m_N * [D] - \frac{(y_N^0 + m_N * [D]) - (y_U^0 + m_U * T)}{1 + \exp(-\frac{\Delta G_{stab} + m_c * [D]}{R * T})}$$

Equation 8. Two-state model with the assumption of linear dependency of emission from native and unfolded protein. *y(D): fluorescence signal; y: y-intercepts; N: native protein; U: unfolded protein; m: slope of the best fit straight line; m_c: cooperativity; [D]: denaturant concentration (mol/l); T: temperature (K); ΔG_{stab}: free enthalpy of denaturant induced unfolding (kcal/mol); R: universal gas constant (8.314 J mol^{-1} K^{-1})*

2.5.3.3. Acrylamide fluorescence quenching

Quenching of the above described intrinsic tryptophan fluorescence allows the evaluation of the protein dynamics and solvent-exposure of local protein environments containing tryptophan residues (Phillips et al., 1986; Tallmadge et al., 1989). For acrylamide quenching,

increasing concentrations of acrylamide were added to 5 µM protein samples and incubated overnight to assure quenching equilibrium. Spectra were measured as described above. Fixed wavelength measurements were carried out utilizing an Infinite M Nano microplate reader (Tecan) using 280 nm and 338 nm for excitation and emission, respectively. F_0 represents the fluorescence emission intensity without acrylamide. The calculated Stern-Volmer quotient F_0/F, where F is the fluorescence intensity at the respective acrylamide concentration, shows the accessibility of a given protein for acrylamide.

2.5.3.4. Thioflavin T and fibril formation assay

Thioflavin T (ThT) is an extrinsic fluorescent dye to probe for amyloid fibril formation (Gade Malmos et al., 2017). ThT is able to bind to amyloid fibril independent of the predecessor protein and is therefore commonly used to monitor amyloid fibril formation. ThT assays were performed in triplicates in black PS 96 well microplates (#437112, Nunc, ThermoFisher Scientific) and measured with a Genios microplate reader (Tecan) at 440 and 480 nm excitation and emission wavelengths, respectively. To start the assay with defined samples and prevent seed contamination, monomer isolation was performed prior to the ThT assay by loading highly concentrated protein on a Superdex 75 10/300 GL and collecting the monomer peak fraction. Other peaks fraction, dimers as well as higher molecular weight species were discarded. Assays were conducted with 20 µM monomeric protein, 10 µM ThT in PBS buffer pH 7.4 with 0.05 % NaN3 with a final volume of 200 µl per well. Microplates were covered with a Crystal Clear PP sealing foil. Each well was measured every 30 min for 10 to 30 days. Between the measurements, the 96 well microplates were incubated at 37 °C under continuous orbital shaking at 225 rpm within the plate reader. For data analysis, means of the replicates were calculated and individually normalized for better comparison. The kinetic parameter $t_{1/2}$ was determined by ascertain the time (d) when 50 % of the maximal ThT signal amplitude was reached.

2.6. Transmission Electron Microscopy

Transmission electron microscopy (TEM) was performed to assure the presence of amyloid fibrils in case of positive ThT assays and vice versa. TEM is among others, one of the complementary techniques of choice to verify fibril formation since the ThT binding assay is

prone to false positive results (Gade Malmos et al., 2017). For TEM sample preparation, 10 µl samples were taken from the ThT assay wells and placed on a 200-mesh activated copper grid and incubated for 1 min. The samples were washed twice with 10 µl H_2O to prevent phosphate artifacts and negatively stained with 8 µl of a 1.5 % (w/v) uranyl acetate solution for 1 min. Excess solutions on the grid were removed with a filter paper. Micrographs were recorded with a JEM-1400 Plus transmission electron microscope (JEOL) at 120 kV.

2.7. Size Exclusion Chromatography HPLC

Analytical size exclusion chromatography (SEC) was used to determine oligomerization kinetics of IgM, IgA and fusion variants. For oligomerization kinetics, protein samples at defined time points were load on a Superdex 200 10/300 GL column run by a Shimadzu HPLC system at a flow rate of 0.5 ml/min equipped with a SPD-20A UV detector. Prior to the measurement, the column was equilibrated with the elution buffer, PBS pH 7.4 with 0.05 % NaN_3 to prevent bacterial contamination at room temperature. Various concentrations and volumes as indicated were loaded. Oligomerization was monitored by analyzing peak shifts in the time-course chromatograms. Molecular mass of the protein peaks was determined by SEC-MALS as described in chapter 2.7.1.

2.7.1. Size Exclusion Chromatography coupled with Multi-Angle Light Scattering

The equipment of the HPLC system described above with a multi-angle light scattering (MALS) detector enables to determine absolute molar mass moments for protein species at any elution volume of a given chromatogram (Folta-Stogniew, 2005a; Wen et al., 1996). Static light scattering provides information about the hydrodynamic radius of a given protein at a specific elution volume. In parallel, the concentration for the same protein species is determined by either UV absorbance or differential light refraction. Further theoretical considerations and the exact calculation of the absolute molar mass is previously described (Folta-Stogniew, 2005b). A Superdex 75 10/300 GL column connected to a Shimadzu HPLC system equipped with a Dawn Helios II multi-angle light scattering detector, a RID-20A detector and a SPD-20A detector were employed for the determination of the absolute mass of V_L, C_L, LC and C_H2 variants. The column was equilibrated for at least 24 h to obtain stable light scattering baseline signals before data collection. Determination of inter-detector delay volumes, band

broadening correction and light-scattering detector normalization were performed using a standard 1 mg/ml BSA solution run, according to the manufacturer's protocol. 50 µM protein samples (50 µl) were loaded on the column. All experiments were performed at room temperature at a flow rate of 0.45 ml/min in PBS buffer containing 0.05 % NaN3. Data analysis was performed using the ASTRA 5.3.4.20 software according to the manufacturer's protocol.

2.8. Hydrogen/Deuterium Exchange Mass Spectrometry

Hydrogen/Deuterium exchange mass spectrometry (H/DX-MS) is a combined biochemical method which allows to gain information on protein dynamics, structure and interactions (Ahn et al., 2013; Engen et al., 2011). Therefore, accessible hydrogen atoms are exchanged by deuterium upon incubation in a deuterium containing buffer followed by a quantitative averaging mass spectrometry analysis to monitor the peptide fragments of a protein in which H/D exchange was possible. Hydrogen/deuterium exchange mass spectrometry experiments were performed using a fully automated system equipped with a Leap PAL RTC, a Waters ACQUITY M-Class UPLC, a HDX manager, and the Synapt G2-S ESI-TOF mass spectrometer, as described elsewhere (Zhang et al., 2014). Protein samples with a concentration of 30 µM were diluted in a ratio of 1:20 with deuterium oxide containing PBS pH 7.4 and incubated at 20 °C for 10 s, 1 min, 10 min, 30 min and 2 h. To stop the labeling reaction and denature the sample, the labeled protein was diluted 1:1 in H/DX quenching buffer at 1 °C. Digestion was performed on a Waters Enzymate BEH Pepsin Column (2.1 x 30 mm) at 20 °C. Peptides were trapped and subsequently separated on a Waters ACQUITY UPLC BEH C18 column (1.7 µm, 1.0 × 100 mm) with acetonitrile plus 0.1% formic acid (v/v) and H_2O plus 0.1% formic acid (v/v) gradient. Trapping and chromatographic separation were carried out at 0 °C to minimize back-exchange. Eluting peptides were directly subjected to the time-of-flight mass spectrometer by electrospray ionization. Before fragmentation by MSE and mass detection in resolution mode, the peptide ions were additionally separated by drift time within the mobility cell. Data processing was performed using the Waters Protein Lynx Global Server PLGS and DynamX.

2.9. Nuclear Magnetic Resonance Spectroscopy (performed by Dr. Manuel Hora, Matthias Brandl. Tejaswini Pradhan and Prof. Bernd Reif, Department Chemie, TUM)

Nuclear magnetic resonance (NMR) is structural biology method, which relies on the quantum mechanical properties of the nuclei of atoms depending on their microenvironment. In the protein chemistry field, it is used to obtain information about the protein structure and dynamics. NMR experiments in this work were performed in collaboration with the group of Prof. Bernd Reif (TUM). Measurements were performed by Dr. Manuel Hora, Tejaswini Pradhan (C_L, V_L and LC) and Matthias Brandl (C_H2).

^{1}H, ^{13}C, ^{15}N triple resonance spectra were acquired on an AVIII 600 MHZ spectrometer with a cryogenic triple resonance gradient probe. ^{1}H, ^{15}N-HSQC NMR spectra were acquired on a Bruker AVIII 500 MHz or 600 MHz spectrometer. Both spectrometer were equipped with a cryogenic triple resonance gradient probe. The respective ^{15}N- or ^{13}C, ^{15}N-labeled proteins were dissolved in PBS buffer supplemented with 10% D_2O. Backbone assignments were based on three-dimensional HNCA (Grzesiek and Bax, 1992; Kay et al., 1994; Schleucher et al., 1993), HNCACB (Muhandiram and Kay, 1994a; Wittekind and Mueller, 1993) and HN(CO)CACB (Grzesiek and Bax, 1993; Muhandiram and Kay, 1994b) spectra. All spectra were processed using Bruker TopSpin, and analyzed using CcpNmr (Vranken et al., 2005). The protein concentrations were in the range of 60 - 1200 µM for two-dimensional experiments and in the range of 600 - 1200 µM for three-dimensional experiments.

^{1}H-^{15}N chemical shift perturbations (CSPs) for the C_H2 study were calculated using equation 9 (Williamson, 2013):

$$\Delta\delta^{1H,15N} = \sqrt{\frac{1}{2}(\Delta\delta_{1H}^2 + (0.14\Delta\delta_{15N})^2)}$$

$$\text{with } \Delta\delta_{Nuc} = \delta_{Variant\,1} - \delta_{Variant\,2}$$

Equation 9. Calculation of chemical shift perturbations used for the C_H2 study. *$\Delta\delta^{1H,15N}$: chemical shift perturbation; $\Delta\delta_{1H}$: chemical shift differences in ppm for the ^{1}H dimension; $\Delta\delta_{15N}$: chemical shift differences in ppm for the ^{15}N dimension.*

^{13}C chemical shift perturbations relative to a random coil peptide (secondary chemical shifts) were calculated using the difference between the observed chemical shift and the random coil chemical shift for the respective residue type (Wishart and Sykes, 1994).

Chemical shift changes for V_L, C_L and LC variants were calculated using equation 10:

$$\Delta\delta^{\mathrm{res}} = \sqrt{(\Delta\delta^{1\mathrm{H}})^2 + \frac{1}{25}(\Delta\delta^{15\mathrm{N}})^2}$$

Equation 10. Chemical shift changes for the light chain studies. *$\Delta\delta^{res}$: chemical shift chnages; $\Delta\delta^{1H}$: chemical shift differences in ppm for the ^{1}H dimension; $\Delta\delta^{15N:}$ chemical shift differences in ppm for the ^{15}N dimension.*

For measurement of residual dipolar couplings (RDCs), we employed 320 µM protein solutions in a buffer containing 10 mM NaH_2PO_4, 50 mM NaCl, 0.5 mM EDTA and 10 % D_2O at pH 6.5. A first IPAP-type ^{1}H,^{15}N-HSQC experiment was conducted to determine isotropic J couplings, then Pf1 phage (ASLA Biotech, Riga) was added to a final concentration of ~ 8 mg/ml, producing a D_2O splitting of 6.0 Hz for LC wt and 11.9 Hz for LC R108A. Experiments were recorded and processed using TopSpin 3.2 (Bruker BioSpin) and analyzed with CcpNmr analysis 2.4 (Vranken et al., 2005).

2.10. RDC-based ensembles (performed by Prof. Carlo Camilloni, Università degli studi di Milano)

RDCs based ensembles were modelled for LC wt and LC R108A variants making use of Metadynamic Metainference (M&M) (Bonomi et al., 2016a) and the theta-method (Camilloni and Vendruscolo, 2015). M&M enables introducing experimental information into Molecular Dynamics (MD) simulations optimally balancing their information content by means of Bayesian Inference. This allows taking into account at the same time force field inaccuracies as well as the possible sources of error affecting experimental data like random noise, systematic errors etc., leading, at least in principle, to structural ensembles that do not depend on the initial choice of the MD force field and that are in agreement with some given experimental evidence (Löhr et al., 2017).

In M&M multiple molecular dynamics (MD) simulations (replicas) are run in parallel coupled together by a potential that, by using Bayesian inference, optimally balance the information content of the experimental data with that of the MD force field (Bonomi et al., 2016b) while at the same time the overall sampling of the conformational space is enhanced by Metadynamics (Laio and Parrinello, 2002).

MD simulations were performed for LC wt and LC R108A using GROMACS 2016 (Abraham et al., 2015), PLUMED 2 (Tribello et al., 2014) and the ISDB PLUMED module (Bonomi and Camilloni, 2017). The two systems were prepared from the PDB 1FH5 (chain L) (Augustine et al., 2001) applying the following mutations using SCWRL4 (Krivov et al., 2009) (E17D and Y87F in both cases and R108A for the second system). The systems were solvated with ~ 25000 TIP4P 2005 (Abascal and Vega, 2005) water molecules and 2 or 3 Na^+ ions in a dodecahedron box of ~ 765 nm^3. The force field employed was the Amber03W (Best and Mittal, 2010). After an energy minimization followed by a NVT and NPT thermalization at 298 K and 1 atm the systems were simulated in the NPT ensemble (using velocity rescaling (Bussi et al., 2007) and Parrinello-Rahman (Parrinello and Rahman, 1981)) for 16 ns and 16 conformations were extracted as the starting point for the 16 replicas employed in the M&M simulations.

M&M simulations with RDCs were run as described previously (Löhr et al., 2017). A complete setup to reproduce the simulations, including starting conformations, topologies, GROMACS parameters file and PLUMED input files, can be downloaded from GitHub (https://github.com/carlocamilloni/papers-data). Shortly, a preliminary optimization of the relative orientation of the 16 replicas with respect to the Z-axis was obtained by maximizing the correlation between calculated and measured RDCs. The Metainference potential was then turned on making use of a Gaussian noise per data point (Bonomi et al., 2016b) and a scaling factor sampled from a prior uniform distribution (Löhr et al., 2017). The sampling of the conformational space was boost by Metadynamics using Parallel Bias and sharing the bias among the replicas (Bonomi et al., 2016a; Pfaendtner and Bonomi, 2015) using a BIASFACTOR of 20 and an initial energy deposition rate of 2.5 kJ/mol/ps. In particular because the aim is that of sampling the inter-domain dynamics, we selected the following collective variables (CVs) as defined in PLUMED: the 1) ALPHABETA and 2) DIHCOR defined over the psi angles of the linker residues; 3) the DISTANCE between the center of masses of the two domains defined using the Cα atoms; 4) the TORSION angle between the major axis of the two domains; and 5) the DHENERGY (Debye-Hückel energy) between the two domains calculated using the changed side-chains (Do et al., 2013). The width of the Metadynamics bias for the above CVs was set to 0.1, 0.1, 0.05 nm, 0.1, and 0.2 kJ/mol, respectively. For both systems each replica was evolved for 200 ns, for a total nominal time of 3.2 μs of simulation per system studied.

The collected sampling was eventually reweighted taking into account the final bias in the quasi-static regime (Branduardi et al., 2012) so to obtain a statistical ensemble of conformations with each conformer associated to a specific statistical weight.

2.11. Molecular Dynamics Simulations (performed by Maria Daniela Pulido Cendales and Prof. Martin Zacharias, Department Physik, TUM)

Molecular dynamics (MD) simulations were carried out and analyzed by means of the Amber16 simulation package (Case et al., 2016). Free unrestrained MD simulations and umbrella sampling (US) simulations were performed to analyze the stability of the C_H2 antibody domain employing the pmemd.cuda module of Amber16 (Götz et al., 2012). Simulations were performed on the wild type C_H2 as well as the variants C_H2-S and C_H2-SK (all based on the crystal structure PDB: 3HKF). Each protein was solvated in TIP3P water (Jorgensen et al., 1983) in a periodic octahedron box with a minimum distance of protein atoms to the box boundary of 10 Å. The ff14SB force field was employed and Na^+ and Cl^- ions were added to neutralize the system and reach an ion concentration of 0.15 M. Energy minimization of each system was performed with the sander module of Amber16 (1500 minimization cycles). The systems were heated in steps of 100 K (10 ps per step) to a final temperature of 300 K with the solute non-hydrogen atoms harmonically restraint to the start structure. All bonds involving hydrogen atoms were kept at optimal length. In additional 6 steps the harmonic restraints were removed stepwise. For the production simulations hydrogen mass repartitioning (HMR) was employed allowing a time step of 4fs (instead of 2 fs used during heating and equilibration). Unrestrained production simulations were extended to 1 μs for each system. Coordinates were saved every 2 ps. The root-mean-square deviation (RMSD) from the experimental structure and solvent accessible surface area per residue (SASA), using the LCPO algorithm, (Weiser et al., 1999) were calculated using the Amber cpptraj module. The change in mean SASA per residue was calculated as difference of average SASA (per residue) of each variant vs. wild type. It was smoothed by averaging the results over a window of 10 consecutive residues along the whole sequence.

By using the US method, it is possible to efficiently extract free energy changes along a coordinate of interest by forcing it to overcome possible energy barriers. In order to estimate the influence of the C-terminal β-strand segment (residues 94-98) on the folding stability of

the protein we defined the reaction coordinate as the distance d between the centers of mass of the C-terminus (delimited by residues 94 to 98) and the rest of the protein (delimited by residues 1 to 91). The C-terminus of the protein can be dissociated gradually by applying a harmonic penalty potential with the force constant K around a reference distance dref. A total of 25 umbrella windows with different values for K and dref were simulated for each biomolecule to obtain overlapping histograms in d for reliable calculation of the potential of mean force (PMF: free energy change along d). The three sets of generated umbrella windows were: i) 11 consecutive simulations with dref varying between 10.0 Å and 20.0 Å with a step of 1.0 Å and a force constant of K = 2.5 kcal / $Å^2$ mol, ii) 11 consecutive simulations with dref varying between 12.0 Å and 17.0 Å with a step of 0.5 Å and a force constant of K = 3.5 kcal / $Å^2$ mol, iii) 3 consecutive simulations with dref varying between 12.5 Å and 13.5 Å with a step of 0.5 Å and a force constant of K = 5.0 kcal / $Å^2$ mol. In each umbrella window simulations of 50 ns at 300 K were performed (time step 2 fs), whereby positional restraints with a force constant of Kpos = 0.02 kcal / Å2 mol were applied to non-hydrogen atoms in residues 1 to 91. A PMF was along the reaction coordinate d was calculated using the WHAM algorithm (Kumar et al., 1992).

3. The antibody light chain linker is an important regulator for the orientation and stability of the two constituent domains

As described in the introduction, it is largely unknown how the C_L domain and the LC linker affect the amyloid fibril formation propensity of amyloidogenic V_L variants in the context of the full-length LC. To address this question, the linker connecting the two individual domains of the MAK33 kappa LC was the main subject of this study. Since the presence of the LC linker residue R108 is important for both, V_L and C_L integrity (Nokwe et al., 2015), Arg108 was substituted by either alanine (LC R108A) or glutamate (LC R108E) to test the impact of this residue in the context of the LC. In addition, variants with an additional glycine (LC link+G) or a glycine-alanine spacer (LC link+GAGAG) between Arg108 and Ala109 were created to allow a higher degree of LC linker flexibility and assess the importance of the conservation of the kappa LC linker. It was not possible to completely delete the linker as this variant aggregated quantitatively during refolding and purification (variant LC Δlink). The aligned sequences of MAK33 V_L, LC and C_L, and the crystal structure of the MAK33 kappa LC (PDB: 1FH5, chain L) are provided in Figure 7 (Augustine et al., 2001). Noteworthy, the LC structure derived from the structure of the entire Fab fragment, which comprises the LC complexed with the heavy chain domains V_H and C_H1 (PDB: 1FH5, chain H). Major parts of this chapter were published in parallel to the preparation and acceptance of this thesis in the Journal of Molecular Biology (Weber et al., 2018a).

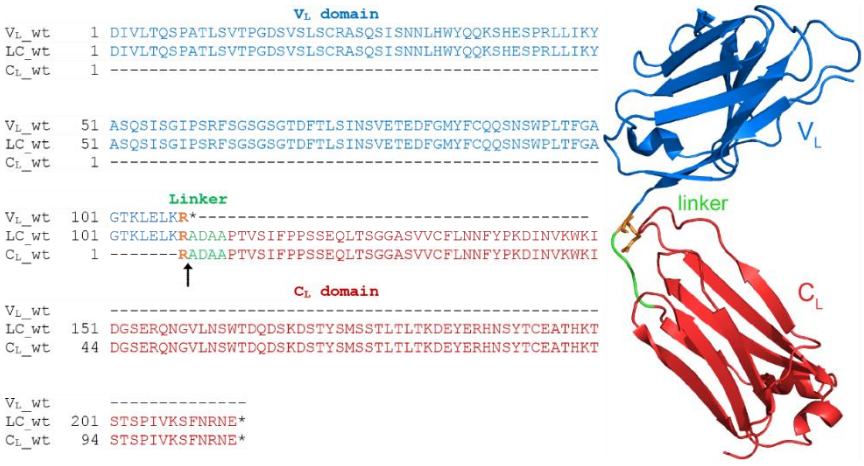

Figure 7. Sequence alignment and crystal structure of the MAK33 kappa LC. *Alignment of the amino acid sequences of V_L wt, LC wt and C_L wt and the crystal structure of the MAK33 light chain (PDB: 1FH5, chain L) indicating the V_L domain (blue), the light chain linker (green), the C_L domain (red) and residue R108 (orange). Arg108 was mutated to alanine or glutamate. Additional linker residues (LC link+G, LC link+GAGAG) were introduced between Arg108 and Ala109 (arrow) to assess the importance of the LC linker.*

3.1. Results

Far-UV and near-UV CD spectroscopy was performed to characterize the secondary and tertiary structures of the V_L, LC and C_L variants. Far-UV CD measurements (Figure 8A) revealed a characteristic β-strand-rich secondary structure for all variants indicated by a minimum at 218 nm. The recorded CD spectra match the respective wild type spectra. Near-UV CD experiments (Figure 8B) were used to assess the tertiary structure. All spectra showed similar ellipticity characterized by a minimum at ~ 275 nm. Only minor signal changes between the variants were observed, indicating a similar tertiary structure. Thus, the introduced mutations did not alter the typical Ig fold, confirming a generally non-disruptive character of the variants, which was the basis to perform further analysis.

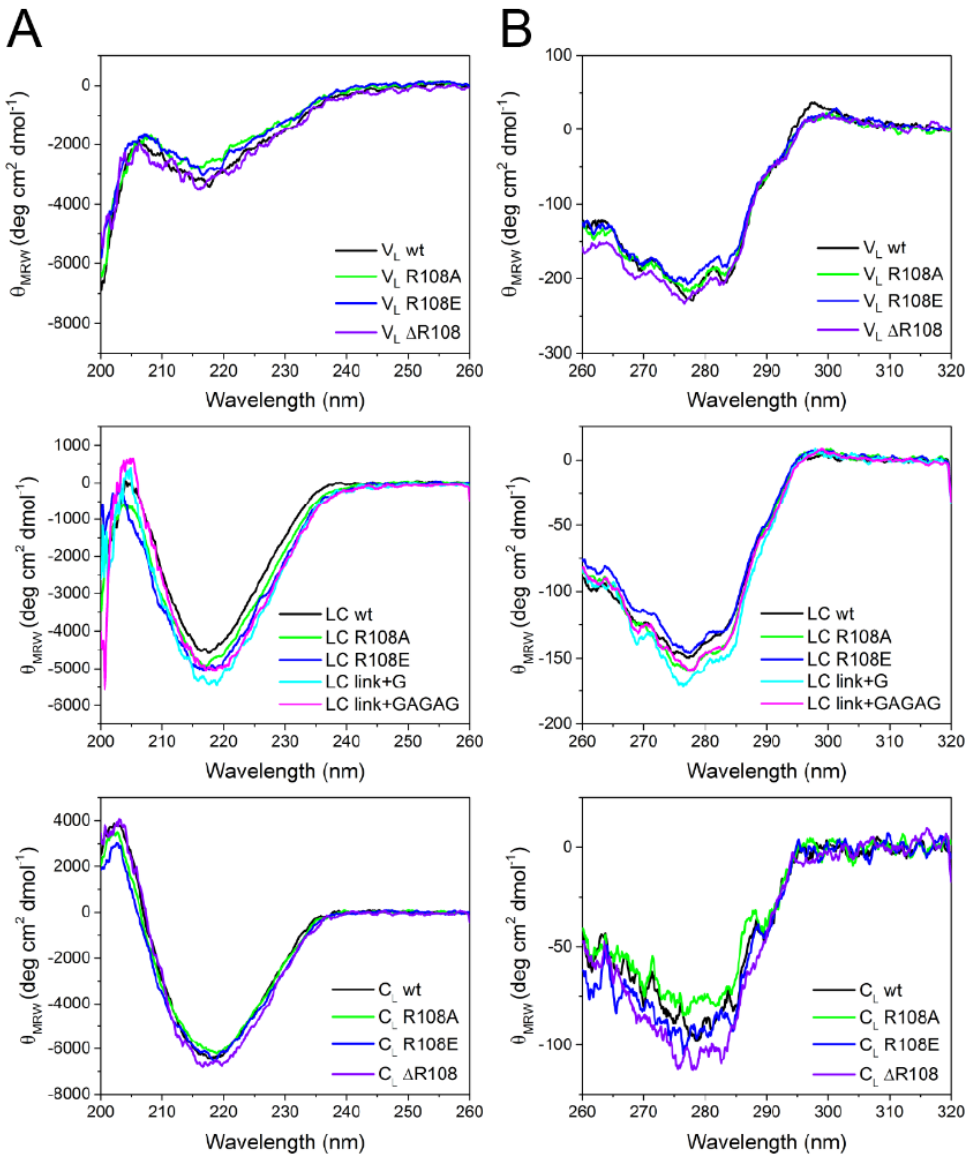

Figure 8. Secondary and tertiary structure analysis. *Far-UV (A) and near-UV (B) CD spectra of V_L (top), LC (middle) and C_L (bottom) variants. The introduced mutations did not significantly alter the secondary or tertiary structure of the variants with respect to their wild type protein.*

SEC-MALS allows the determination of absolute molar masses for isolated protein species. In this study, the technique was used to assess the oligomeric state, i.e. the quaternary structure of the variants (Figure 9). All V_L, LC and C_L variants were monomeric (Table 2). For V_L wt, a minor fraction of larger species, most likely dimers as judged from the elution volume, was detected. This fraction is represented by a shoulder of the main peak. Due to a very low concentration of this species and the poor separation from the main peak, it was not possible to determine the exact molecular mass of the shoulder peak.

In summary, the quaternary structure of V_L, LC and C_L is not affected by mutating R108 or extending the LC linker. Consequently, the data show that the introduced mutations did not alter the native structure of the variants.

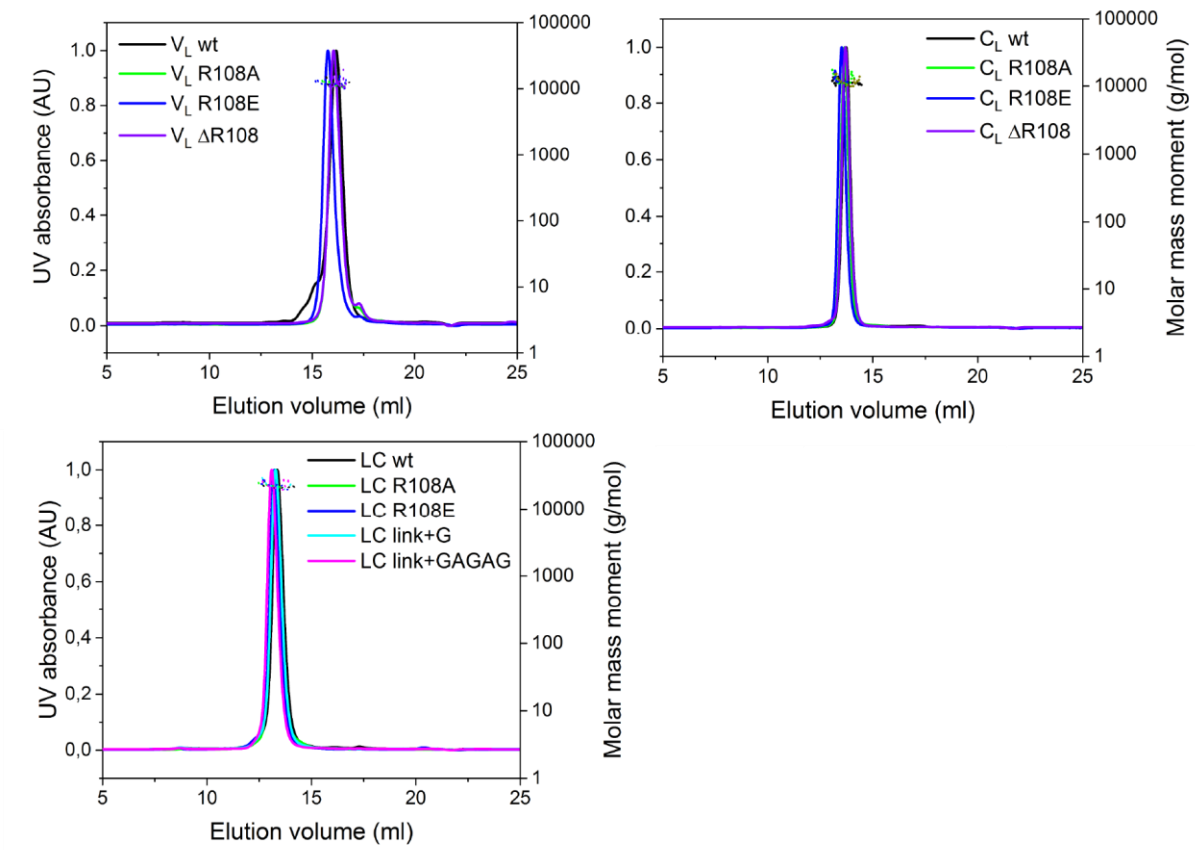

Figure 9. Quaternary structure analysis. *SEC-MALS chromatograms of all V_L (top left), LC (bottom) and C_L (top right) variants show the monomer as the predominant species. Only V_L wt shows a minor fraction of higher molecular weight species represented by a small shoulder of the main peak. 50 µl of 50 µM samples were injected.*

Variant	Molecular mass (kDa)		
	V_L	LC	C_L
wt	11.7	20.9	11.2
R108A	11.9	22.9	11.5
R108E	12.6	23.4	12.6
ΔR108	11.5	N/A	10.6
link+G	N/A	21.2	N/A
link+GAGAG	N/A	21.3	N/A

Table 2. SEC-MALS-derived molecular masses. *Molecular masses in kDa of the predominant peaks of the size-exclusion chromatograms as determined by SEC-MALS. The calculated molecular masses using the ProtParam tool are 11.8 kDa for V_L wt, 23.4 kDa for LC wt and 11.8 kDa for C_L wt. All variants were monomeric. N/A: construct does not exist.*

To assess the impact of the LC linker and the presence of the C_L domain in the full-length LC context, the conformational stability of the V_L, LC and C_L variants was determined by CD thermal unfolding transitions (Figure 10 and Table 3). V_L wt and V_L R108E are the most stable V_L variants exhibiting melting temperatures (T_ms) of 51.8 ± 0.3 °C and 51.9 ± 0.8 °C, respectively. V_L R108A (T_m 50.8 ± 0.6 °C) revealed a slightly decreased conformational stability by ~ 1 °C. The Arg108 deletion variant V_L ΔR108 (T_m 48.2 ± 0.6 °C) showed the lowest T_m among the V_L variants, which is in agreement with the literature (Nokwe et al., 2015). C_L wt (T_m 53.8 ± 0.5 °C) and C_L R108A (T_m 53.7 ± 0.2 °C) are the most stable variants among the investigated proteins. The R108E (T_m 51.5 ± 0.3 °C) mutation in C_L decreased the stability by ~ 2 °C, which is more pronounced than the impact of the destabilizing R108A mutation in V_L (decrease of ~ 1 °C). Similar to the effect on the V_L variant, the deletion of Arg108 in C_L (T_m 50.2 ± 0.4 °C) results in a ~ 3 °C lower conformational stability. Thus, for both individual domains, the R108 deletion variants, V_L ΔR108 and C_L ΔR108, reveal the highest decrease in conformational stability. This is in agreement with a study by Nokwe et al. showing that Arg108 is important for the integrity of both, V_L and C_L (Nokwe et al., 2015).

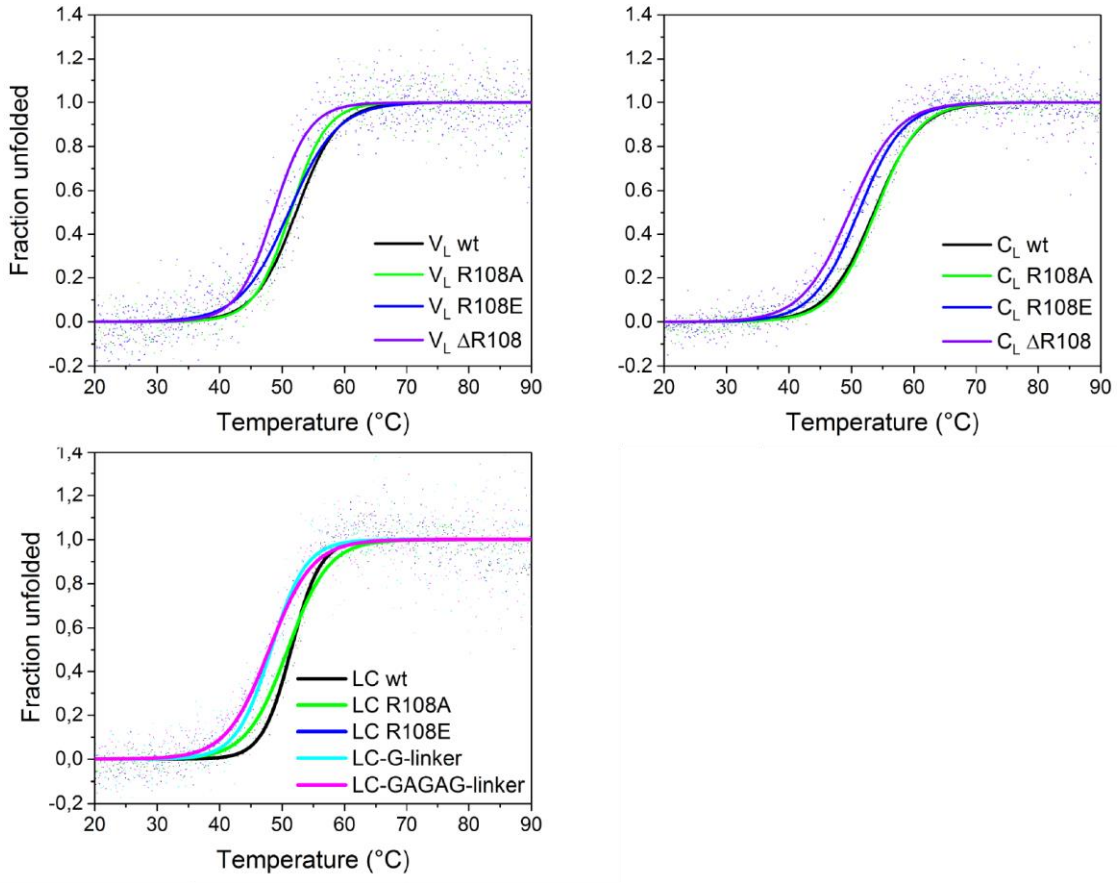

Figure 10. Conformational stability. *Thermal unfolding transitions followed by CD spectroscopy reveal a distinct impact of certain mutations on V_L (top left), LC (bottom) and C_L (top right). The heating rate was 0.5 °C/min at a constant wavelength of 205 nm. Solid lines represent the Boltzmann fit of the monitored data to determine the melting temperature (T_m).*

	Melting temperature (°C)		
Variant	**V_L**	**LC**	**C_L**
wt	51.8 ± 0.3	50.8 ± 0.7	53.8 ± 0.5
R108A	50.8 ± 0.6	49.7 ± 0.9	53.7 ± 0.2
R108E	51.9 ± 0.8	47.6 ± 0.4	51.5 ± 0.3
ΔR108	48.2 ± 0.6	N/A	50.2 ± 0.4
link+G	N/A	49.2 ± 0.7	N/A
link+GAGAG	N/A	48.5 ± 0.4	N/A

Table 3. Conformational stability. *Melting temperatures (T_m) of all V_L, LC and C_L variants as determined by thermal unfolding transitions. The heating rate was 0.5 °C/min. The error is the standard deviation of three technical replicates. N/A: construct does not exist.*

For the full-length LC variants, comparable results were obtained. Like for the V_L and C_L domains, LC wt (T_m 50.8 ± 0.7 °C) showed the highest stability among the LC variants. LC R108A (T_m 49.7 ± 0.9 °C) as well as the extension variants LC link+G (T_m 49.2 ± 0.7 °C) and LC link+GAGAG (T_m 48.5 ± 0.4 °C) are slightly less stable than the wild type, but more stable than LC R108E (T_m 47.6 ± 0.4 °C). Interestingly, this suggests that the dislocation of Arg108 relative to the C_L domain by introducing a glycine (LC link+G) or a glycine-alanine spacer (LC link+GAGAG) has the same impact as the substitution of Arg108 by alanine (LC R108A). This is to be expected as the additional residues were introduced between Arg108 and Ala109, resulting in a spatial separation of R108 and the C_L domain. It was not possible to purify the LC linker deletion variant (LC Δlink) due to quantitative protein aggregation during protein refolding. As the residues next in sequence to Arg108 are Ala109, Asp110, Ala111 and Ala112, a LC ΔR108 variant would possibly result in the same degree of destabilization as LC R108A since the LC linker comprises an alanine-rich cluster. The data show that mutating the important linker residue Arg108 to alanine has no significant impact on protein stability, neither for the individual domains V_L and C_L, nor for the full-length LC. The decrease of T_m is comparably low for V_L and LC (both decreased by ~ 1 °C), suggesting that the R108A mutation is causing a conformational destabilization within V_L, but is not affecting C_L. Interestingly, the substitution of Arg108 by glutamate resulted in a T_m decrease of ~ 2.5 to 3 °C for both, LC R108E (T_m 47.6 ± 0.4 °C) and the C_L R108E variant (T_m 51.5 ± 0.3 °C). In contrast, V_L R108E (T_m 51.9 ± 0.8 °C) is completely unaffected by this mutation. Thus, deleting the positive charge of Arg108 (R108A) affects exclusively V_L, whereas the introduction of a negative charge (R108E) affects the C_L domain. Changes in the length of the LC linker (LC link+G, LC link+GAGAG) slightly decreased the stability of the LC, similar to the R108A mutation. Arg108 substitutions by alanine or glutamate (LC R108A, LC R108E), however, did not uniformly impact the conformational stability of the LC. Interestingly, the data show that the stability of the LC is generally lower than the stability of the individual V_L and C_L or the average T_m of both domains (Table 3). This is true for the wild type proteins as well as for the variants. For LC R108E, the decrease in conformational stability can be attributed to the individual C_L domain, whereas the slightly decreased stability of LC R108A is likely caused by the V_L domain.

The fluorescent dye thioflavin T (ThT), was utilized as a probe to monitor amyloid fibril formation (Figure 11A). Samples showing a positive ThT signal were further analyzed by transmission electron microscopy (TEM) (Figure 11B). This procedure is necessary as ThT is known to be prone to false positive results, however, it is still the gold standard among amyloid dyes (Gade Malmos et al., 2017). For all variants with a positive ThT signal, the TEM analysis revealed fibrillar structures. For comparison of the kinetic profiles, the time point when 50 % of the maximal ThT signal (t_{50}) was determined after normalization of the ThT fluorescence intensity (Table 4). After 15 days of incubation under gentle shaking in PBS pH 7.4 buffer at 37°C, all V_L variants formed fibrils albeit with different lag times (Figure 11A). Both, V_L R108A (t_{50} 6.99 ± 0.08 d) and V_L R108E (t_{50} 6.85 ± 0.20 d) exhibited comparable t_{50} values and the fastest fibril formation kinetics. Surprisingly, the Arg108 deletion variant V_L ΔR108 (t_{50} 7.52 ± 0.45 d) formed fibrils only slightly slower than the R108 substitution variants V_L R108A and V_L R108E, although V_L ΔR108 exhibited a significantly lower conformational stability. Fibril formation of V_L wt (t_{50} 9.58 ± 0.47 d) was ~ 2.5 to 2 d slower than that of the Arg108 substitution and the deletion variants, respectively. All V_L variants revealed different fibril morphologies and degrees of fragmentation as observed by TEM. Particularly, TEM micrographs of V_L ΔR108 fibrils showed strongly fragmented fibrillar clusters. In contrast, no fibril formation was observed for the C_L variants, including C_L wt, C_L R108A and C_L R108E (Figure 11A), which was confirmed by TEM analysis (not shown). Even the destabilized variant C_L ΔR108 was resistant to amyloid formation under the given conditions. This indicates that C_L wt is intrinsically resistant against fibril formation, which is in agreement with the literature (Feige et al., 2008). To test how the presence of the C_L domain affects mutations shown to enhance V_L amyloidogenicity, ThT assays with the full-length LC variants carrying the respective mutations (R108A, R108E) were performed (Figure 11A). Interestingly, none of the LC variants showed ThT binding or fibrils in the TEM analysis. In contrast to the V_L wt, the LC wt is resilient against fibril formation under the experimental conditions. Although the mutation R108E decreases the T_m of the LC, it did not render the LC amyloidogenic, comparable to the respective C_L variants. LC variants with an extended LC linker, LC link+G and LC link+GAGAG, did not form fibrils. Moreover, destabilizing mutations in the linker did not have a negative impact on the integrity of C_L and LC. In the full-length LC context, the C_L domain retains its resistance against amyloid fibril formation and expands it to V_L via the LC linker even in the presence of unfavorable, destabilizing mutations. Moreover, extending the LC linker by

up to five non-polar amino acids, does not decrease the integrity of the LC, although the conformational stability is slightly affected.

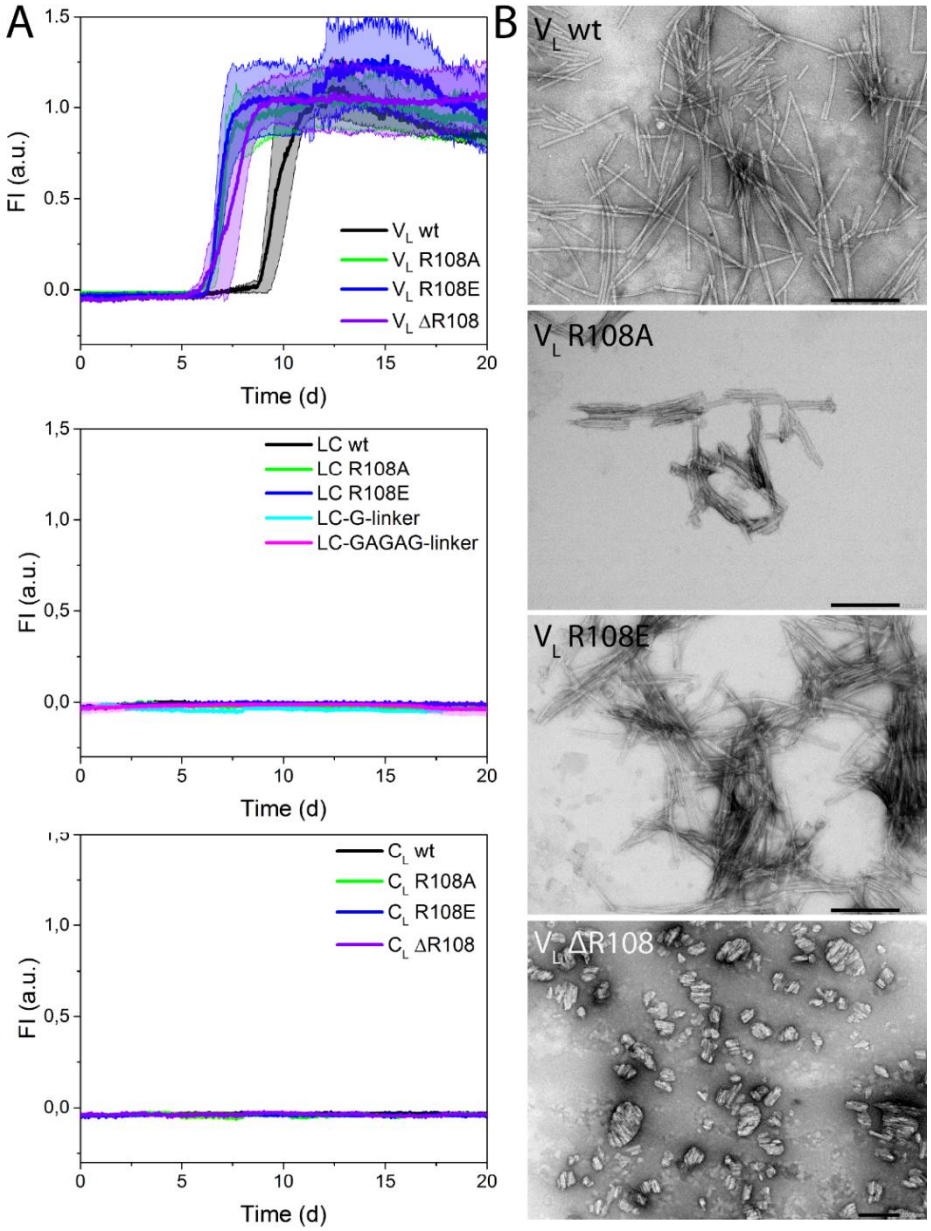

Figure 11. Amyloid fibril formation propensity. A) *ThT fibril formation kinetic plot of the V_L (top), LC (middle) and C_L (bottom) variants. All exhibited altered fibril formation propensity upon mutating Arg108. "Shadows" indicate standard deviation of at least two technical replicates. For better kinetic comparison, the fibril formation curves were individually normalized.* ***B)*** *TEM micrographs of ThT-positive samples. TEM micrographs of V_L wt, V_L R108A and V_L R108E were obtained using 50,000 x magnification, while V_L ΔR108 was observed at 30,000 x magnification. The bar represents 200 nm.*

Variant	t_{50} (days)		
	V_L	LC	C_L
wt	9.58 ± 0.47	no fibrils	no fibrils
R108A	6.99 ± 0.08	no fibrils	no fibrils
R108E	6.85 ± 0.20	no fibrils	no fibrils
ΔR108	7.52 ± 0.45	N/A	no fibrils
link+G	N/A	no fibrils	N/A
link+GAGAG	N/A	no fibrils	N/A

Table 4. Fibril formation kinetics. t_{50} values represent the time point when 50 % of the total fibril formation was reached. Data was taken from ThT fibril formation kinetics. The error is the standard deviation of two technical replicates. N/A: construct does not exist.

In order to gain a structural understanding of the results, NMR experiments were conducted by Dr. Manuel Hora in the group of Prof. Bernd Reif, TUM. All variants produced spectra indicative of folded proteins. Despite some minor chemical shift changes, all spectra were similar enough to confirm that the variants adopt the characteristic Ig fold (Figure 12). The largest chemical shift differences between the V_L wt and LC wt NMR spectra were observed for Thr80 and Gly84 (Figure 13A). These residues also displayed considerable chemical shift changes in the mutants LC R108A and LC R108E. Comparison of C_L wt and LC wt showed a prominent shift difference for Thr172. In a previous study, Nokwe et al. found two distinct conformations for C_L wt, which resulted in two sets of NMR resonances (Nokwe et al., 2015). It was shown that the R108 deletion variant C_L ΔR108 caused a shift to the weakly populated C_L wt conformer (also present in Figure 13). In LC wt, T172 yields a significant chemical shift difference in comparison to the C_L wt major peak. Still, the shift was much closer to the major population peak than to the minor population. This is supported by the LC wt Asn138 peak being very close to the main conformation of C_L wt. Similar observations were made for LC R108A. While two weak signals for Asn138 were observed in this case, the position of the Thr172 peak indicates a conformation closer to the major state of C_L wt. In contrast, LC R108E exhibits chemical shifts for Asn138 and Thr172, which are in good agreement with the weakly populated conformation of C_L wt (Figure 14).

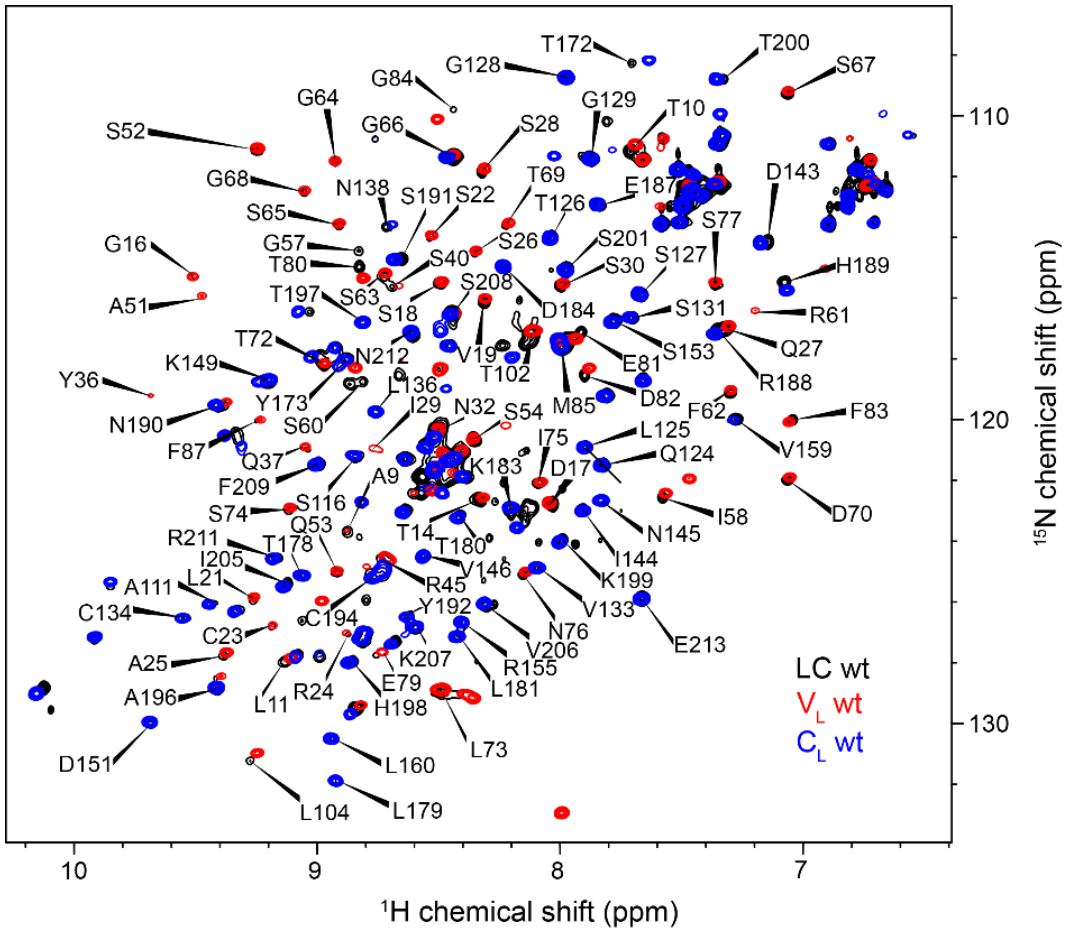

Figure 12. $^{1}H,^{15}N$-HSQC spectra. *Spectra of V_L wt (red), LC wt (black) and C_L wt (blue) reveal similar chemical shifts for the residues of the individual domains, V_L and C_L, compared with the full-length LC. The assignment of the residues to the respective peak signals was previously performed and utilized for the analysis of variants in this study (Nokwe et al., 2015). Data recorded and processed by Dr. Manuel Hora (group of Prof. Bernd Reif, TUM).*

In summary, the conformations of LC wt and LC R108A are more similar to the major population, albeit with some differences, while the R108E mutation caused a similar conformational change as the C_L deletion ΔR108. Due to their close structural proximity, R108 likely interacts with Thr172 or Asp170 in LC wt, providing stabilizing contacts (Figure 13C). However, the latter residue (Asp170) was not assigned in the NMR spectra. In the case of LC R108A, these favorable interactions are absent. This would result in an increased flexibility of the Asp170 loop and a relaxation of the rigid orientation between both domains. In LC R108E, the attractive interactions might be replaced by repelling forces, thereby pushing the two individual domains apart.

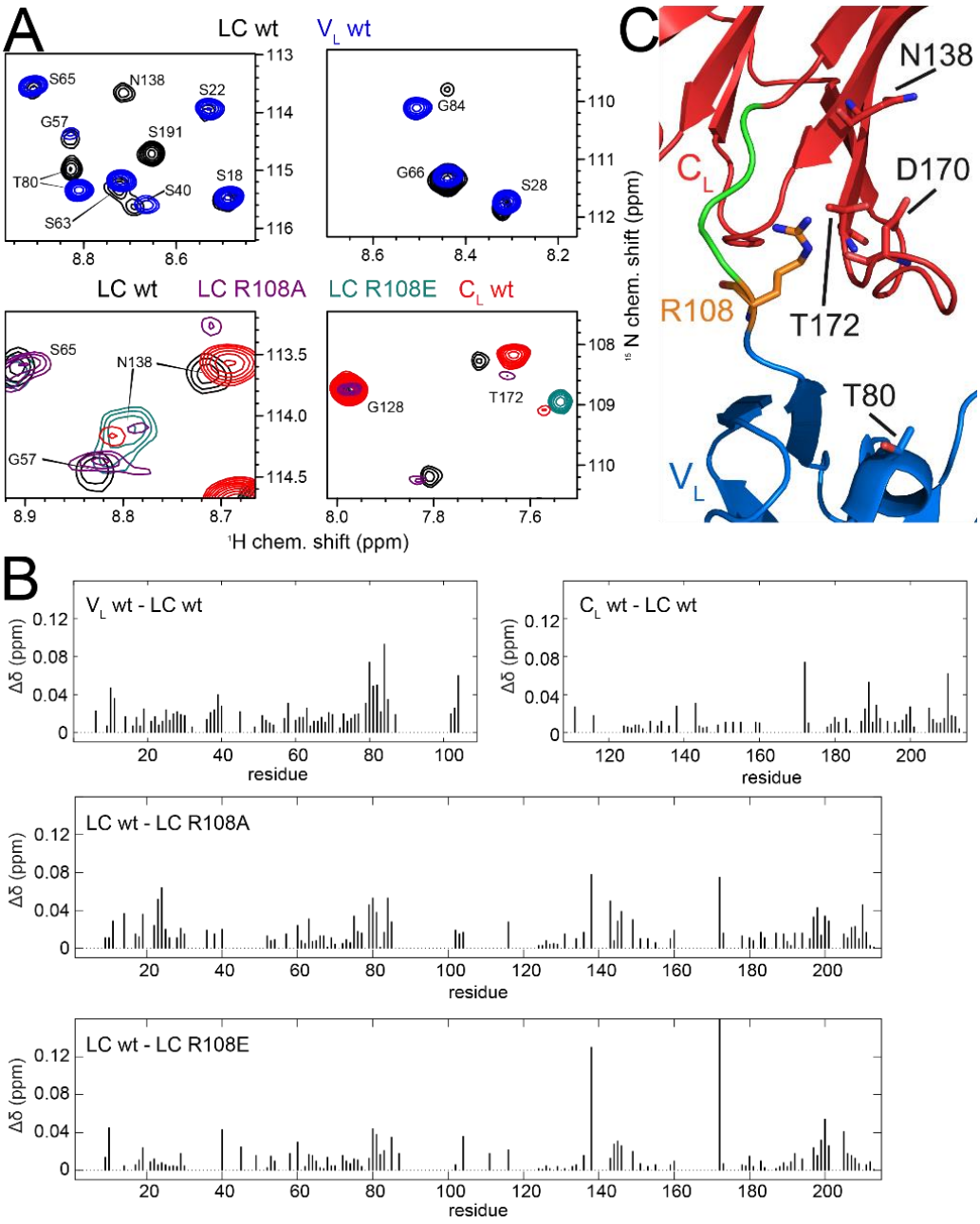

Figure 13. NMR spectra and chemical shift changes. A) *Top: Chemical shift differences between LC wt (black) and V_L wt (blue). Bottom: Chemical shifts differences between LC wt (black), LC R108A (purple), LC R108E (green) and C_L wt (red). Nokwe et al. observed previously that C_L wt undergoes slow chemical exchange between two conformations. In the spectra, this exchange is shown for residues Asn138 and Thr172, with a major and a minor population. LC wt Asn138 and Thr172 displayed similar chemical shifts as the major population of C_L wt. For Asn138 of LC R108A, two weak peaks were visible, however they were close to the noise. Residue Thr172 of LC R108A indicated this variant to adopt a conformation similar to the major population of C_L wt. In contrast, LC R108E produced chemical shifts close to those of the minor population of C_L wt.* ***B)*** *Chemical shift differences between V_L, C_L and LC constructs. For C_L wt, the major populations for residues Asn138 and Thr172 were used to calculate chemical shift differences.* ***C)*** *Structural representation of the LC domain interface. In addition to the LC linker residue Arg108 (orange, other linker residues in green), important residues in V_L (Thr80) and C_L (Asn138, Asp170, Thr172) are highlighted. Data recorded and processed by Dr. Manuel Hora (group of Prof. Bernd Reif, TUM).*

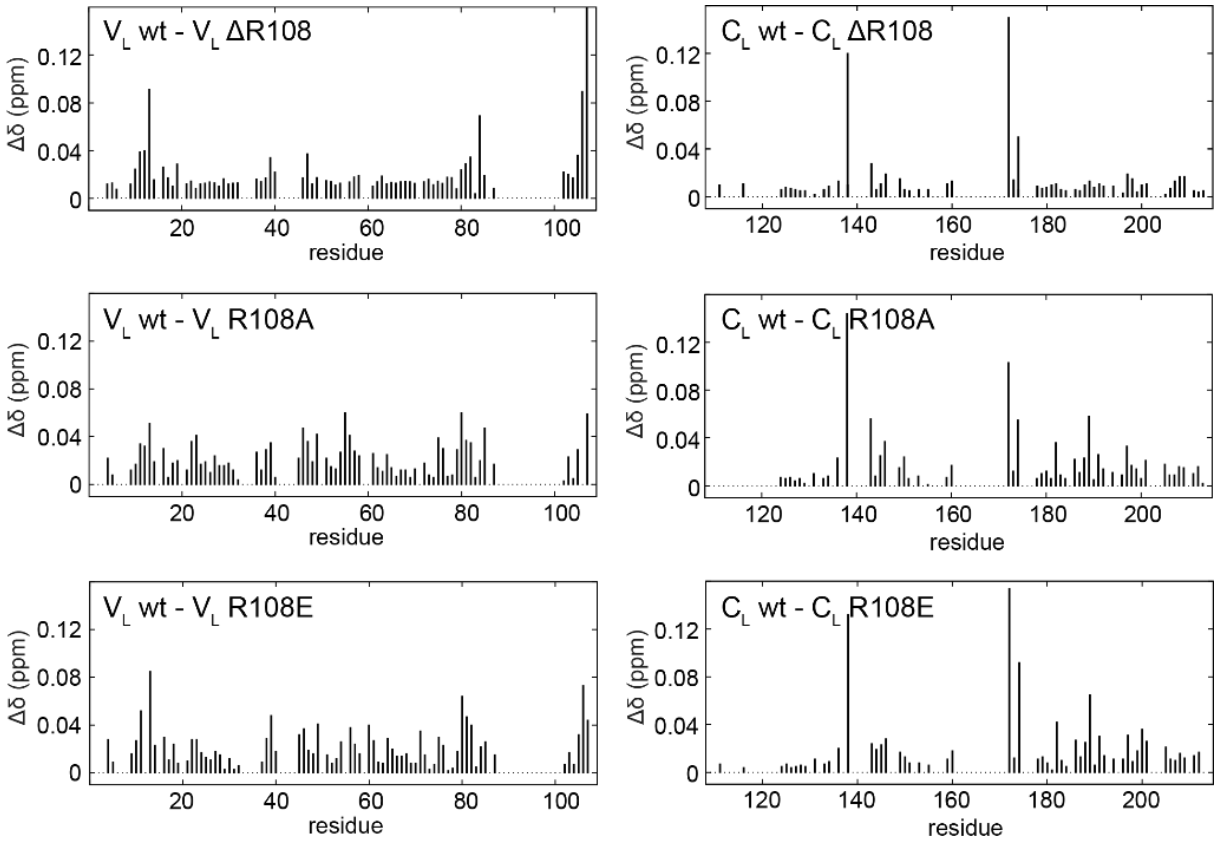

Figure 14. Chemical shift differences between V_L (left) and C_L (right) variants and the respective wild type. *While the R108A and R108E mutations in V_L cause comparable chemical shifts differences, the deletion of R108 (V_L ΔR108) leads to more significant chemical shift changes in the C-terminus of V_L. In contrast, substitution and deletion of Arg108 cause similar chemical shift changes in C_L particularly impacting Asn138, Asp170 and Thr172. For the C_L variants, the major populations for residues asn138 and Thr172 were used to calculate chemical shift differences. Data recorded and processed by Dr. Manuel Hora (group of Prof. Bernd Reif, TUM).*

In order to confirm the hypotheses about different domain orientations affected by the linker residues, residual dipolar couplings (RDCs) were measured. These NMR parameters report on relative orientations of nuclear spins to an alignment medium (Chen and Tjandra, 2012). In total, 67 RDCs for LC wt and 63 RDCs for LC R108A could be determined (Figure 15 and Table 5). The experiments yield different RDC values for LC wt and LC R108A. Since chemical shift differences between these variants were rather small, it was considered unlikely that the different RDCs result from changes in the domain structures. Instead, the RDCs report most likely on different relative orientations of C_L and V_L with respect to each other.

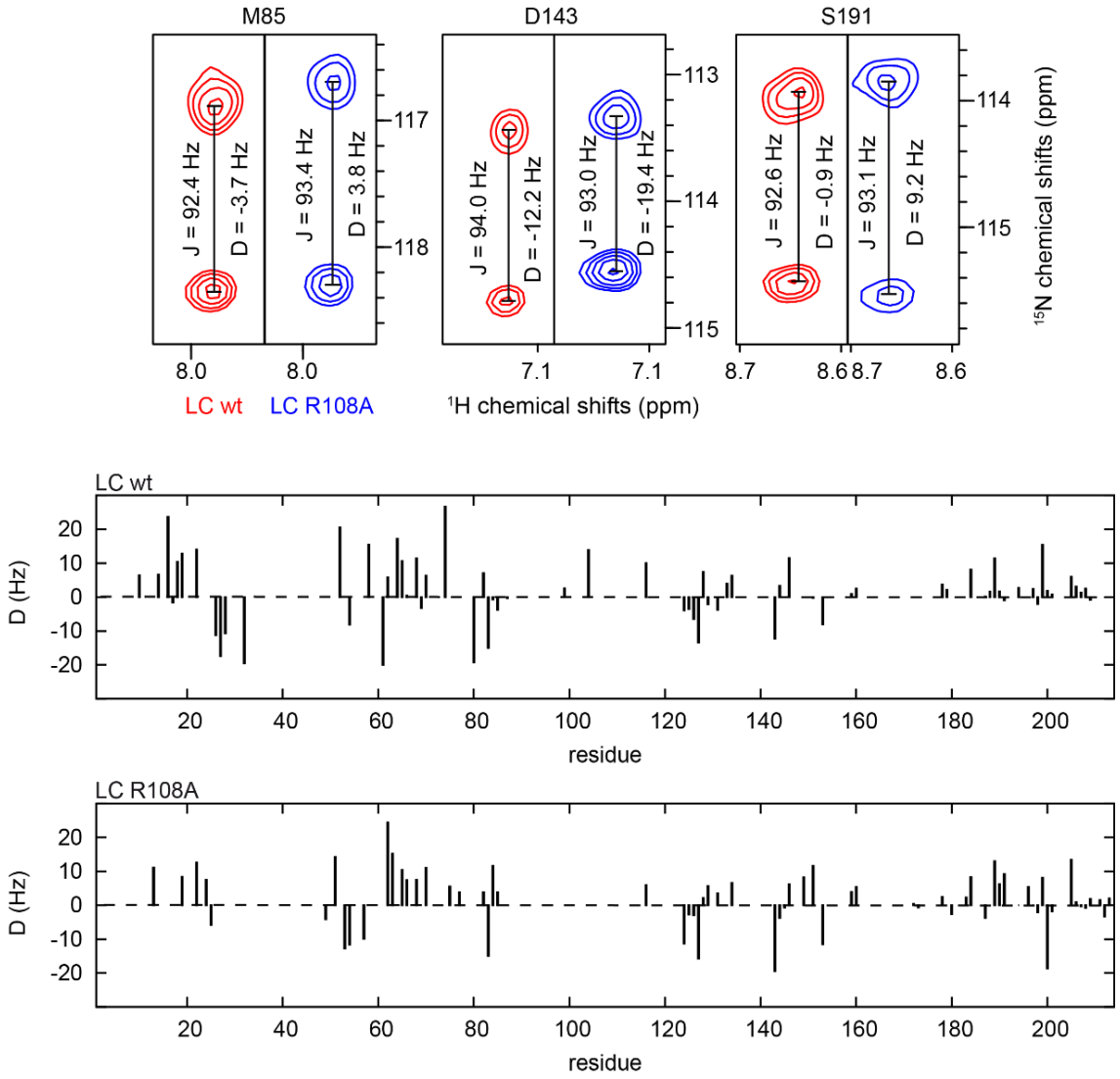

Figure 15. Residual dipolar couplings spectra for LC wt and LC R108A. *The size of the doublet splitting of the cross peak is dependent on the sum of J (scalar coupling, ~ 92 Hz for $^{1}J_{HN}$) and D (dipolar coupling, dependent on orientation). Subtraction of ^{15}N chemical shift differences measured with and without alignment medium yields the RDC values. RDCs were measured for residues along the whole sequence, yielding information about the relative orientation of the V_L and C_L domain. Data recorded and processed by Dr. Manuel Hora (group of Prof. Bernd Reif, TUM).*

	LC wt	LC R108A		LC wt	LC R108A		LC wt	LC R108A
residue	D (Hz)		residue	D (Hz)		residue	D (Hz)	
10	6.4		**70**	6.3	11.0	**159**	0.9	3.9
13		11.1	**74**	26.7		**160**	2.5	5.4
14	6.6		**75**		5.5	**172**		0.4
16	23.6		**77**		3.8	**173**		-0.6
17	-1.6		**80**	-19.2		**178**	3.7	2.4
18	10.4		**82**	7.0	3.8	**179**	2.1	
19	12.8	8.4	**83**	-15.0	-14.9	**180**		-2.6
22	14.0	12.6	**84**	-0.7	11.6	**183**		2.3
24		7.5	**85**	-3.7	3.8	**184**	8.1	8.3
25		-5.8	**87**	-0.3		**187**	0.1	-3.7
26	-11.2		**99**	2.6		**188**	1.6	
27	-17.4		**104**	13.9		**189**	11.4	13.0
28	-10.6		**116**	10.0	5.9	**190**	1.7	6.2
32	-19.5		**124**	-3.9	-11.3	**191**	-0.9	9.2
49		-4.1	**125**	-3.5	-2.7	**194**	2.7	
51		14.2	**126**	-6.5	-2.9	**196**		5.4
52	20.5		**127**	-13.4	-15.7	**197**	2.4	
53		-12.8	**128**	7.4	2.1	**198**	-2.0	-2.0
54	-8.1	-11.6	**129**	-2.1	5.6	**199**	15.4	8.1
57		-9.9	**131**	-3.7	3.5	**200**	1.8	-18.6
58	15.5		**133**	3.9		**201**	0.7	-1.8
61	-20.0		**134**	6.3	6.6	**205**	6.0	13.4
62	5.8	24.4	**143**	-12.2	-19.4	**206**	3.1	0.9
63		15.2	**144**	3.3	-3.7	**207**	1.3	-0.3
64	17.2		**145**		-0.7	**208**	2.5	-0.8
65	10.6	10.4	**146**	11.5	6.1	**209**	-0.8	1.8
66	0.4	7.4	**149**		8.2	**211**		1.5
68	11.4	7.5	**151**	-0.2	11.6	**212**		-3.3
69	-3.2		**153**	-8.0	-11.5	**213**		2.0

Table 5. Residual dipolar couplings values for LC wt and LC R108A for the respective residue. *RDCs were measured for residues along the whole sequence, yielding information about the relative orientation of the VL and CL domain. Data recorded and processed by Dr. Manuel Hora (group of Prof. Bernd Reif, TUM).*

To obtain a structural interpretation of the effect of the R108A mutation on the dynamics of the two LC domains, Metadynamic Metainference (M&M) simulations and the theta-method for the interpretation of the experimentally derived RDCs were utilized and performed by Prof. Carlo Camilloni, University of Milano. Two M&M simulations were performed to compare the conformational space accessible for LC wt and LC R108A. The two resulting conformational ensembles show that the single domains maintain a stable fold (in both cases we have average RMSDs of ~ 1.4 and ~ 1.8 Å for the backbone with respect to the crystal structure of V_L and C_L, respectively), while showing marked differences in the global dynamics (Figure 16). The free energy surfaces are represented as a function of the relative orientation of the two domains defined by the two angles phi and psi as shown in the cartoon representation Figure 16 bottom). Remarkably, while in LC wt and LC R108A the two domains can fully rotate around the linker (phi angle), the R108A mutation allows the two domains to adopt more open orientations to each other (psi angle), however with a preference for relative close orientation (psi 1.5 - 2.5 rad). For the phi angle, both systems prefer angles between -2 and 0 rad, as well as around 1 rad for LC wt. Of notice, the X-ray structure falls into the low free-energy region, adopting psi and phi angles of ~ 1.7 and -1, respectively (Figure 16, yellow square). From a structural perspective, R108 is able to form a salt-bridge with Asp170 (absent in the crystal structure) that is 75 % populated in the ensemble, while the R108A substitution increases the average distance between the residues Ala108 and Asp170 from 5 Å to 8 Å. Consequently, a contact between these two residues is only 20 % populated in LC R108A. In contrast, the salt-bridge Lys103 - Asp165 formed in the crystal structure is only weakly populated in the solution ensembles (2 % and 8 % for LC wt and LC R108A, respectively).

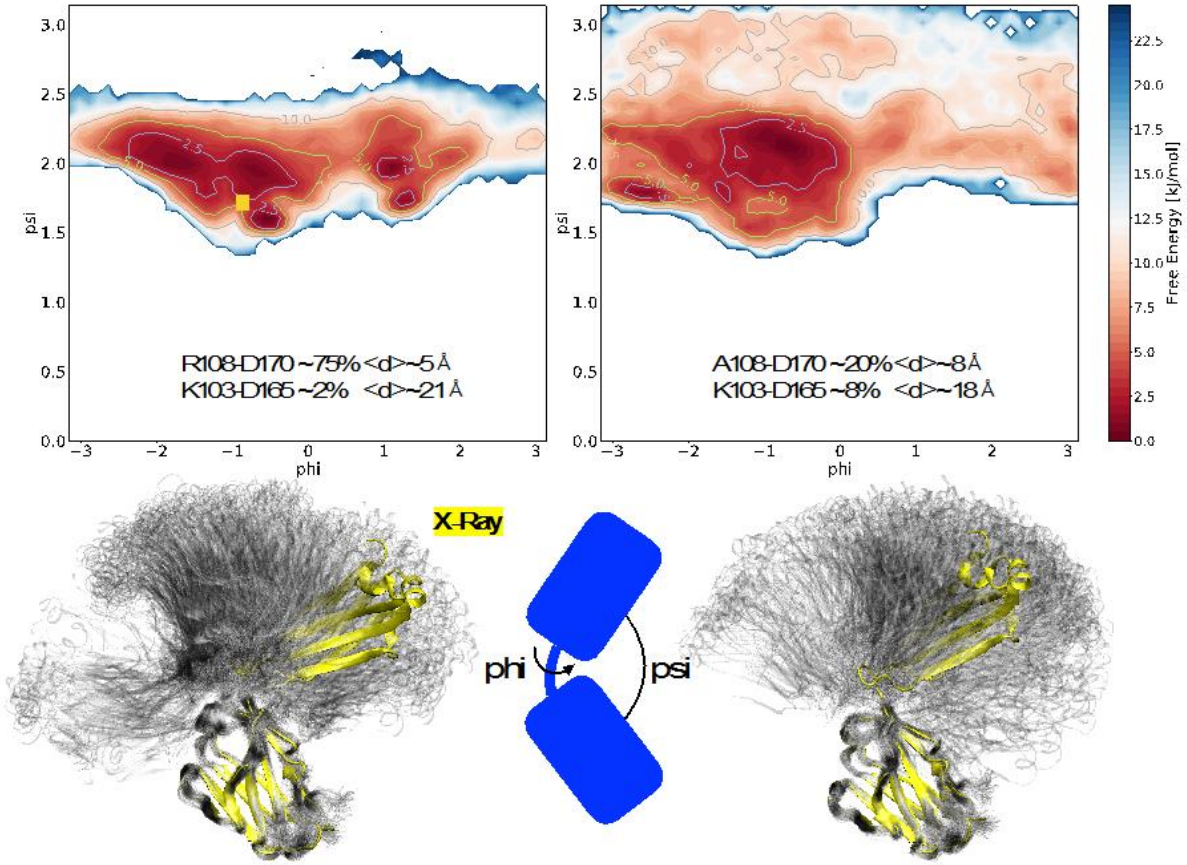

Figure 16. RDC-based ensembles. *Top: Free energy surfaces as a function of the domains relative orientation for LC wt (left) and LC R108A (right). The populations and average distance of two salt-bridges is also reported. In particular the salt-bridge Arg108 - Asp170 (not formed in the crystal) is strongly populated in the wild type and abolished by the mutation. The Lys103 - Asp165 salt-bridge that is formed in the crystal is instead only marginally populated in both ensembles. Bottom: Comparison of the conformational ensembles for LC wt (left) and LC R108A (right) based on RDCs. The ensembles (grey) are represented below and compared with the crystal structure (PDB: 1FH5, chain L) in yellow. On average the effect of the R108A mutation is that of increasing the accessible conformational space towards more open conformations. RDC data recorded by Dr. Manuel Hora (group of Prof. Bernd Reif, TUM) and processed by Prof. Carlo Camilloni, University of Milano.*

The analysis of the contact probabilities shows a remarkable similarity between the contacts within the two domains (V_L and C_L) for both, LC wt and LC R108A, confirming the common interactions maintaining the Ig fold (Figure 17). In the case of LC R108A, a number of inter-domain contacts are lost, which is in agreement with the more open conformations accessible upon mutating R108 to A108. Noteworthy, the overall amount of inter-domain contacts in both LC variants is relatively low, highlighting the importance of the LC linker to covalently connect V_L and C_L.

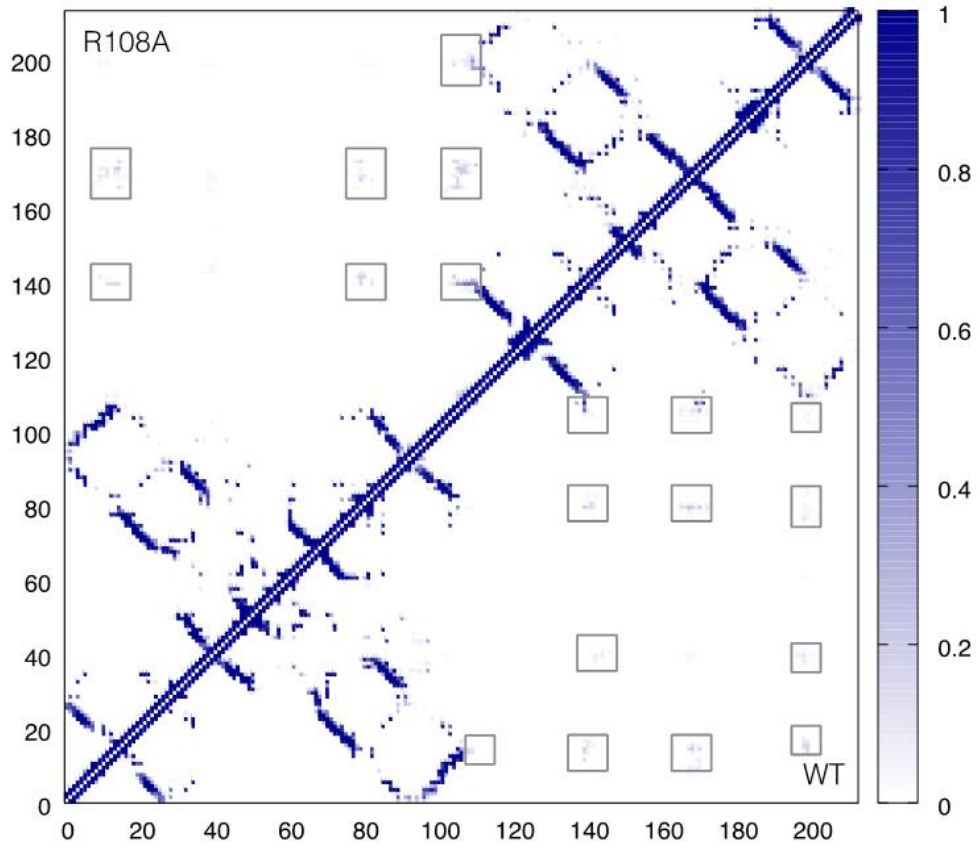

Figure 17. Probability contact maps for LC wt (bottom right) and LC R108A (top left). *The overall effect of the R108A mutation is that of decreasing the populations of all contacts between the V_L and C_L domains. Highlighted are clusters of contacts formed between the two domains with a population > 5%. RDC data recorded by Dr. Manuel Hora (group of Prof. Bernd Reif, TUM) and processed by Prof. Carlo Camilloni, University of Milano.*

3.2. Discussion

Since the individual V_L domain and full-length LC, have been found deposited in amyloid plaques, it is important to determine the amyloidogenic properties of V_L, LC and C_L in comparison (Buxbaum et al., 1990; Dasari et al., 2015; Glenner et al., 1970; Klimtchuk et al., 2010; Lavatelli et al., 2008). In this study, the impact of the C_L domain on the V_L domain was analyzed in the LC framework with a particular focus on the importance of the LC linker in this context.

Based on previous work on the isolated V_L and C_L domains, the presence of Arg108 in the linker between the two domains was suspected to be of crucial importance for the full-length LC (Nokwe et al., 2015). Substituting the linker residue Arg108 by alanine, however, does not substantially affect the stability of V_L or LC. Since the positively charged side chain of R108 does not undergo distinct intramolecular interactions in V_L as judged from the crystal structure (PDB: 1FH5, L chain), it was expected that its replacement by alanine does not affect the domain structure. Moreover, missing electrostatic interactions could also be compensated by Lys107, resulting in similar structural properties of the V_L variants as determined for LC wt and LC R108A. For LC R108A, however, the decrease of polar interactions between Arg108 and its potential interaction partners Tyr140, Asp165, Asp170 and Thr172 in the C_L domain of the LC most likely results in a less rigid orientation of the V_L and C_L domains. RDC-based simulations revealed that there are less transient inter-domain contacts between C_L and V_L for LC R108A than for LC wt. As a result, the structure of LC R108A exhibits increased flexibility, shifting towards opened conformations compared to LC wt. In general, the overall amount of inter-domain contacts in LC wt and LC R108A is relatively low and transient, highlighting the importance of the LC linker to covalently connect V_L and C_L.

Substituting Arg108 by a negatively charged residue, here glutamate, reduces the stabilities of C_L and LC, but not that of V_L. In contrast to V_L, there are likely intramolecular interactions of Arg108 with the C_L domain, i.e. polar interactions with Tyr140, Asp165, Asp170 and Thr172, which anchor the exposed “bc” and “de” loops of C_L to the LC linker, thereby keeping V_L and C_L in close proximity. Interestingly, the deletion of R108 or its substitution by glutamate (C_L ΔR108, C_L R108E) result in a more pronounced decrease in stability than the substitution of R108 by alanine (C_L R108A). Thus, structurally necessary polar interactions of Arg108 can be maintained by the backbone carbonyl oxygen of Ala108 as implied by the RDC-based

simulations. The introduction of a negative charge (Glu108), however, leads to a repulsive effect. This is still present in the full-length LC as indicated by a similar decrease in T_m and NMR chemical shifts observed for Asn138 and Thr172 in LC R108E, which are similar to those of the destabilized variant C_L ΔR108 (Nokwe et al., 2015). In contrast, for LC R108A and LC wt the chemical shifts for these residues were found to be close to the major and more stable C_L conformer. Hence, a conformational shift of the C_L domain and an open orientation of V_L and C_L upon introducing Glu108 are likely. In that context, Nokwe et al. found the deletion of Arg108 in V_L to result in a partially solvent-exposed hydrophobic core (Nokwe et al., 2015). The NMR and RDC data indicate that this is not the case for the full-length LC as there are no substantial chemical shift changes or loss of signals for the core residues of any LC variant studied. Surprisingly, the conformational stability of the LC is generally lower than the average of the stabilities of its individual domains suggesting a negative influence of the observed transient inter-domain contacts and the connection via the linker. Linker extensions by one or five non-polar residues (LC link+G, LC link+GAGAG) resulted in a further decrease in stability consistent with the idea that in these variants Arg108 is sterically separated from its potential interactions partners (Tyr140, Asp165, Asp170 and Thr172), similar to the missing Arg108 in the variant LC R108A.

Although some mutations destabilized and locally altered the conformation of the LC as well as the relative orientation of its domains, no LC variant tested formed fibrils even after 20 days of incubation. This supports the view that the conformational stability does not necessarily correlate with the fibril formation propensity of a given protein (Blancas-Mejía et al., 2015; Camilloni et al., 2016; Marin-Argany et al., 2015; Nokwe et al., 2016). In contrast, all V_L variants studied readily fibrillated: V_L R108A, V_L R108E and V_L ΔR108 revealed comparable fibril formation kinetics, although only V_L ΔR108 exhibits a decreased conformational stability. This emphasizes the importance of Arg108 in particular for the structural integrity, not necessarily for the stability, of the V_L domain. Its presence in V_L seems to inhibit structural transitions that lead to fibril formation. It could well be that Arg108 helps stabilizing a specific folding intermediate as V_L can pursue two distinct folding pathways, each comprising an intermediate, which folds to the native state (Simpson et al., 2009). It was suggested that specific folding intermediates shift towards amyloid formation (Kelly, 1998; Simpson et al., 2009; Uversky and Fink, 2004). The finding that a C_L domain protecting V_L from fibrillation in the context of a kappa LC is consistent with results on a lambda LC (Klimtchuk et al., 2010). For other LCs no

pronounced effect was described (Blancas-Mejía et al., 2015). Hence, the underlying mechanism is more complex and additional aspects have to be factored in, such as kinetically or thermodynamically controlled fibril reactions (Blancas-Mejía et al., 2014, 2015; Marin-Argany et al., 2015; Saelices et al., 2015) as well as proteolysis (Morgan and Kelly, 2016).

C_L did not form any fibrils, neither the wild type protein nor the comparably instable C_L ΔR108 variant, which is in agreement with the literature (Feige et al., 2008). A major difference between the two domains is the presence of helical elements in C_L, which are missing in the V_L domain. These helices rapidly initialize folding of the C_L domain by orientating the β-sheets correctly (Feige et al., 2008; Qin et al., 2007; Simpson et al., 2009). This seems to make C_L intrinsically more resistant against misfolding and amyloid formation than V_L. The data show that in the LC context, C_L protects V_L from fibril formation even in the presence of amyloidogenic linker mutations. Therefore, C_L possibly modulates the folding pathway of V_L to maintain the native fold. This might be achieved by stabilizing the local fold of the C-terminal segment of V_L, which forms β-strands in amyloid fibrils derived from V_L domains as demonstrated by MAS solid-state NMR (Hora et al., 2017; Piehl et al., 2017). The fact that the full-length LC does not form fibrils suggests that an unrestrained C-terminus is a key requirement for the transition from natively folded monomers to amyloid fibrils. Supporting this notion, an important role of the C-terminus in fibrillation initiation was shown for the Aβ peptide (Misra et al., 2016). Thus, the LC is optimized for preventing off-pathway reactions while maintaining the conformational flexibility required for antigen binding and the interaction with the heavy chain.

4. The presence of the C_L domain modulates the amyloidogenicity of the full-length light chain in a mutation-dependent manner

The fibril formation mechanisms of the V_L variants, V_L I2E, V_L S20N and V_L R61A have been investigated earlier (Nokwe et al., 2014, 2016). Noteworthy, the mutations S20N and R61A have been associated with AL amyloidosis as reported elsewhere (Connors et al., 2007; Stevens et al., 1995). In V_L kappa, but not in V_L lambda domains, an Ile residue at position 2 is highly conserved. It is crucial for assuring the thermodynamic stability of the entire domain as well as its amyloid fibril formation propensity (Nokwe et al., 2014). Upon the substitution of Ile2 by glutamate, the hydrophobic core of the protein was found to be more solvent-exposed resulting in an overall decrease in conformational stability. Another mechanism was identified for V_L R61A, which revealed a substantially decreased conformational stability (Nokwe et al., 2016). The substitution of Arg61 leads to the accumulation of non-native intermediates, which are highly susceptible for amyloid fibril formation. Interestingly, Nokwe et al. identified a different mechanism for the wild type-like stable variant V_L S20N (Nokwe et al., 2016). The asparagine establishes non-native intramolecular interactions particularly with serine and threonine residues in its proximity. In contrast to V_L I2E and V_L R61A, the native fold is thereby not destabilized, but the interactions induced by S20N stabilize fibrillar structures, consequently shifting the equilibrium towards fibril formation.

In this study, variants of the full-length LC containing the described known amyloidogenic mutations (I2E, S20N and R61A) were employed to investigate the impact of these residues in the LC framework, which comprises both, the C_L and the V_L domain, covalently linked by the LC linker. As shown in chapter three, the C_L kappa domain can prevent the V_L kappa domain from fibril formation in the context of the full-length LC. The data obtained in this study suggest that this is not necessarily applicable for all full-length LCs as an opposite effect of the presence of C_L could be observed for I2E and particularly for R61A compared to other LC variants (cf. chapter 3).

4.1. Results

To assess the impact of the introduced mutations on the secondary and tertiary structure of the V_L and LC variants, far-UV (Figure 18, top) and near-UV (Figure 18, bottom) CD spectroscopy was performed. For the V_L variants, no impact of the introduced mutations I2E, S20N and R61A on the secondary structure was found. All proteins revealed wild type-like far-UV CD spectra. The same was true for the corresponding LC variants, which all exhibited the characteristic Ig fold, represented by a single minimum at approx. 218 nm indicative for β-strand-rich proteins. In contrast, the near-UV CD spectra revealed a decreased signal for V_L R61A and less pronounced for LC R61A compared to the corresponding wild type and the other variants. For the R61A V_L and LC variants, the altered signal was monitored between 270 and 285 nm wavelengths, which indicates changes in the tertiary structure, particularly for aromatic residues such as tryptophan. In conclusion, all V_L and LC variants as well as C_L wt were properly folded and exhibited the characteristic Ig fold. Thus, the introduced mutations do not substantially change the conformation of the LC. However, for V_L R61A and LC R61A, a slight difference in the tertiary structure was observed.

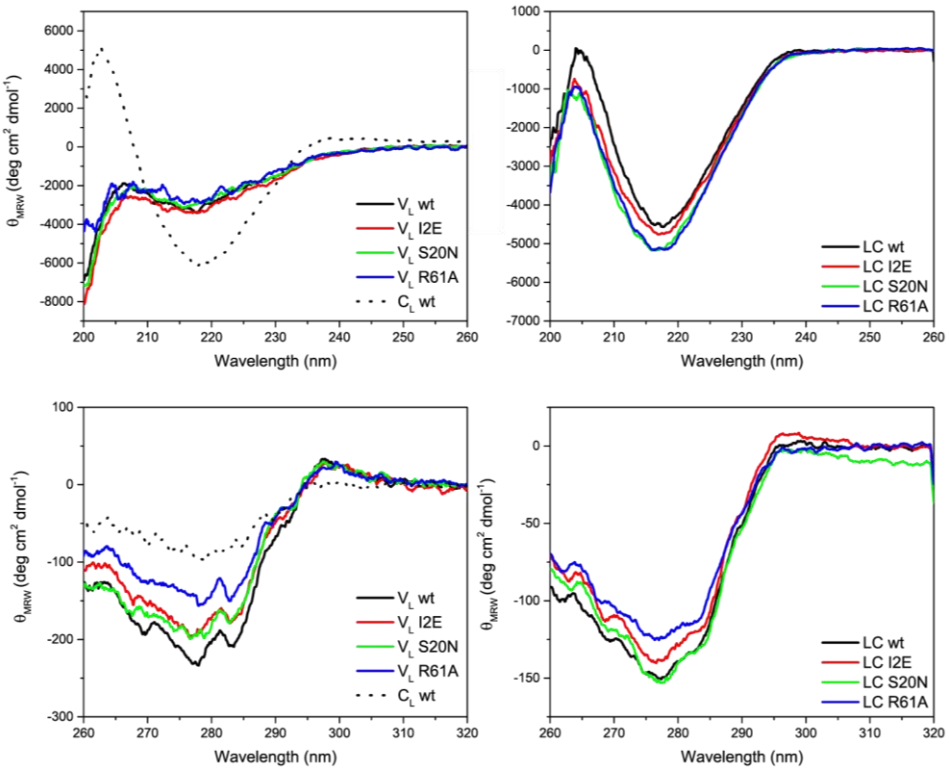

Figure 18. Secondary and tertiary structure analysis. *Far-UV (top) and near-UV (bottom) CD spectra of V_L variants and C_L wt (left) as well as LC variants (right). The introduced mutations did not significantly alter the secondary structure of the variants with respect to their wild type protein. Interestingly, upon the introduction of the R61A mutation in V_L, the ellipticity is decreased between 265 and 290 nm. For LC R61A this effect was less pronounced.*

To assure a homogenous oligomeric state of all variants, SEC-MALS was performed to assess the molecular mass of the variants and thereby characterize the quaternary structure (Figure 19). All investigated V_L variants as well as C_L wt were predominantly monomeric as judged by the molecular mass of the major species (Table 6). For V_L wt and V_L S20N, a minor proportion of a higher molecular weight species was observed in the SEC-MALS chromatograms represented by a shoulder of the main peak. Due to the low concentration of that minor fractions, it was not possible to determine their molecular masses. For the LC variants, the determined molecular masses correspond to monomers. In summary, the introduced mutations did not alter the oligomeric state of the V_L and LC variants.

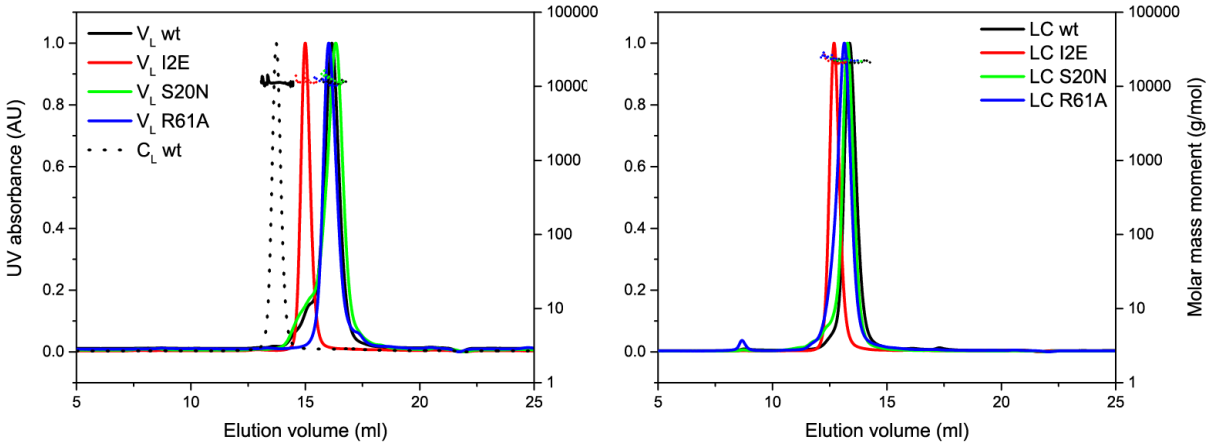

Figure 19. Quaternary structure analysis. *SEC-MALS chromatograms of V_L variants and C_L wt (left) as well as LC variants (right) reveal the monomer as the predominant species. V_L wt and V_L S20N show a minor fraction of higher molecular weight species represented by a small shoulder of the main peak. Of notice, the monomer peak for V_L I2E is shifted towards lower elution volumes although the calculated mass corresponds to a monomer. A similar shift was observed for LC I2E, however, less pronounced. 50 µl of 50 µM samples were injected.*

Variant	Molecular mass (kDa)		
	V_L	LC	C_L
wt	11.7	20.9	11.2
I2E	12.1	22.9	N/A
S20N	11.7	22.7	N/A
R61A	12.3	23.4	N/A

Table 6. SEC-MALS-derived molecular masses. *Molecular masses in kDa of the predominant peaks of the size-exclusion chromatograms as determined by SEC-MALS. The calculated molecular masses using the ProtParam tool are 11.8 kDa for V_L wt, 23.4 kDa for LC wt and 11.8 kDa for C_L wt. All variants were monomeric. N/A: construct does not exist.*

As many amyloid precursor proteins reveal a low conformational stability (Blancas-Mejía et al., 2014; Nokwe et al., 2014), an important part of the characterization of aggregation-prone proteins is the analysis of the conformational stability. Here, thermal unfolding transitions monitored by CD spectroscopy were performed (Figure 20 and Table 7). All V_L variants showed a decreased conformational stability compared to V_L wt. V_L R61A revealed the lowest T_m (31.9 °C) of all variants, followed by V_L I2E (T_m 42.5 °C). V_L S20N (T_m 50.3 °C) was found to be similarly stable as V_L wt (T_m 51.8 °C). These results are consistent with the literature (Nokwe et al., 2014, 2016). C_L wt (T_m 53.8 °C) revealed the highest stability of all investigated proteins. For the full-length LC variants, LC R61A (T_m 43.2 °C) and LC I2E (T_m 48.7 °C) exhibited the lowest stabilities. Like for V_L, LC wt (T_m 50.8 °C) and LC S20N (T_m 52.4 °C) revealed similar melting temperatures.

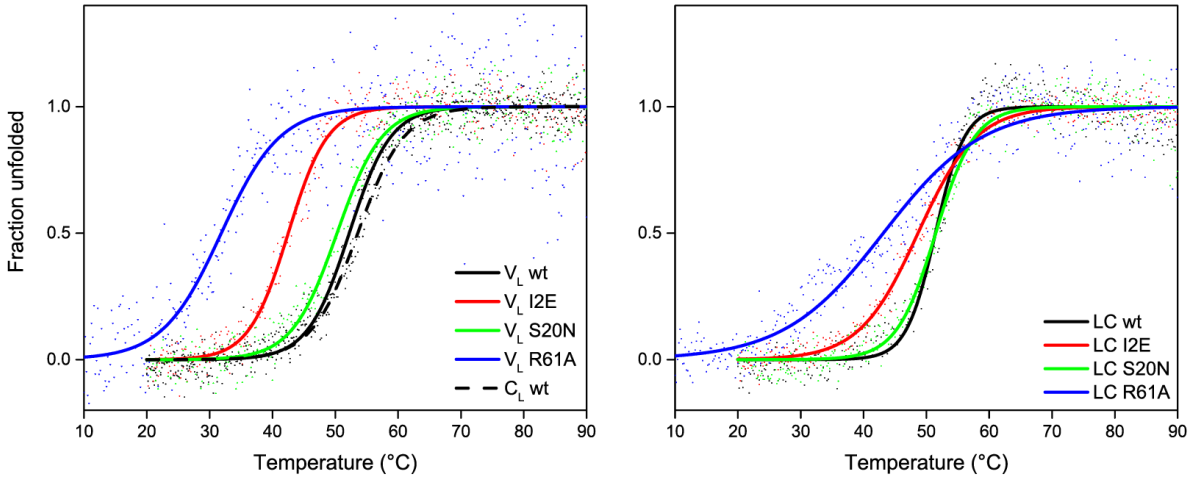

Figure 20. Conformational stability. *Thermal unfolding transitions followed by CD spectroscopy reveal a distinct impact of specific mutations on the V_L (left) and LC (right) variants. The CD signal was recorded at a constant wavelength of 205 nm during heating with a rate of 0.5 °C/min. Solid lines represent the Boltzmann fit of the monitored data to determine the melting temperature (T_m).*

Variant	Melting temperature (°C)		
	V_L	LC	C_L
wt	51.8 ± 0.3	50.8 ± 0.7	53.8 ± 0.5
I2E	42.5	48.7	N/A
S20N	50.3	52.4	N/A
R61A	31.9	43.2	N/A

Table 7. Conformational stability. *Melting temperatures (T_m) of V_L and LC variants as well as C_L wt as determined by CD thermal unfolding transitions in °C. The heating rate was 0.5 °C/min. The Error indicates the standard deviation of three technical replicates. N/A: construct does not exist.*

Interestingly, the presence of the C_L domain increased the T_ms of V_L I2E and V_L R61A in the context of the full-length LC, however to a different extent. Moreover, it was observed that the slope of LC I2E and LC R61A variants was less steep than for their corresponding V_L variants, V_L I2E and V_L R61A. This suggests that the two domains, V_L and C_L, unfold individually rather than cooperatively, although they are covalently bonded by the LC linker. This is supported by the thermal unfolding of LC R61A, which shows two distinct transitions (Figure 20, right). The two transition midpoints are at approx. 35 °C and 50 °C corresponding to the melting temperatures of the individual V_L and C_L domains. In summary, the I2E and R61A mutations decrease the conformational stability of the V_L domain. This effect is still present, but less substantial in the full-length LC variants.

As CD thermal unfolding transitions cannot precisely monitor the individual unfolding of the two constituent LC domains, V_L and C_L, it is possible that the slope, which reflects the unfolding cooperativity, is altered by the introduced mutations. Therefore, a theoretical approach was utilized to arithmetically determine the unfolding transition curves of the individual domains derived from the recorded data of the full-length LC variants (Figure 21). The comparison of experimental and arithmetical thermal unfolding transitions allows to detect interactions between the two LC domains, V_L and C_L, or stability changes upon covalent linkage of both domains. This method can be applied with the assumption that in the absence of interactions between C_L and V_L, the sum of the experimental thermal unfolding transitions of the two constituent LC domains represents the transition curve of the full-length LC. Thus, not only the T_ms, but also the transition slope of the LC should be represented by the sum of V_L and C_L. For V_L wt and V_L I2E a high degree of similarity was observed for arithmetically and experimentally determined unfolding transitions. Both, the T_ms (T_m^{cal} and T_m^{exp}) as well as the slope (i.e. cooperativity) of the V_L, LC and C_L variants match well for wild type and I2E (Table 8). For variants comprising the S20N mutation, the transitions curves revealed different slopes, particularly for V_L S20N and C_L, not for LC S20N. The most substantial differences between the experimental and calculated unfolding transitions were observed for V_L R61A and LC R61A. The theoretical slope of V_L R61A, derived from LC R61A and C_L wt, was steeper than the experimental transition of V_L R61A. The R61A mutation is similarly impacting the unfolding cooperativity of the V_L domain for V_L R61A and LC R61A. However, the T_m^{cal} (49.4 °C) of LC R61A is approx. 6 °C higher than T_m^{exp} (43.2 °C) indicating a rather destabilized V_L R61A domain in the full-length LC framework. In summary, by arithmetically determining the

theoretical transition curves, it was possible to detect alterations for V_L R61A and LC R61A, showing a substantial destabilizing effect upon the presence of the C_L domain. Thus, destabilizing inter-domain interactions between V_L R61A and the C_L domain or a strong effect of the covalent binding are likely.

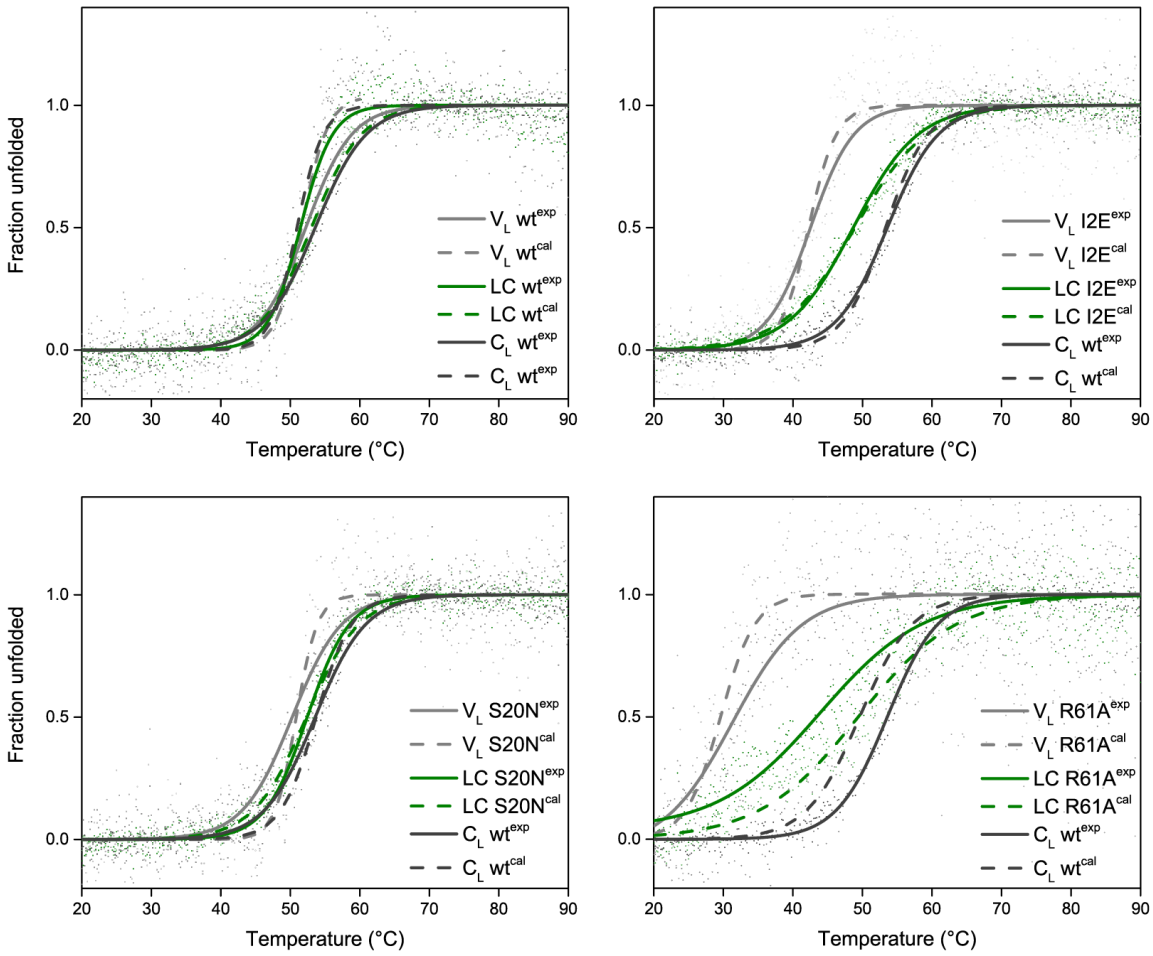

Figure 21. Experimental (^{exp}) and arithmetical (^{cal}) conformational stabilities of LC wt (top left), I2E (top right), S20N (bottom left) and R61A (bottom right) and the constituent domains, V_L and C_L. *Experimental (^{exp}) thermal unfolding transitions were observed using CD spectroscopy. A summary of all experimental CD thermal unfolding transitions is provided in Figure 20 and Table 7. Calculated (^{cal}) thermal transitions are derived from subtraction of the corresponding constituent domains as described in the results section.*

Variant	Melting temperature (°C)					
	V_L		LC		C_L	
	T_m^{exp}	T_m^{cal}	T_m^{exp}	T_m^{cal}	T_m^{exp}	T_m^{cal}
wt	52.2	51.7	51.4	53.0	53.6	50.1
I2E	42.5	42.2	48.7	48.9	N/A	53.3 (wt)
S20N	50.3	50.9	52.4	52.3	N/A	53.7 (wt)
R61A	31.9	29.6	43.2	49.4	N/A	49.6 (wt)

Table 8. Experimental (^{exp}) and calculated (^{cal}) conformational stability. *Experimental melting temperatures (T_m^{exp}) of all V_L and LC variants as well as C_L wt as determined by CD thermal unfolding transitions in °C. Calculated melting temperatures (T_m^{cal}) were arithmetically determined as described in the methods chapter. N/A: construct does not exist.*

To assess how the introduced mutations affect the amyloid fibril formation propensity, ThT assays were performed. As already shown for other protein, MAK33 V_L wt is able to form amyloid fibrils on different time scales dependent on the experimental conditions (Feige et al., 2008; Nokwe et al., 2014). Here, the same effect was found for the other V_L variants, V_L I2E, V_L S20N and V_L R61A, when comparing the t_{50} with the literature (Nokwe et al., 2014, 2016). All V_L variants readily formed amyloid fibrils as detected by the ThT assays and verified by TEM analysis (Figure 22). V_L I2E exhibited the fastest fibril formation kinetic (t_{50} 3.40 ± 0.22 d), followed by V_L R61A (t_{50} 3.97 ± 0.53 d) (Table 9). V_L wt (t_{50} 9.58 ± 0.47 d) and V_L S20N (t_{50} 13.17 ± 0.02 d) were significantly slower in amyloid formation. In contrast, C_L wt did not form any fibrils, which is consistent with the literature (Feige et al., 2008). For the corresponding LC variants, significantly altered fibril formation propensities were monitored. LC wt and LC S20N did not form any amyloid fibrils, even if incubated for 20 d (data not shown). In contrast, LC I2E and LC R61A revealed a significantly faster t_{50} than the corresponding V_L variants. LC R61A revealed the fastest kinetic (t_{50} 0.19 ± 0.03 d), more than 3 days faster than V_L R61A (t_{50} 3.97 ± 0.53 d). LC I2E (t_{50} 1.99 ± 0.21 d) formed fibril approx. 1.5 d faster than V_L I2E (t_{50} 3.40 ± 0.22 d). Noteworthy, the fibril morphology of the V_L and LC proteins was altered upon the extension by the C_L domain. Additionally, the ThT fluorescence intensity of the raw data was lower for V_L R61A and LC I2E, indicating a lower fibril concentration for these two proteins compared to the other variants studied (data not shown).

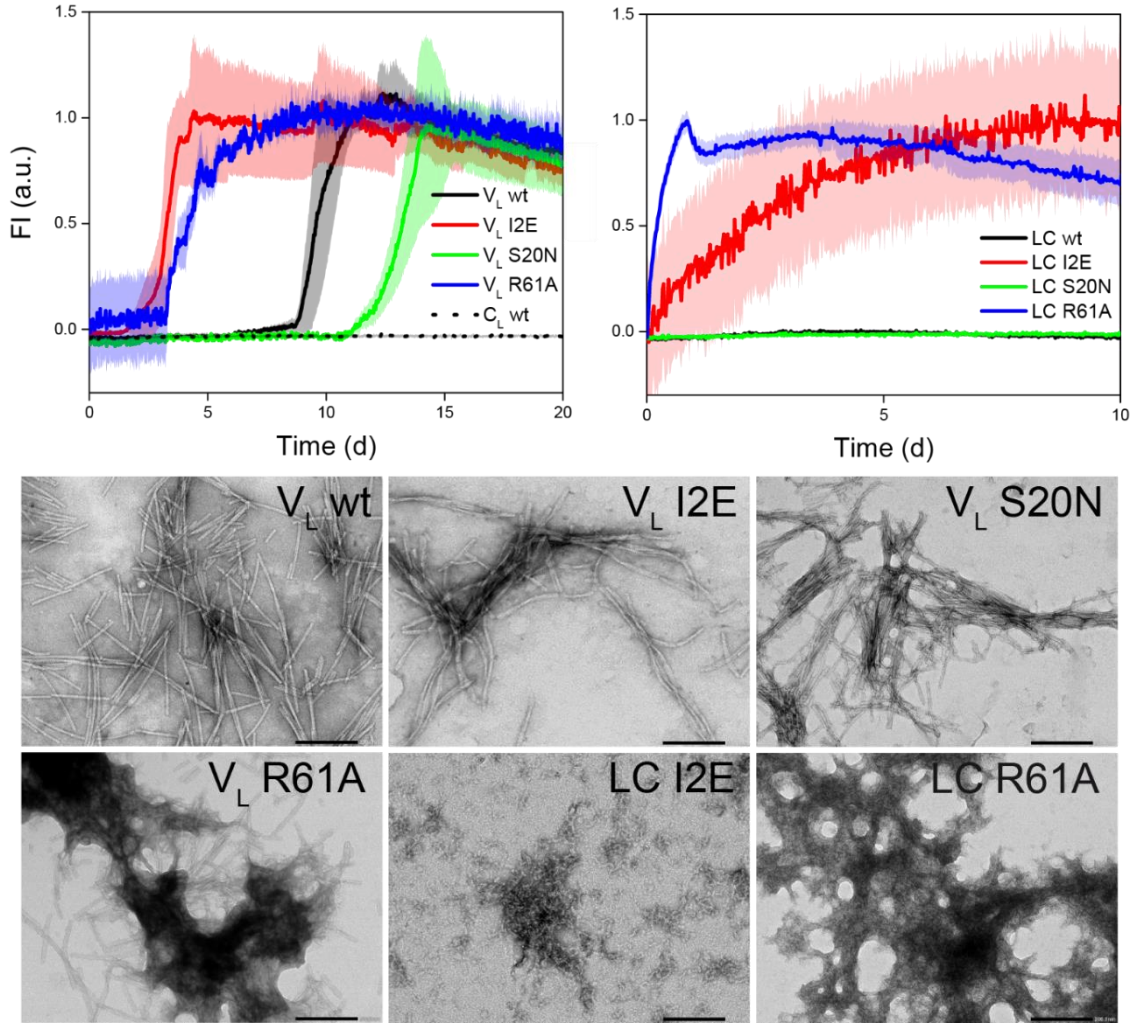

Figure 22. Amyloid fibril formation propensity. *Top: fibril formation kinetics of V_L variants and C_L wt (left) as well as LC variants (right) followed by ThT fluorescence. "Shadows" indicate standard deviation of at least two technical replicates. For better kinetic comparison, the fibril formation curves were individually normalized. Bottom: TEM micrographs of ThT-positive samples. TEM micrographs were obtained using 50,000 x magnification. The bar represents 200 nm.*

Variant	t_{50} (days)		
	V_L	**LC**	**C_L**
wt	9.58 ± 0.47	no fibrils	no fibrils
I2E	3.40 ± 0.22	1.99 ± 0.21	N/A
S20N	13.17 ± 0.02	no fibrils	N/A
R61A	3.97 ± 0.53	0.19 ± 0.03	N/A

Table 9. Fibril formation kinetics. *t_{50} values represent the time point when 50 % of the total fibril formation was reached. Data was taken from ThT fibril formation kinetics. The error represents the standard deviation of at least two technical replicates. N/A: construct does not exist.*

To understand the impact of the introduced mutations and the presence of C_L on the V_L domain in the full-length LC framework from a structural point of view, NMR experiments were performed by Dr. Manuel Hora and Tejaswini Pradhan (group of Prof. Bernd Reif, TUM). Chemical shifts changes derived from HSQC experiments revealed substantial changes within the V_L domain between V_L wt and LC wt (Figure 23 and Figure 24). Since the conformational stabilities of V_L R61A and LC R61A are untypically low (Table 7), the measured HSQCs revealed significantly altered signals at the measurement temperatures of 25 °C and 30 °C (data not shown). This is most likely due to a partially or fully unfolded V_L R61A domain, in both, the individual V_L protein or in the corresponding full-length LC. As this is rather artificial, the experimental conditions need to be further optimized for these variants since lower temperatures result in decreased HSQC peaks. By analyzing the chemical shift changes between V_L wt and its corresponding LC variants, it is possible to determine the relative impact of the C_L domain and the introduced mutations on the V_L domain (Figure 23). Interestingly, many residues experienced chemical shift changes for the wild type proteins. Most substantial changes were observed for the residues 10 - 11, 79 - 87 and 103 - 105. Compared to LC wt, the V_L domain within the LC I2E and LC S20N variants exhibited less chemical shift changes. Interestingly, among the regions showing changes in all variants, are the above-mentioned residues representing a part of two β-strands (10 - 11 and 103 - 105) and an α-helix (79-87). These secondary structure elements are structurally close to each other and interact by polar contacts as judged from the crystal structure (PDB: 1FH5, chain L). In the folded state, the β-strand comprising the C-terminal residues 103 - 105 are in the center of these three secondary structure elements. In the full-length LC, the C-terminal residues are extended by the LC linker and the C_L domain. Moreover, among the residues exhibiting pronounced chemical shift changes in all variants are Glu80, Thr81 und Glu82, which form a conserved salt bridge to Arg61 (Nokwe et al., 2016).

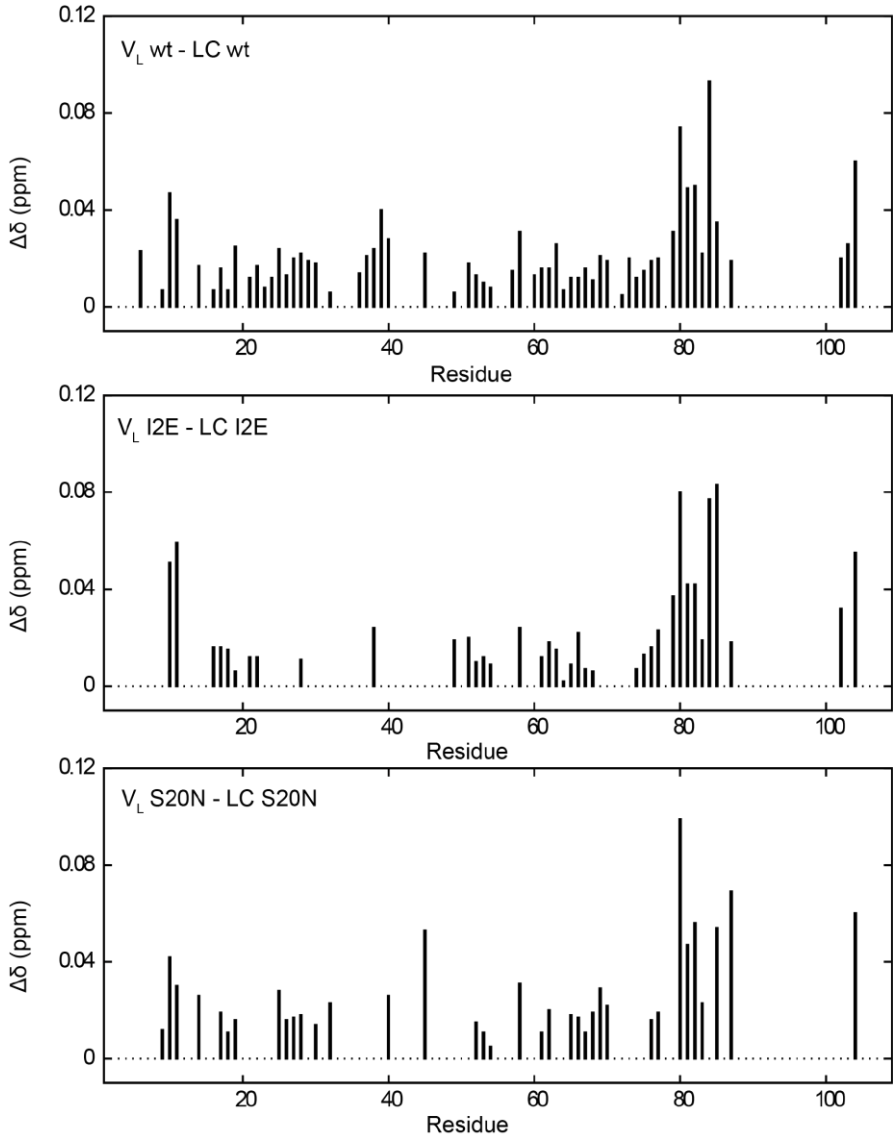

Figure 23. Chemical shift differences of corresponding LC and V_L variants. *NMR identified regions showing similar chemical shift differences upon the presence of the C_L domain. In general, more pronounced differences were observed for V_L wt than for its variants. Data recorded and processed by Dr. Manuel Hora and Tejaswini Pradhan (group of Prof. Bernd Reif, TUM).*

To test the impact of the introduced mutations on the full-length LC, chemical shift changes between LC wt and the LC variants were analyzed by Dr. Manuel Hora (group of Prof. Bernd Reif, TUM) (Figure 24). Substantial changes were observed within the V_L domain. Interestingly, only minor chemical shift changes were observed within C_L domain, which does not contain any of the introduced mutations. Two main regions of chemical shift changes were identified for both, LC I2E and LC S20N. The first set of signals is distributed around residue 20, which is a β-strand and two loops. The other region is distributed around residue 70, containing two more β-strands, which are in close proximity to the first defined region. For LC S20N, intense chemical shifts changes compared to LC wt were additionally observed for the C-terminal region of the C_L domain.

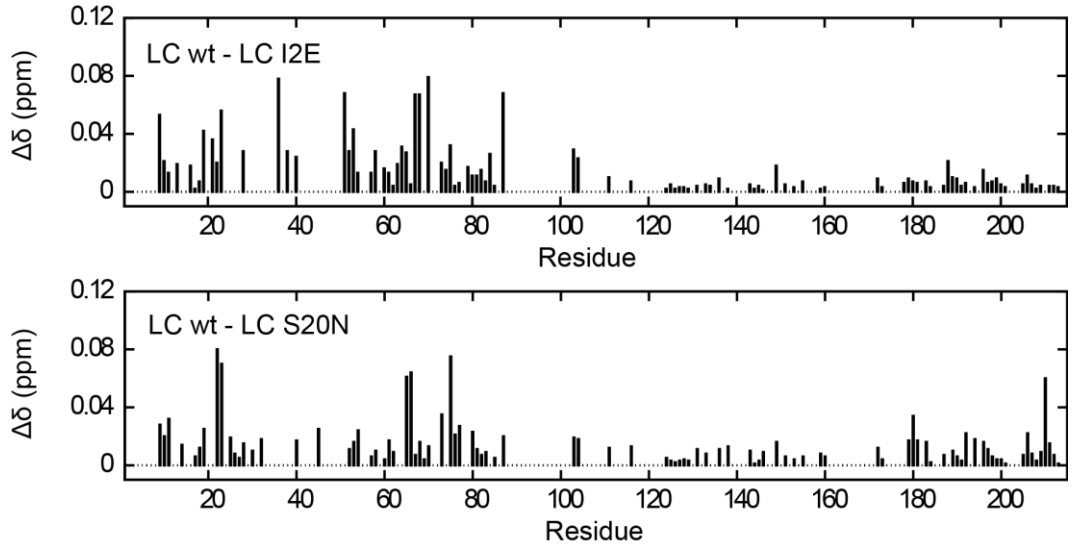

Figure 24. Chemical shift differences between LC wt and LC variants. *NMR analysis revealed more pronounced chemical shift differences within the V_L domain (residue 1-108), less within the C_L domain (residue 109-213). Moreover, the LC variants exhibited altered chemical shift changes compared to LC wt. For LC I2E, more chemical shift changes were observed than for LC S20N. Interestingly, for LC S20N changes in the C-terminal region of C_L were observed. Data recorded and processed by Dr. Manuel Hora and Tejaswini Pradhan (group of Prof. Bernd Reif, TUM).*

To analyze substantially faster fibril formation kinetics of LC I2E and particularly LC R61A in detail and gain deeper structural information, homology modelling was performed using the Phyre2 web server (Kelley et al., 2015) (Figure 25). For comparison, the crystal structure of MAK33 LC wt (PDB: 1FH5, chain L) was employed, which was crystalized in the context of the assembled Fab fragment (Augustine et al., 2001). In the LC wt, the side chain of Arg61 is involved in polar interactions with Glu81 and Asp82 (yellow dashed lines in Figure 25). Both, Glu81 and Asp82 are part of a short α-helix, which is located structurally close to the C_L domain. However, as judged from the crystal structure of the LC wt, there are no interactions between this α-helix of V_L with the C_L domain. The homology modelled structure of LC R61A indicates minor conformational shifts induced by the absence of the arginine side chain. As a result of the structural rearrangements, the interactions between Ala61 and Glu81 as well as Asp82 are extinguished, which leads to an altered orientation of the Glu81-containing α-helix towards a closer proximity to C_L. As a result of that, polar interactions between Ser40 and Asp165 as well as Thr80 and Ser168 are possible (Figure 25). As the conformational stability and structural integrity are substantially decreased by the R61A substitution (Table 7 and Table 9), it is reasonable that these effects are caused by the missing interaction between

residue 61 and the α-helix containing Glu81 and Asp82. Moreover, the newly acquired likely polar interactions between V_L and C_L (Ser40 - Asp165 and Tht80 - Ser168) might interfere with the structural integrity of the V_L domain. Thus, as a result of the structural analysis and comparison of LC wt and the homology modelling derived LC R61A, four LC R61A double mutation variants were created and characterized: LC R61A S40P, LC R61A T80A, LC R61A S168A and LC R61A K169A (Figure 26). Thr80, Ser168 and Lys169 were substituted by alanine to erase polar interactions. The S40P substitution was introduced because proline is the most frequent residue at position 40 as indicated by the abYsis database tool for identification of unusual residues (data not shown) (Swindells et al., 2017).

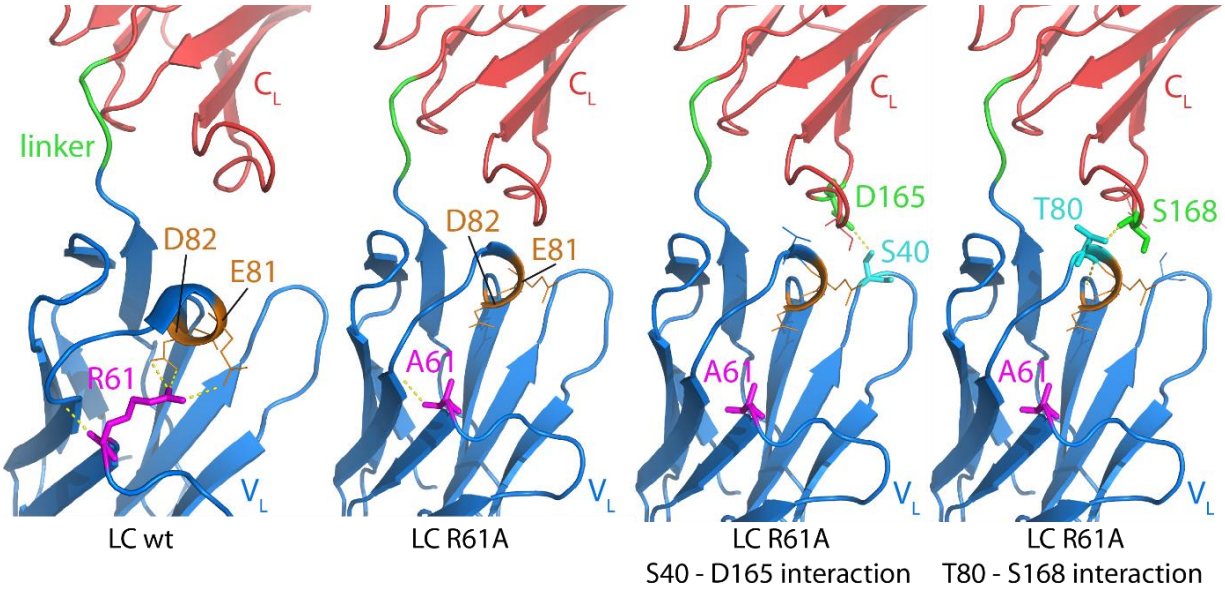

Figure 25. Structural comparison of LC wt and the homology modelled LC R61A as well as likely polar interactions between V_L and C_L induced by the R61A mutation. *Homology modelling was employed to gain structural understanding of the impact of the R61A mutation on the conformation of the MAK33 LC. The crystal structure of LC wt (PDB: 1FH5, chain L) was employed for comparison. Homology modelling was performed using the Phyre2 web server (Kelley et al., 2015). Yellow dotted lines indicate polar interactions between atoms of the respective residues. There are likely polar interactions between Thr80 (V_L) and Ser168 (C_L) as well as between Ser40 (V_L) and Asp165 (C_L).*

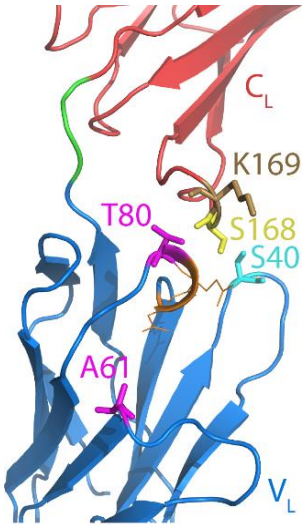

Figure 26. LC R61A double mutant variants. *The structure shows the homology modelling derived LC R61A with the highlighted residues additionally mutated in this study, i.e. LC R61A double mutants: Ser40 (turquoise), Thr80 (magenta), Ser168 (yellow) and Lys169 (brown).*

The LC R61A double mutation variants were characterized according to the other protein employed in this study and compared to LC R61A. The variants except LC R61A S40P revealed similar secondary structure as judged from far-UV CD spectra (Figure 27). For LC R61A S40P the far-UV CD spectrum showed a more pronounced minimum at approx. 218 nm wavelength. By comparing near-UV CD spectra to assess the tertiary structure of the variants, a substantial difference between LC R61A S40P and all other variants was observed (Figure 27). Like monitored for the far-UV spectrum, the near-UV spectrum revealed alterations of the structure for LC R61A S40P, particularly between 265 and 290 nm, indicating minor changes in the secondary structures and more pronounced alteration in the tertiary structures.

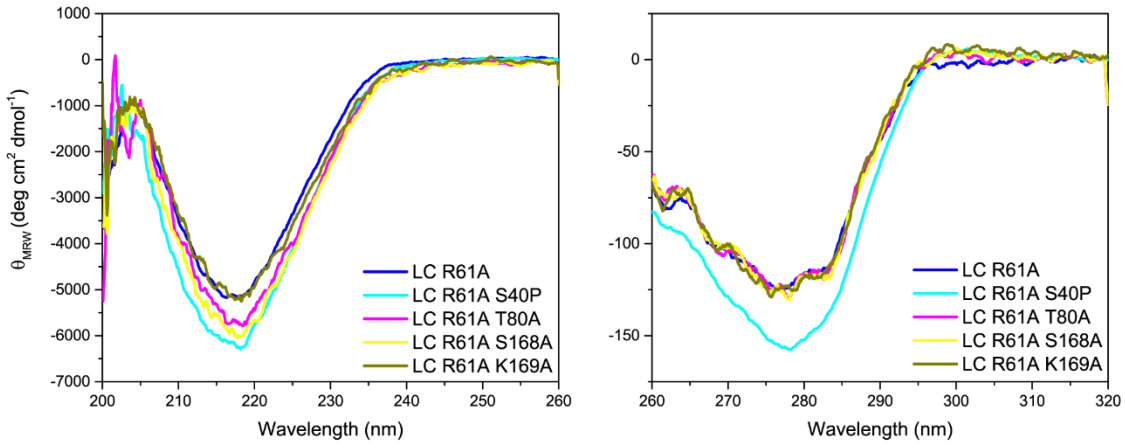

Figure 27. Secondary and tertiary structure analysis. *Far-UV (left) and near-UV (right) CD spectra of LC R61A double mutation variants.*

SEC-MALS analysis of the LC R61A double mutation variants revealed similar elution volumes for all variants (Figure 28), matching the elution volume of LC R61A, which was shown to be monomeric (Figure 19). Interestingly, for all LC R61A double mutation variants, a minor fraction of high molecular weight species at approx. 8 ml elution volume was observed. Although the concentration of this fraction was too low to determine its molecular mass, it is most likely that this fraction corresponds to aggregates indicated by an intense light scattering signal, characteristic for very large molecules, e.g. aggregates or fibrils (light scattering traces not shown). This indicates that all LC R61A variants aggregate as the fraction at approx. 8 ml is not present in the SEC chromatograms of LC wt, LC I2E and LC S20N. The calculated molar masses for all LC R61A variants except for LC R61A K169A correspond to monomers (Table 10). SEC-MALS analysis revealed a molecular mass of 49.5 kDa for LC R61A K169A, which suggests this variant to be dimeric at the given concentration.

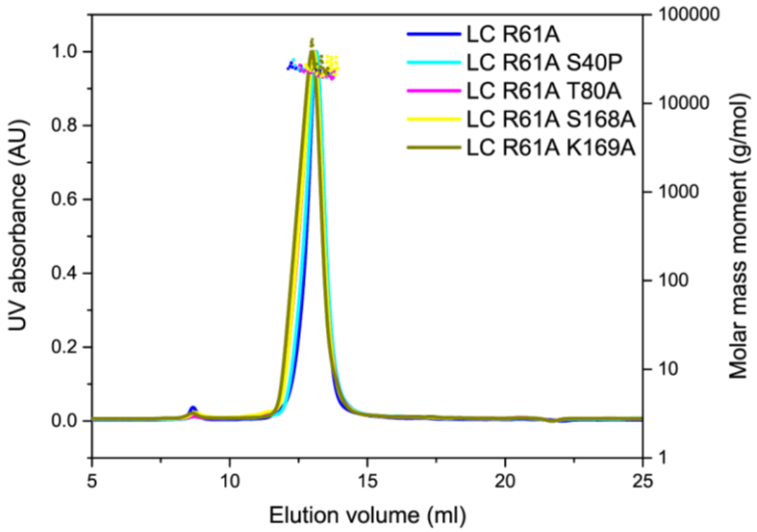

Figure 28. Quaternary structure analysis. *SEC-MALS chromatograms of LC R61A double mutation variants reveal all proteins to have the same elution volumes. The calculated molecular masses for all variants indicate monomers. The only exception of that is LC R61A K169A, which main peak exhibited a calculated molecular mass of 49.5 kDa, corresponding to a dimer. However, elution profile and calculated mass of this variant clearly mismatch. 50 µl of 50 µM samples were injected.*

Variant	Molecular mass (kDa)		
	V_L	LC	C_L
wt	11.7	20.9	11.2
R61A	12.3	23.4	N/A
R61A S40P	N/A	20.9	N/A
R61A T80A	N/A	21.3	N/A
R61A S168A	N/A	23.2	N/A
R61A K169A	N/A	49.5	N/A

Table 10. SEC-MALS-derived molecular masses. *Molecular masses in kDa of the predominant peaks of the size-exclusion chromatograms as determined by SEC-MALS. The calculated molecular masses using the ProtParam tool are 11.8 kDa for V_L wt, 23.4 kDa for LC wt and 11.8 kDa for C_L wt. All variants were monomeric. Of notice, the SEC-MALS chromatogram for LC R61A K169A revealed an identical elution time as the other investigated LC variants, although the calculated molecular mass is indicative for a dimer. N/A: construct does not exist.*

The conformational stability of the LC R61A double mutation variants was investigated by thermal unfolding transitions followed by CD spectroscopy (Figure 29). Surprisingly, LC R61A S40P (T_m 38.0 °C) exhibited a further decreased conformational stability compared to LC R61A (T_m 43.2 °C) (Table 11). All other R61A variants revealed similar T_ms and unfolding cooperativities. Thus, the substitution of Ser40 by proline negatively affected the conformational stability, while the other additional mutations did not show substantial effects.

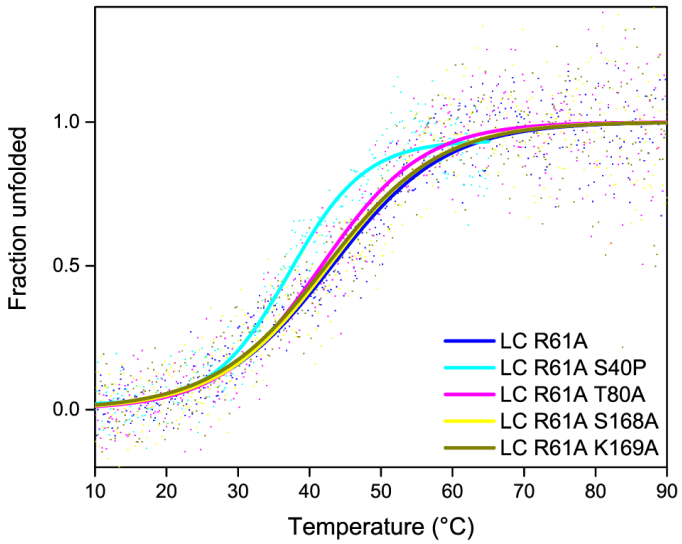

Figure 29. Conformational stability. *Thermal unfolding transitions followed by CD spectroscopy of LC R61A double mutation variants. The CD signal was recorded at a constant wavelength of 205 nm during heating using a rate of 0.5 °C/min. Solid lines represent the Boltzmann fit of the monitored data to determine the melting temperature (T_m).*

Variant	Melting temperature (°C)		
	V_L	LC	C_L
wt	51.8 ± 0.3	50.8 ± 0.7	53.8 ± 0.5
R61A	31.9	43.2	N/A
R61A S40P	N/A	38.0	N/A
R61A T80A	N/A	41.7	N/A
R61A S168A	N/A	42.7	N/A
R61A K169A	N/A	42.3	N/A

Table 11. Conformational stability. *Melting temperatures (T_m) of the LC R61A double variants compared to V_L, LC and C_L wt as determined by thermal unfolding transitions followed by CD spectroscopy. The heating rate was 0.5 °C/min. The error indicates the standard deviation of three technical replicates. N/A: construct does not exist.*

Like for LC wt, LC I2E, LC S20N and LC R61A, ThT fibril formation experiments were performed to investigate the structural integrity and fibril formation propensity of the LC R61A variants (Figure 30). All variants formed amyloid fibrils after less than 5 days of incubation at 37 °C in PBS pH 7.4 under gentle shaking. Interestingly, the kinetics were affected by the introduction of the double mutations (Table 12). LC R61A S40P (t_{50} 0.09 ± 0.01 d) formed fibrils even faster than LC R61A (t_{50} 0.19 ± 0.03 d) and any other protein in this study. LC R61A S168A (t_{50} 0.39 ± 0.01 d) and LC R61A K169A (t_{50} 0.41 ± 0.01 d) showed similar kinetics, which formed fibrils approx. 2-fold slower than LC R61A. LC R61A T80A (t_{50} 0.51 ± 0.02 d) revealed the highest t_{50} of all LC R61A variants. Noteworthy, LC R61A K169A exhibited the longest time to reach the maximal ThT signal, suggesting a comparable fast fibril formation initiation for all variants, but altered recruitment of monomers to prolong the fibrils (Figure 30).

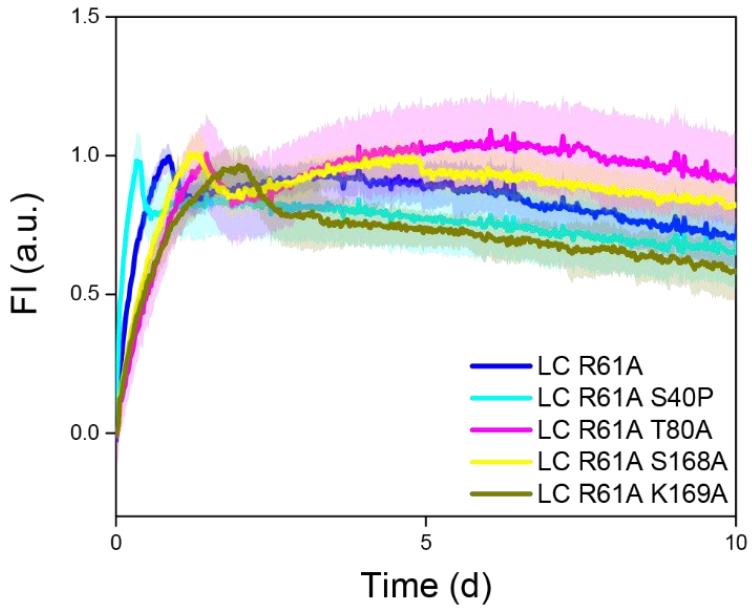

Figure 30. Amyloid fibril formation propensity. *Fibril formation kinetics of LC R61A double mutation variants followed by ThT fluorescence revealed altered fibril formation kinetics dependent on the specific mutation. "Shadows" indicate standard deviation of at least two technical replicates. For better kinetic comparison, the fibril formation curves were individually normalized.*

Variant	t_{50} (days)		
	V_L	**LC**	**C_L**
wt	9.58 ± 0.47	no fibrils	no fibrils
R61A	3.97 ± 0.53	0.19 ± 0.03	N/A
R61A S40P	N/A	0.09 ± 0.01	N/A
R61A T80A	N/A	0.51 ± 0.02	N/A
R61A S168A	N/A	0.39 ± 0.01	N/A
R61A K169A	N/A	0.41 ± 0.01	N/A

Table 12. Fibril formation kinetics. *t_{50} values represent the time point when 50 % of the total fibril formation was reached. Data was taken from ThT fibril formation kinetics. The error represents the standard deviation of at least two technical replicates. N/A: construct does not exist.*

4.2. Discussion

In early amyloid research, a decreased conformational stability of the native precursor protein, e.g. by mutations, was invariably correlated with a higher fibril formation propensity (Chiti, 2000; Hurle et al., 1994; Ramirez-Alvarado et al., 2000; Safar et al., 1993). However, there is growing evidence that conformational stability is only one of many factors giving rise to amyloid fibril formation and therefore the stability of a given protein does not necessarily correlate with its amyloid formation propensity (Chiti and Dobson, 2017; Marin-Argany et al., 2015; Nokwe et al., 2016). Moreover, recently more full-length LCs, not only V_L fragments, have been found deposited in amyloid plaques (Buxbaum et al., 1990; Dasari et al., 2015; Glenner et al., 1970; Klimtchuk et al., 2010; Lavatelli et al., 2008). As a consequence of both, it is important to determine the amyloidogenic properties of V_L and LC and how this is affected by the presence of the C_L domain.

The amyloidogenic mutations I2E, S20N and R61A employed in this study were earlier characterized, however, exclusively in the V_L framework, but not in the context of the full-length LC (Helms and Wetzel, 1996; Nokwe et al., 2014, 2016). Therefore, the fibril formation mechanism for the V_L I2E, V_L S20N and V_L R61A variants are known (discussed at the beginning of chapter 4). However, for LC I2E and LC R61A the proposed mechanism by Nokwe et al. do not explain the surprisingly faster fibril formation of the full-length LC variants compared to the corresponding V_L proteins, which is not reflected by other biophysical properties such as secondary, tertiary, quaternary structure or conformational stability. Using a simple arithmetical approach, it was possible to calculate T_ms, which can be expected if the LC variants experience no major inter-domain interactions between C_L and V_L. If this was the case, the thermal unfolding transition of LC should be the sum of the corresponding V_L and C_L proteins. Particularly LC R61A exhibited alterations between the experimentally and arithmetically derived unfolding curves indicating interactions within the protein, which affect the conformational stability. Interestingly, for LC I2E both curves, experimental and calculated, matched well for V_L, LC and C_L. A well-known prerequisite for proteins to enter the amyloid formation pathway is the conversion to partially unfolded intermediate states (Chiti, 2000; Chiti and Dobson, 2017; Ramirez-Alvarado et al., 2000). As many reported mechanisms differ in the detailed molecular pathway leading to partially unfolded species susceptible for fibril formation, even minor structural changes are able to interrupt the structural integrity of a protein (Blancas-Mejía and Ramirez-Alvarado, 2016; Blancas-Mejía et al., 2014; Marin-Argany

et al., 2015; Nokwe et al., 2014, 2016). This study reports that the C_L domain is able to structurally stabilize V_L variants, which are able to form fibrils (V_L wt, V_L S20N and cf. chapter 3). In contrast, for specific variants (V_L I2E and V_L R61A), an opposite effect was observed leading to accelerated fibril formation upon the presence of the covalently linked C_L domain. As suggested by the analysis of experimental and arithmetical unfolding transitions, the opposite effect is likely due to inter-domain interactions between C_L and V_L, facilitating the conversion of the native state to partially unfolded intermediates. This is supported by the decelerated fibril formation resulted from substituting a second residue besides R61A for the variants LC R61A T80A, LC R61A S168A and LC R61A K169A. However, all LC R61A variants still readily formed fibrils faster than V_L R61A implying that the additionally investigated residues on their own are not key. Most likely, as the exposed C_L loop containing Asp165, Ser168 and Lys169 is rich in polar and charged residues, the unfavorable interactions leading to accelerated fibril formation are maintained by other nearby residues. Since the data did not imply inter-domain interactions for LC I2E, but for LC R61A, it is likely that there are two different modes of interaction. Including the molecular fibril formation mechanism of V_L I2E and V_L R61A, and the assumption that for LC I2E and LC R61A the probability for the occurrence of partially unfolded states is increased, it might be that primarily the V_L domain is involved in fibril formation, even in the LC context. This is supported by the fact that C_L is intrinsically resistant to fibril formation (Feige et al., 2008). The data suggest that non-native C_L - V_L interactions increase partial unfolding of V_L. In parallel, C_L is potentially not substantially involved in fibril formation, leading to shorter fibrils and lower fibril concentrations for LC compared to V_L. For LC R61A S40P it can be concluded that the loss of structural integrity compared to LC R61A and LC wt resulted in a decreased conformational stability and accelerated fibril formation, most likely by reinforcing the underlying molecular mechanism found for V_L R61A.

In summary, potential interactions between V_L and C_L in the full-length LC are responsible for an increased fibril formation propensity of LC I2E and LC R61A. The facilitating residues in V_L are in the α-helix containing Glu81 and Asp82 as well as the loop comprising S40. The interface of C_L is formed by a loop including Asp165, Ser168, Lys169A and other potential polar interacting residues. Thus, further structural studies are necessary to determine the impact of each single residue on the structural integrity of the LC and the fibril formation propensity.

This study reveals new insights in the interplay of V_L and C_L as another key factor to understand the emergence of fibrils and the progress of AL amyloidosis.

5. The role of C-terminal residues for the stability of a murine IgG1 C_H2 domain

The structural elements determining antibody domain stability are not sufficiently understood. For the V_L domain, Nokwe et al. showed that the nature of the terminal residues is of importance for its stability and integrity (Nokwe et al., 2014, 2015). To test the impact of C-terminal residues on an antibody domain that is followed by a linker and another Ig domain, the MAK33 IgG1 C_H2 domain was utilized. C_H2 variants in this study were extended by amino acids naturally occurring C-terminally of the MAK33 C_H2 domain, i.e. they connect it to the C_H3 domain (Buckel et al., 1987; Feige et al., 2009a) (Figure 31). In total, four variants were investigated, C_H2 wild type (henceforth referred to as "C_H2") and variants extended by one to four residues: C_H2-S, C_H2-SK and C_H2-SKTK. Major parts of this chapter were published in parallel to the preparation and acceptance of this thesis in the Journal of Biological Chemistry (Weber et al., 2018b).

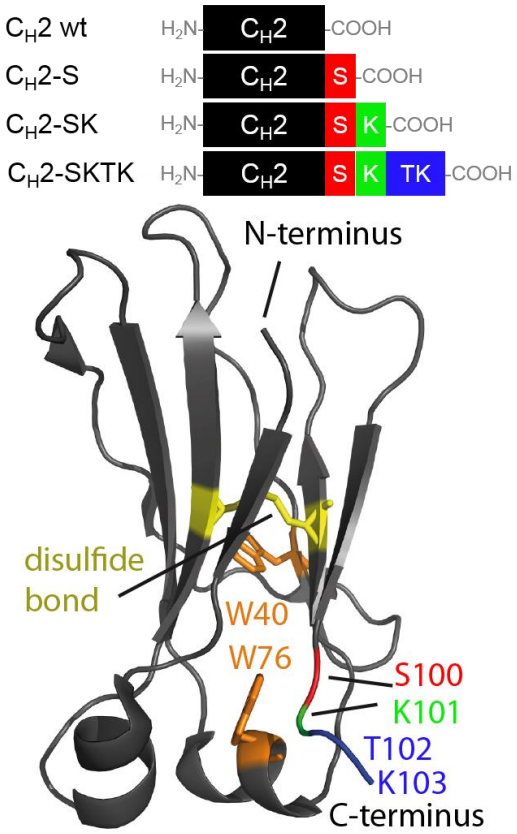

Figure 31. Scheme and crystal structure of C_H2 variants. *Top: Scheme of the used C_H2 variants. Bottom: Crystal structure (PDB: 3HKF) derived from the entire Fc fragment (comprising C_H2 and C_H3) with highlighted important features of the C_H2 domain: Trp40 and Trp76 (orange), intramolecular disulfide bond (yellow), Ser100 (red), Lys101 (green), Thr102 and Lys103 (blue). The N-terminus is in mature antibodies connected to the C_H1 domain, while the C-terminus is extended by the C_H3 domain.*

5.1. Results

To test whether the C_H2 variants are properly folded, far- and near-UV CD spectra were recorded (Figure 32). The far-UV spectra of all C_H2 variants exhibit two minima at approximately 210 nm and 230 nm and are highly similar, indicating identical secondary structures for all variants studied. Thus, up to four additional residues at the C-terminus do not lead to changes in the secondary structure of C_H2. However, near-UV CD spectra show a significant change in ellipticity, particularly between 275 nm and 295 nm compared to the wild type protein. These wavelengths are indicative of changes in the environment of aromatic residues, particularly Tyr and Trp (Boxer et al., 2004; Kelly et al., 2005). Since there are no Tyr residues in C_H2, the change in ellipticity at around 280 nm suggests differences in the microenvironment of the Trp residues. There are two Trp residues present in C_H2, Trp40 and Trp76 (Figure 31). Trp40 is part of a β-strand in close proximity to the intramolecular disulfide bond between Cys24 and Cys84. Its side chain is buried in the hydrophobic core of C_H2. Trp76 is located in a solvent-exposed α-helix structurally close to the C-terminus (Feige et al., 2009a). Trp40 is highly conserved in all antibody domains, whereas Trp76 is exclusively conserved in Ig domains which are part of the Fc fragment (Feige et al., 2010).

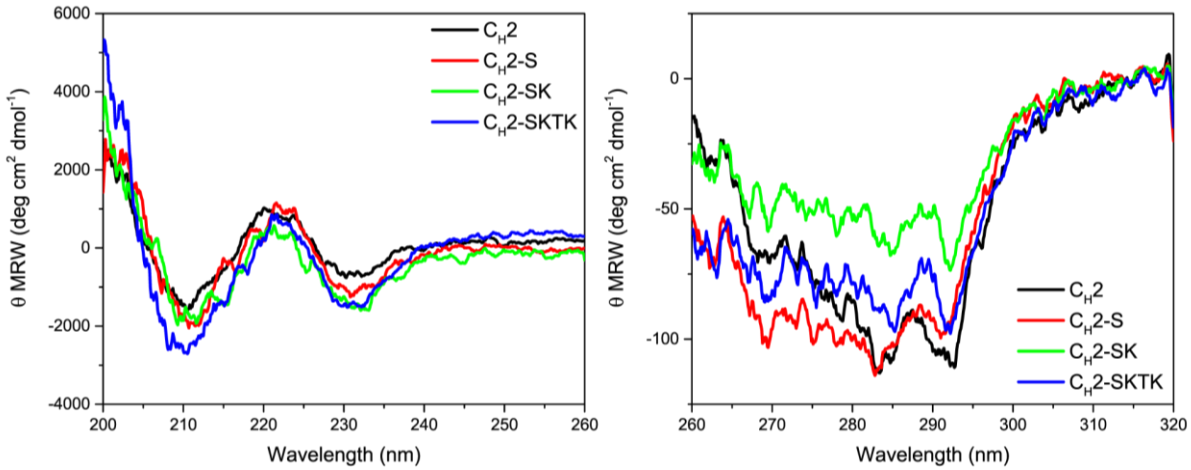

Figure 32. Far-UV and near-UV CD spectra. *Far (left) and near (right) UV CD spectra of C_H2 variants. All variants share the same wild-type like secondary structure. The tertiary structure, however, exhibits significant differences between the C_H2 variants with maximal distinction between 275 and 295 nm, which indicates changes in the tertiary structure especially for Trp residues.*

Taken together, all C_H2 variants exhibit wild type-like secondary structure, but differences in the tertiary structure compared to the wild type. The data indicate a change in the environment of the Trp residues, suggesting that either one or both Trp residues are located differently in the C_H2 extension variants.

C_H2 wild type is known to be almost exclusively monomeric (Feige et al., 2004). Upon characterization of the variants, it was tested by SEC-MALS whether the C_H2 variants were affected in their oligomeric state. All investigated variants shared a similar elution profile with one major and one minor peak (Figure 33). SEC-MALS proved all variants to be more than 93 % monomeric with a calculated molecular mass between 11.1 and 12.0 kDa for the monomers (Table 13). Since the oligomer concentration was too low, it was not possible to determine the molecular mass of the minor peak. A monomer / oligomer ratio for the C_H2 variants between 100 % / 0 % and 94 % / 6 % was found. There was no correlation of the C-terminal extension and the oligomeric state; only C_H2 and C_H2-SKTK, the shortest and the longest variants tested, respectively, exhibited a small fraction of oligomers.

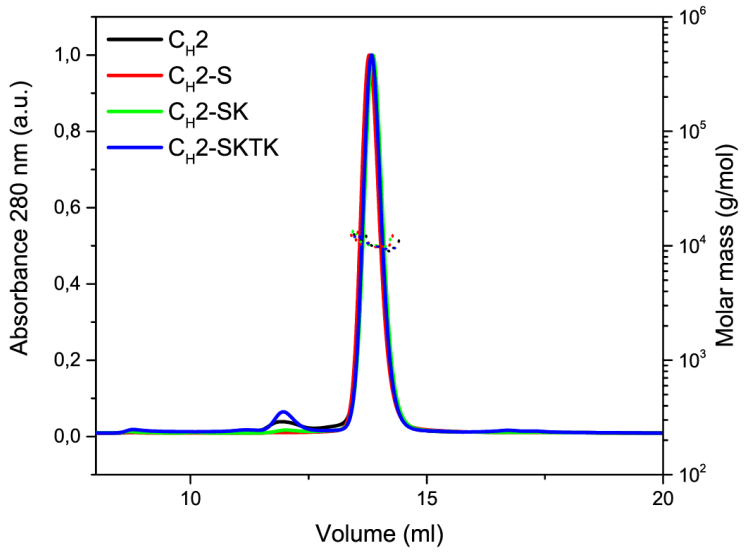

Figure 33. SEC-MALS chromatograms. *SEC-MALS experiments exhibit monomers as the prevalent species for all C_H2 variants. Only a small dimer fraction (< 7 %) was found. Hence, terminal extensions of the C_H2 domains do not alter the quaternary structure. 50 µl of 50 µM samples were injected.*

Variant	Molecular mass of prevalent species (kDa)	Relative distribution monomer : oligomer (%)
C_H2 wt	11.1 ± 0.3	95 : 5
C_H2-S	11.1 ± 0.2	100 : 0
C_H2-SK	12.0 ± 1.4	99 : 1
C_H2-SKTK	11.3 ± 0.3	94 : 6

Table 13. Oligomeric state of the C_H2 variants. *The molecular masses and relative oligomer distribution was determined by SEC-MALS in duplicates using 50 µM protein samples. Representative SEC-MALS chromatograms are shown in Figure 33. The error indicate the standard deviation of three technical replicates.*

To determine the conformational stability of the variants, thermal and chemical equilibrium unfolding transitions were performed (Figure 34). C_H2 wt starts to unfold already at ~ 20 °C with a melting temperature (T_m) of 44.4 ± 0.7 °C (Table 14). C_H2-S is similarly stable, revealing a T_m of 43.6 ± 0.4 °C. Surprisingly, C_H2-SK exhibits a significantly higher T_m of 58.6 ± 0.2 °C. Hence, C_H2-SK is ~ 14 °C more stable than C_H2 and C_H2-S. Thus, this strong increase in stability is accomplished by adding a single Lys residue (Lys101) which is part of the native linker connecting C_H2 with C_H3 in the full-length antibody. C_H2-SKTK, comprising two more linker residues, showed a T_m of 60.1 ± 0.2 °C, which represents an additional stabilization of ~ 1.5 °C compared to C_H2-SK. All spectra showed similar unfolding cooperativity. To complement the thermal transitions, chemical equilibrium unfolding was measured by probing intrinsic tryptophan fluorescence intensity in the presence of increasing concentrations of guanidinium chloride (GdmCl). Disulfide bonds are known to quench the intrinsic fluorescence of adjacent Trp residues (here Trp40, Figure 31) in 'Ig fold' proteins (Goto and Hamaguchi, 1979). Thus, unfolding results in an increase in fluorescence intensity. In agreement with the literature, C_H2 fluorescence emission increased upon denaturation (Feige et al., 2004). The experiments revealed a low stability (transition midpoint ($D_{1/2}$) 1.14 ± 0.03 M) for C_H2, whereas C_H2-SK and C_H2-SKTK were similarly stable with a $D_{1/2}$ of 1.62 ± 0.04 M and 1.65 ± 0.03 M, respectively (Figure 34 and Table 14). In contrast to thermal transitions, C_H2-S was less stable against chemical unfolding ($D_{1/2}$ 0.92 ± 0.92 M) than C_H2. For all variants, the order of stability is the same as observed in thermal unfolding experiments. However, chemical unfolding revealed a more pronounced difference between C_H2 and C_H2-S. Also, the chemical unfolding transitions feature similar cooperativities. Among all tested C_H2 variants, C_H2-SKTK showed the highest stability followed by C_H2-SK. As observed in the thermal unfolding transitions, the presence of Lys101 resulted in a strong increase in conformational stability, proving that Lys101 makes a strong contribution to the integrity of the C_H2 domain. Additional residues of the C_H2 - C_H3 linker (Thr102 and Lys103) increase thermal and chemical resilience slightly further. In contrast, Ser100 without Lys101 seems to slightly decrease the conformational stability.

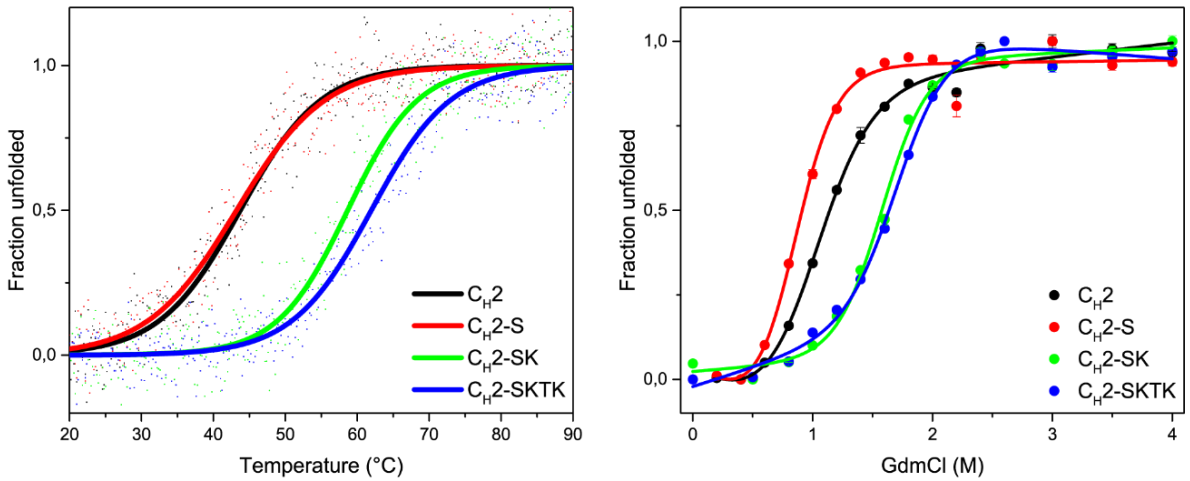

Figure 34. Stability of C_H2 variants. *Thermal (left) and chemical equilibrium (right) unfolding transitions of C_H2 variants. Thermal unfolding transitions were followed by CD spectroscopy. For chemical denaturation, aliquots of the C_H2 variants were incubated for at least 12 hours with increasing concentrations of GdmCl and monitored by intrinsic tryptophan fluorescence. Solid lines in thermal unfolding transitions represent the Boltzmann fit of the monitored data to determine the melting temperature (T_m), while chemical unfolding transitions were analyzed employing a two-state-model as described in the methods chapter.*

Variant	T_m (°C)	$D_{1/2}$ (M)
C_H2 wt	44.4 ± 0.7	1.14 ± 0.03
C_H2-S	43.6 ± 0.4	0.92 ± 0.02
C_H2-SK	58.6 ± 0.2	1.62 ± 0.04
C_H2-SKTK	60.1 ± 0.2	1.65 ± 0.03

Table 14. Stability of the C_H2 variants. *Melting temperatures (T_m) were determined by CD thermal unfolding. Chemical unfolding midpoints ($D_{1/2}$) derived from GdmCl equilibrium unfolding experiments, using intrinsic Trp fluorescence as a probe. T_m and $D_{1/2}$ represent the mean of three and two technical replicates, respectively. Representative thermal and chemical transitions are shown in Figure 34.*

To study the local environment of the Trp residues in more detail, intrinsic tryptophan fluorescence emission spectra were recorded. Different emission intensities as well as shifted wavelengths of the maximum signal between the natively folded C_H2 variants were detected (Figure 35 left, solid lines). C_H2-S and C_H2-SKTK display the highest intensities, followed by C_H2-SK and C_H2. Since the fluorescence intensity is correlated to intrinsic quenching and exposure of the Trp residues, variations most likely reveal differences in the local environment. C_H2 and C_H2-S (both 338 nm) as well as C_H2-SK and C_H2-SKTK (both 331 nm) share the same maximum of the emission peak wavelength, respectively. Variants containing Lys101 (C_H2-SK and C_H2-SKTK) exhibit a pronounced blue shift of ~ 8 nm which is characteristic for a more hydrophobic surrounding of Trp (Vivian and Callis, 2001). Blue shifts and altered fluorescence intensities

indicate different microenvironments of one or both Trp residues for C_H2 variants. Since the two Trp residues are probes for the hydrophobic core (Trp40) and for a solvent-exposed α-helix close to the C-terminus (Trp76) (Figure 31), the structural changes can be assigned to the inner protein core and the α-helix. However, using this technique it is not possible to discriminate between the effects of the two Trp residues. After equilibrium unfolding in the presence of GdmCl, all C_H2 variants exhibit the same Trp spectra including intensities (Figure 35 left, dotted lines).

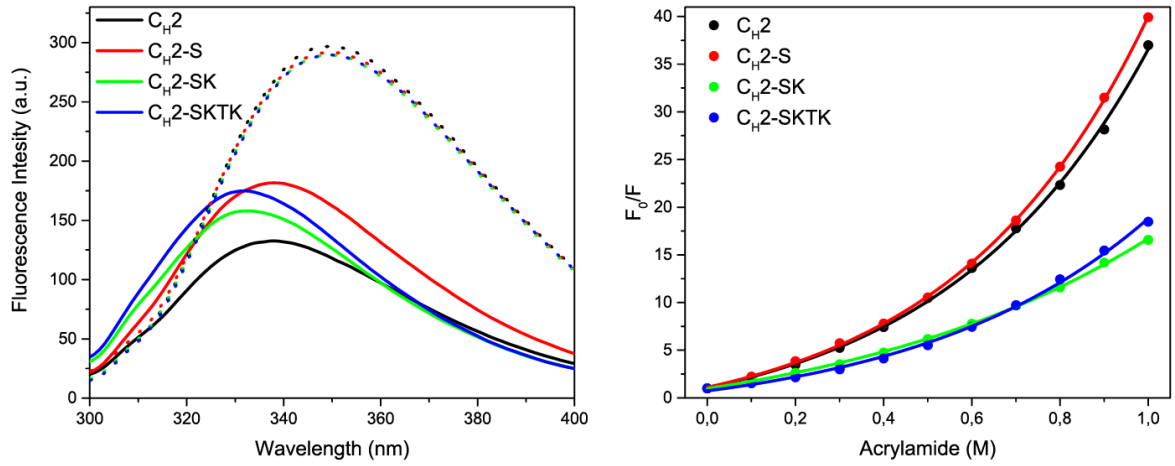

Figure 35. Conformational differences of C_H2 variants. *Left: Intrinsic tryptophan fluorescence of both, Trp40 and Trp76, probes conformational changes. All folded C_H2 variants (solid lines) display different Trp fluorescence intensities at the same concentration (7µM). For comparison, all variants show identical spectra when unfolded using 3.8 M GdmCl (dotted lines). Right: Stern-Volmer plots show different acrylamide quenching profiles for C_H2 variants. Data were analyzed using an exponential fit.*

To gain further insight in conformational changes within the C_H2 variants, the intrinsic Trp fluorescence was quenched using acrylamide (Figure 35 right). This allowed to evaluate the accessibility of both Trp residues and therefore of the hydrophobic core and the α-helix close to the C-terminus. C_H2 and C_H2-S featured a strong exponential relation between fluorescence signal and acrylamide concentration. An acrylamide concentration of 0.1 M caused a ~ 2.2-fold Trp fluorescence quenching. For 1 M acrylamide, 37-fold and 40-fold lower Trp emission intensities for C_H2 and C_H2-S, respectively, were observed. Accordingly, C_H2-SK and C_H2-SKTK exhibit Stern-Volmer plots, which are flatter and less exponential than for the variants without Lys101. Interestingly, the presence of Lys101 at the C-terminus significantly changed the quenching intensity, indicating an altered accessibility for both Trp residues. Exponential Stern-Volmer correlations suggest distinct accessibility of the quencher for two fluorophores within one protein (Phillips et al., 1986; Tallmadge et al., 1989). In contrast, linear plots indicate equal accessibility of acrylamide for two tryptophanes. Thus, the less pronounced the

exponential character, the more similar are the dynamics for both Trp residues. The Trp residues of C_H2 variants containing Lys101 feature ~ 2-fold less quenching than C_H2 variants without Lys101, proving their lower accessibility for acrylamide. Most likely, this alteration is due to a more compact microenvironment of the Trp residues in the presence of Lys101. However, it is not possible to discriminate between the impact of Trp40 and Trp76 on the fluorescence emission. In summary, C_H2-SK and C_H2-SKTK experience locally distinct structural differences compared to wild type and C_H2-S, which is attributed to the presence of Lys101.

To reveal differences in protein dynamics and structure, hydrogen/deuterium-exchange (H/DX) experiments were performed followed by mass spectrometric analysis of C_H2 and C_H2-SKTK (Figure 36, Figure **37** and Figure **38**). Along the entire primary structure, C_H2 showed higher deuterium uptake than C_H2-SKTK for all time points. To focus on the H/DX level alterations between C_H2 and C_H2-SKTK, the relative uptake ratio (C_H2/C_H2-SKTK) was calculated and plotted against the residue number of the C_H2 sequence (Figure 38 left). For distinct regions, particularly the α-helix and its extension (residues 70 - 81) pronounced differences in the relative uptake level were observed (Figure 38 region (3)). These differences are mainly localized in structural vicinity of the C-terminal region of C_H2 (Figure 38 right). In contrast, no substantial effects for the N-terminal part were observed, but a slightly decreased H/D-exchange for the adjacent helix from residue 8 - 15 (Figure 38, region (1)). Particularly two regions, from Pro70 to Glu81 and surprisingly Trp40 revealed strongly decreased H/D-exchange levels for C_H2-SKTK compared to wild type (Figure 38, region (2) and (3)). The affected α-helix (Met72 - Asn78) in proximity to the C-terminus is sequentially located between the two Ig fold-characteristic β-strands e and f, which are both defining among others (β-strands B and C) the hydrophobic core (Bork et al., 1994; Wang, 2013). Thus, both β-sheets making up the β-sandwich-like Ig fold in constant domains (CFG and ABED) experience decreased dynamics due to the presence of Lys101. Moreover, the very C-terminus (Figure 38, region (4)) exhibits a decreased H/D-exchange slightly more pronounced than the first helix (residues 8 - 15). Among the residues which experienced altered H/D-exchange are Trp40 and Trp76. In summary, the presence of SKTK at the C-terminus impacts the dynamics and local structure of a number of secondary structure elements. By correlating the H/DX data with unfolding transitions (Figure 34), it is possible to map the effect to Lys101 since this residue is responsible for the strong increase in conformational stability. Important Ig fold-structures,

which are part of the hydrophobic core and connected by an α-helix, are affected by the C-terminal presence of Lys101, indicating a more compact and protected protein core.

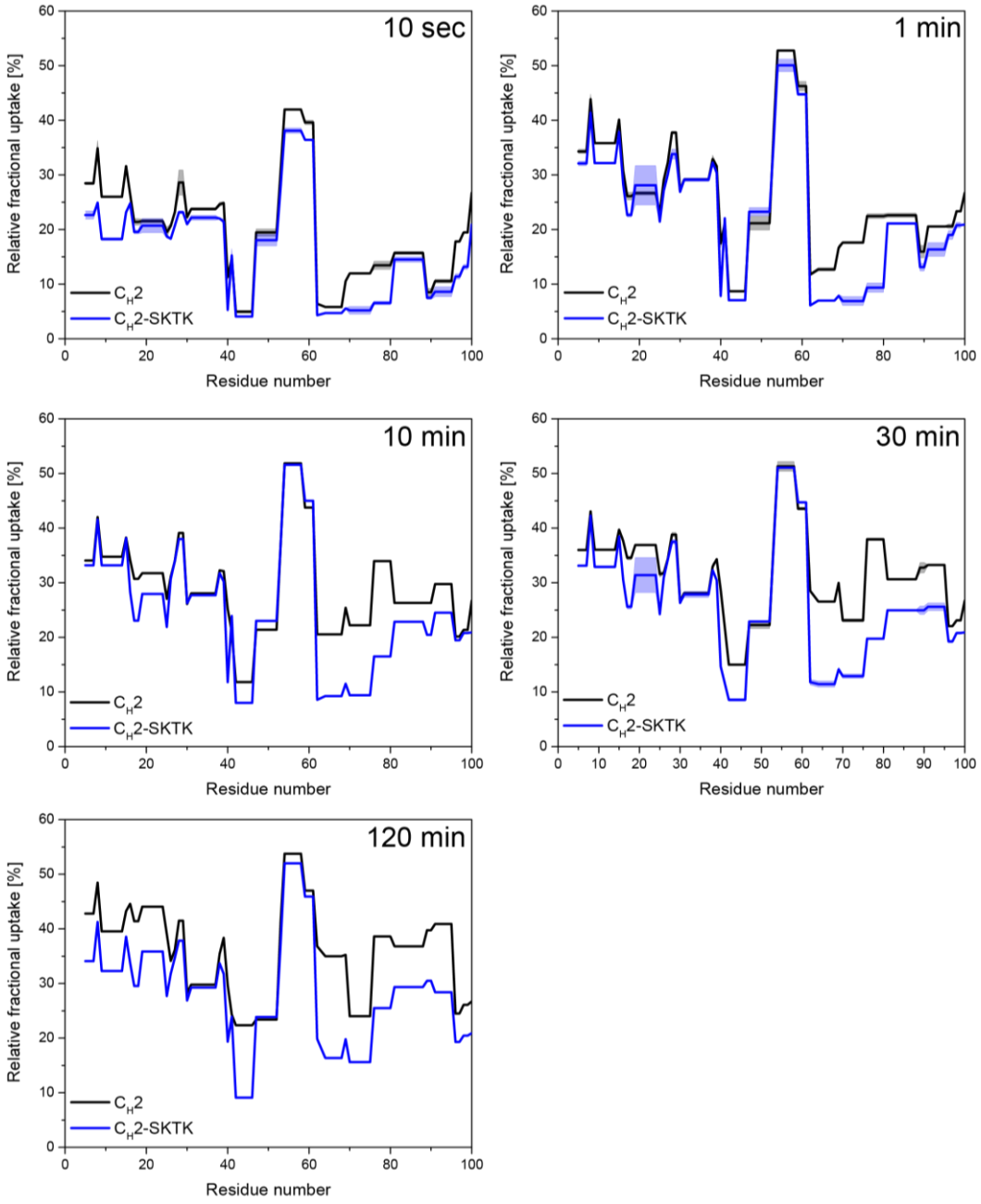

Figure 36. H/DX-derived relative fractional uptakes. *Relative fractional uptake of C_H2 (black) and C_H2-SKTK (blue) for the indicated time points of hydrogen/deuterium-exchange quenching.*

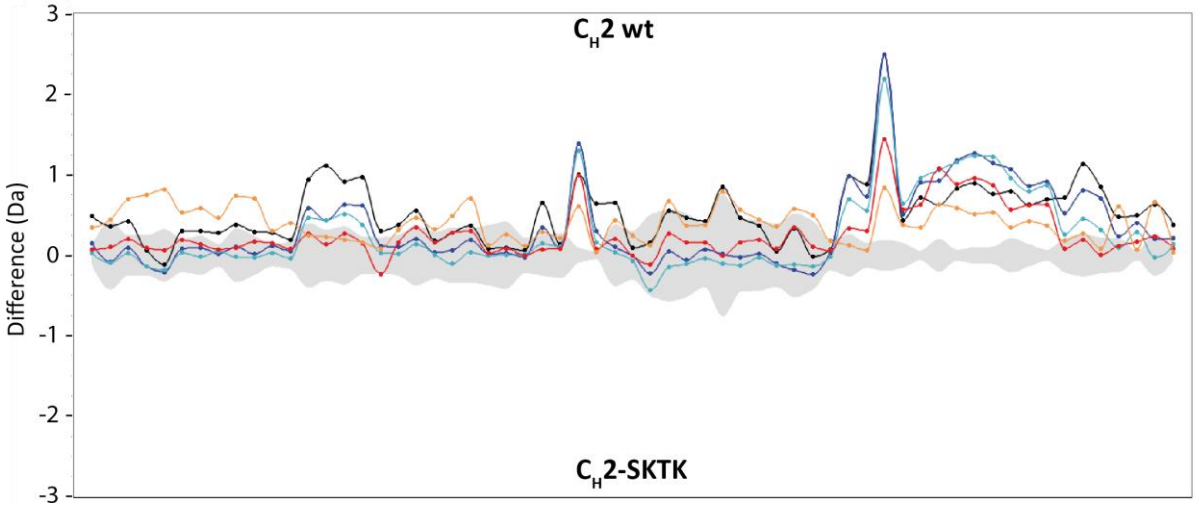

Figure 37. H/DX-derived butterfly plot for C_H2 wt and C_H2-SKTK. *The butterfly plot shows the absolute difference in molecular mass after H/D-exchange compared for C_H2 wt and C_H2-SKTK. Positive and negative values indicate higher uptake for C_H2 wt and C_H2-SKTK, respectively. Colors indicate the time of H/D-exchange quenching for 10 sec (orange), 1 min (red), 10 min (cyan), 30 min (blue) and 120 min (black). The grey area represents the error for mass spectrometry analysis.*

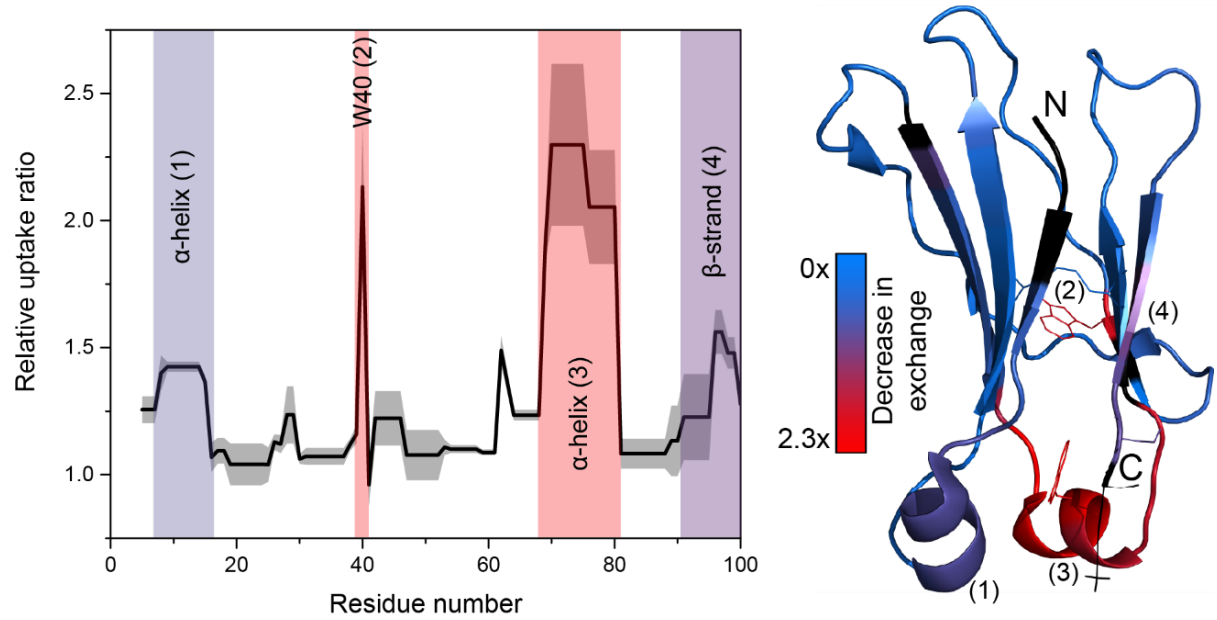

Figure 38. H/DX experiments to probe protein dynamics. *Left: Relative fractional uptake ratio for C_H2/C_H2-SKTK. The colored and numbered areas indicate protein regions with substantial H/D-exchange. Data represent the mean of two technical replicates for each variant; black "shadows" indicate standard deviation. Right: Relative uptake ratio for C_H2/C_H2-SKTK displayed on the crystal structure of C_H2 (PDB: 3HKF). Color gradient from blue (no difference) to red (2.3-fold decrease in exchange relative to C_H2) indicates structural and dynamical alteration between C_H2 and C_H2-SKTK. Black: no H/DX data available for these residues. H/DX experiments for the shown data were quenched after 10 s of exchange.*

^{1}H, ^{15}N HSQC NMR experiments as well as ^{1}H, ^{13}C, ^{15}N triple resonance experiments were performed and processed by Matthias Brandl (group of Prof. Bernd Reif, TUM) using wild type C_H2 and the variant EV-C_H2-SK. For the NMR analysis, the variant EV-C_H2-SK was employed, since the variant C_H2-SK shows peaks doubling for some residues, which is most likely due to insufficient oxidization of the intramolecular disulfide bond (data not shown). However, the N-terminal extension of the protein by two residues (Glu-Val) was found to have no impact on the overall conformation of the C_H2 domain (Figure 39, Figure **40**, Figure **41** and Figure **42**) Additionally, the conformational stability of EV-C_H2-SK is not substantially changed by the introduction of two N-terminal residues (Figure 43). Of notice, Glu and Val occur naturally N-terminal of C_H2 connecting it to C_H1 in a mature full-length antibody heavy chain.

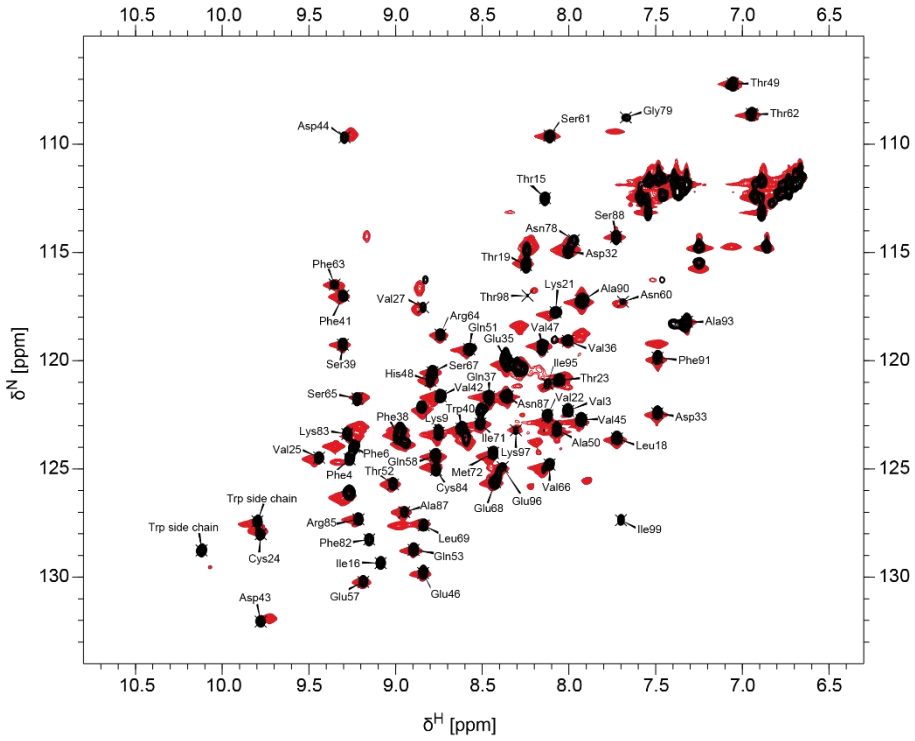

Figure 39. ^{1}H, ^{15}N HSQC spectra of C_H2 (black) and C_H2-S (red). *Resonance assignments for C_H2. It can be seen that the introduction of a C-terminal serine only induces minor changes in the spectrum. Residue numbering according to C_H2 wt. Data recorded and processed by Matthias Brandl (group of Prof. Bernd Reif, TUM).*

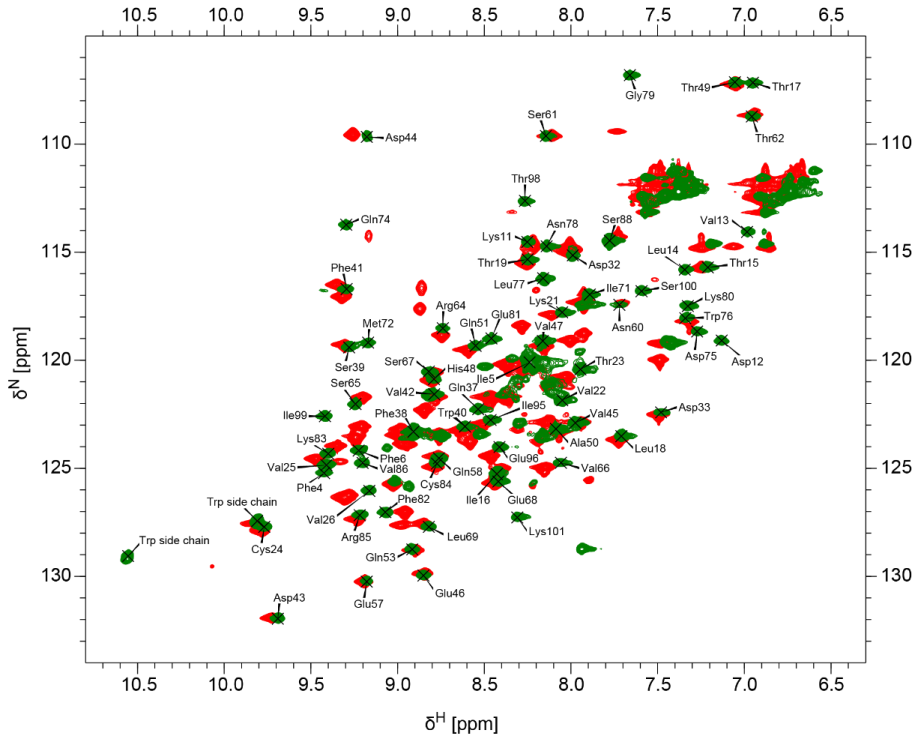

Figure 40. ^{1}H, ^{15}N HSQC spectra of EV-C_H2-SK (dark green) and C_H2-S (red). *Resonance assignments for EV-C_H2-SK. Marked changes can be observed upon the introduction of a C-terminal lysine. EV-C_H2-SK has been employed for NMR analysis since C_H2-SK shows peak doubling for some residues. Residue numbering according to C_H2 wt. Data recorded and processed by Matthias Brandl (group of Prof. Bernd Reif, TUM).*

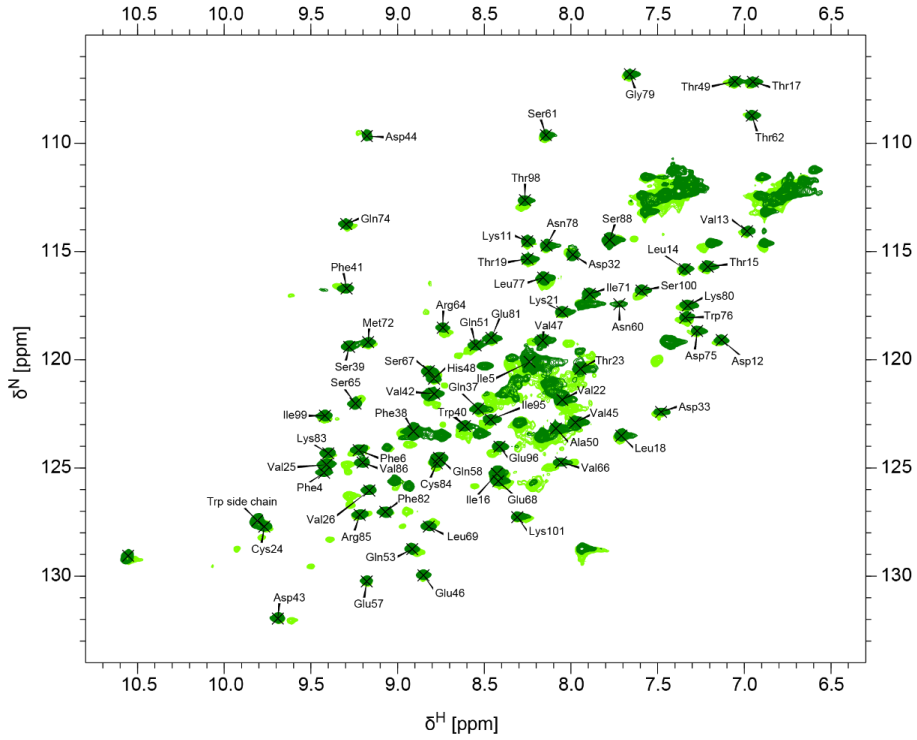

Figure 41. ^{1}H, ^{15}N HSQC spectra of EV-C_H2-SK (dark green) and C_H2-SK (light green). *Resonance assignments for EV-C_H2-SK. Apart from the removal of doubled peaks, the N-terminal extension does not induce major spectral changes. Residue numbering according to C_H2 wt. Data recorded and processed by Matthias Brandl (group of Prof. Bernd Reif, TUM).*

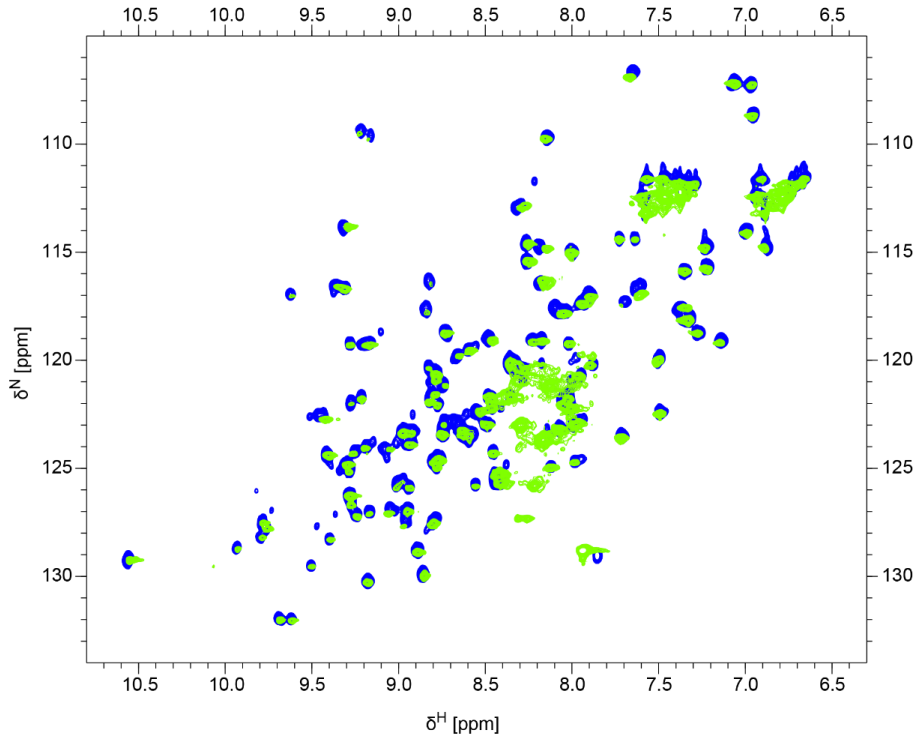

Figure 42. ^{1}H, ^{15}N HSQC spectra of C_H2-SK (light green) and C_H2-SKTK (blue). *As for Figure 41, the spectrum of C_H2-SK shows peak doubling and was therefore not further employed for NMR chemical shift analysis, but instead the variant EV-C_H2-SK. Residue numbering according to C_H2 wt. Data recorded and processed by Matthias Brandl (group of Prof. Bernd Reif, TUM).*

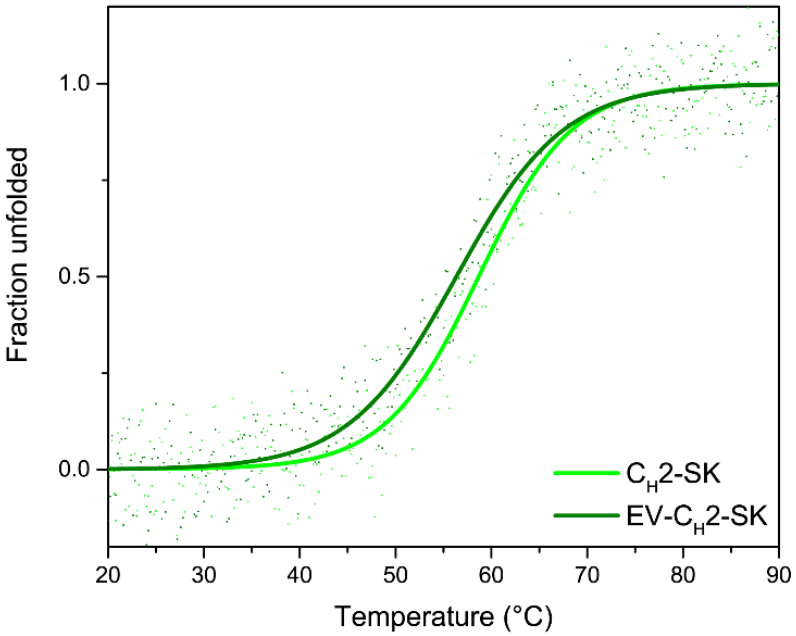

Figure 43. Thermal unfolding transitions of C_H2-SK (light green) and EV-C_H2-SK (dark green). *Thermal unfolding transitions were followed by CD spectroscopy. Both variants show a similar conformational stability, which is decreased for EV-C_H2-SK by ~ 2°C. Solid lines in thermal unfolding transitions represent the Boltzmann fit of the monitored data to determine the melting temperature (T_m).*

Differences in backbone amide ^{1}H and ^{15}N chemical shifts for the two proteins C_H2 and the variant EV-C_H2-SK are indicative for conformational differences (Figure 44). In addition to residues in the C-terminal region (residues 95 - 99), we observe no assignable resonances for most residues located in the α-helices (residues 10 - 14 and 72 - 78) in C_H2 wt. The appearance of assignable resonances for these residues in the extended variant indicates changes in dynamics. This suggests that these three regions change their mode of interaction upon the presence of the Lys101. Consistent with that, H/DX experiments identified all three regions to be impacted by Lys101 (Figure 38). Moreover, intense chemical shift changes for residues 15 and 16 close to the second α-helix were found (Figure 38, region (1)), which showed less intense H/D-exchange.

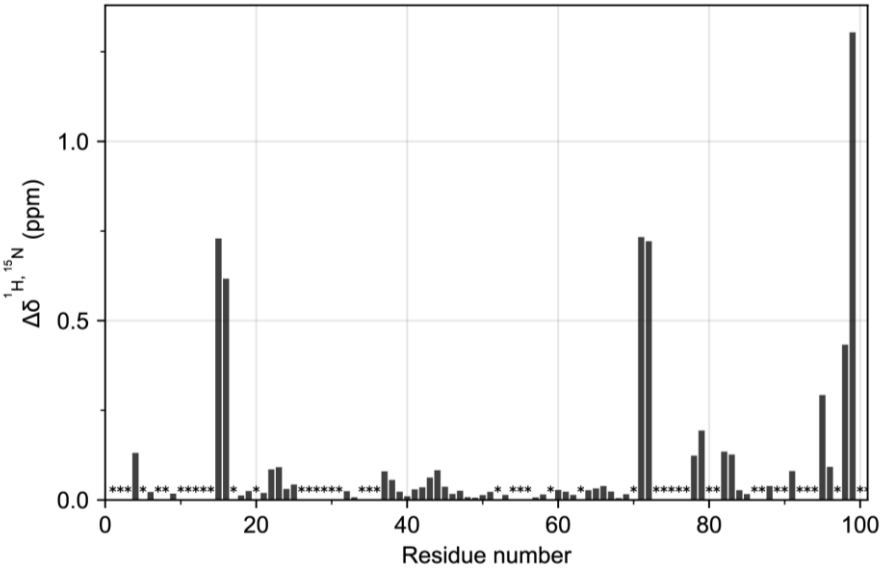

Figure 44. ^{1}H, ^{15}N chemical shift differences observed for C_H2 and EV-C_H2-SK. *Most substantial changes in chemical shifts occur in the very C-terminal part of the protein, while most residues of the two α-helices (10 - 14 and 72 - 78) could not be assigned for C_H2 wt. Residue numbering according to the C_H2 wt sequence. Asterisks indicate unassigned residues. Data recorded and processed by Matthias Brandl (group of Prof. Bernd Reif, TUM).*

Analysis of the $^{13}C^{\alpha}$ chemical shift of the two proteins yields information on changes in the local secondary structure propensity (Spera and Bax, 1991; Wishart and Sykes, 1994). Positive $\Delta^{13}C^{\alpha}$ chemical shift differences [δ(experimental) - δ(random coil)] are indicative for formation of helical structure, while negative shift differences suggest the population of β-sheet structure. Comparing the two C_H2 variants, helical propensity for residues 10 - 13 and 73 - 80 in EV-C_H2-SK was determined, while respective residues could not be assigned in the wild type (Figure 45). This supports the finding that H/D-exchange is substantially decreased for these regions of C_H2-SKTK (Figure 38) and suggests that the α-helix visible in these regions in the

crystal structure (PDB: 3HKF) only form in presence of Lys101. Of notice, the crystal structure of the C_H2 domain derived from the entire Fc fragment, which includes Lys101 and C_H3. In addition, the negative secondary chemical shifts observed for the C-terminal region (residues 95 - 99) suggest the formation of a β-strand structure in EV-C_H2-SK that is not present in the wild type. These additional secondary structural elements, α-helices and β-strand contribute to the overall stabilization of the protein observed after the addition of Lys101 (Figure 34).

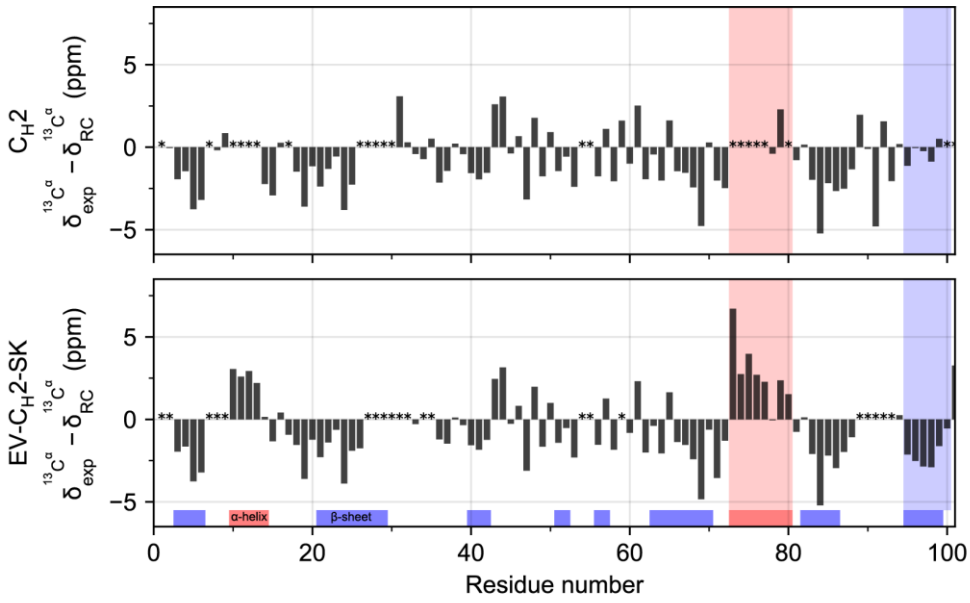

Figure 45. ^{13}Cα chemical shift differences δ(^{13}C, experimental) - δ(^{13}C, random coil). *Random coil chemical shifts were taken from Wishart and Sykes, 1994. Positive values indicate propensity for an α-helix, negative values for a β-strand. The top panel shows the chemical shift differences for wild type C_H2, the bottom panel for EV-C_H2-SK. Blue and red bars at the bottom of the lower panel show the secondary structure elements as indicated in the crystal structure (PDB: 3HKF), shaded areas highlight the major changes in secondary structure propensity. The changes in secondary chemical shifts for residues 11 - 14 and 73 - 80 indicate the formation of two α-helices in the extended variant which were not present in the wild type. In addition, the C-terminal region of EV-C_H2-SK shows propensity to form a β-strand, while the same region in the wild type appears more likely to occupy a random coil conformation. Asterisks indicate unassigned residues. Data recorded and processed by Matthias Brandl (group of Prof. Bernd Reif, TUM).*

To understand the effect of Lys101 in further structural detail, molecular dynamics (MD)-simulations were performed by María Daniela Pulido Cendales (group of Prof. Martin Zacharias, TUM) on the C_H2 wild type, and the extended variants C_H2-S and C_H2-SK in explicit solvent. On a time scale of 1 µs the structures stayed close to the X-ray structure starting geometry with a root mean square deviation (RMSD) of the protein backbone of < 2.5 Å (Figure 46) and no signs of unfolding. However, in the case of the C_H2 wild type and C_H2-S slight shifts in conformation and increased fluctuations were observed compared to C_H2-SK.

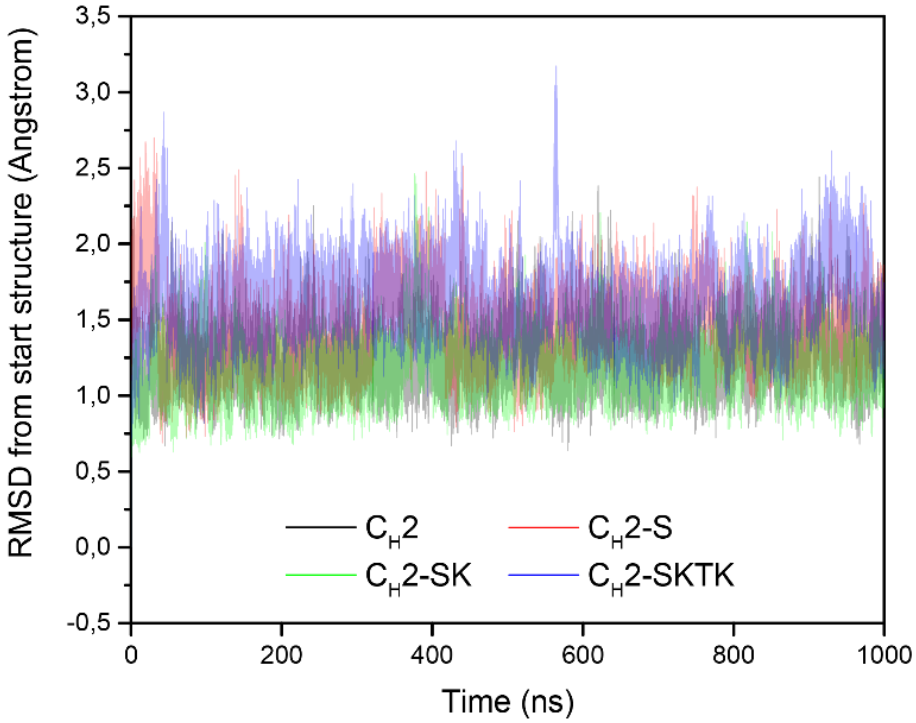

Figure 46. Root mean square deviation (RMSD) from corresponding starting structures during MD-Simulations. *The coordinate RMSD was calculated over a simulation period of 1 µs at 300 K with respect to the starting structure of each of the variants. A RMSD < 2.5 Å indicates a high stability of the conformations during the performed MD-Simulations. Data recorded and processed by María Daniela Pulido Cendales (group of Prof. Martin Zacharias, TUM).*

In order to correlate the simulations to experimentally observed differences in H/DX (Figure 38), the mean solvent accessibility of each residue was calculated during the whole production simulations (Figure 47). Since the H/D-exchange probability is expected to depend on the solvent accessibility, this quantity is well suited for a comparison to the experiment. Indeed, the simulations predict an increased mean solvent accessibility of the region between residues 65 - 80 which corresponds to the short α-helical segment in the vicinity of the C-terminal Lys101 residue. The absence of this stabilizing interaction with the helix dipole in case of the C_H2 wild type and C_H2-S likely influences also the conformation and flexibility of the helical segment and in turn causes an increased H/D-exchange for C_H2 and C_H2-S. Interestingly, the simulations indicate that this is mostly due to a conformational shift and not necessarily due to increased mobility (fluctuations of residues are similar for all C_H2 variants, Figure 46).

To explain the reduced stability of the C_H2 wild type and C_H2-S compared to the C_H2-SK variant, umbrella sampling (US) free energy simulations along a reaction coordinate that results in dissociation of the C-terminal segment from the folded β-strand to form fully solvated conformation was performed. Such a transition indicates the free energy contribution of the

C-terminal fragment to stabilize the folded structure (compared to an unfolded solvated form). The difference between the calculated free energy for the variants vs. the wild-type is a measure of the loss in stabilization due to the presence of the folded C-terminal strand.

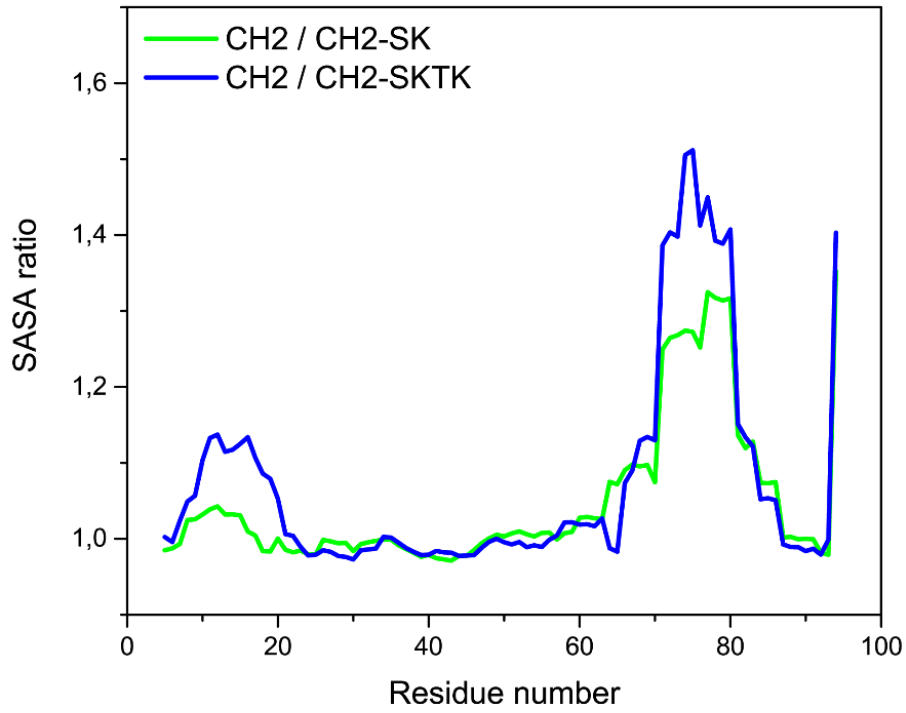

Figure 47. Mean solvent accessibility during MD simulations. *The ratio in mean solvent accessibility (SASA) of residues in C_H2 wild type relative to C_H2-SK (green) and relative to C_H2-SKTK (blue) was calculated for each residue over a simulation period of 1 µs at 300 K. A SASA ratio > 1 indicates increased mean accessibility relative to C_H2-SK or relative to C_H2-SKTK, respectively (obtained as sliding window average over 10 consecutive residues). María Daniela Pulido Cendales (group of Prof. Martin Zacharias, TUM).*

The simulations predict a similar reduced stabilization of C_H2 and C_H2-S relative to C_H2-SK and C_H2-SKTK, respectively, by ~4 kcal·mol^{-1} (compare the difference in plateau free energy values reached in case of C_H2 and C_H2-S vs. C_H2-SK and C_H2-SKTK variants upon dissociation of the C-terminal strand in Figure 48). The significant free energy difference in favor of the C_H2-SK and C_H2-SKTK variants agrees qualitatively well with the observed differences in melting temperature of the two variants vs. wild type. Inspection of snapshots from different US windows indicates that the origin of the free energy difference seems to be indeed the interaction of the Lys101 side chain at the C-terminus with the adjacent helical segment (residues 72 - 81). This interaction is still visible in the intermediate US windows that result in the largest free energy increase (Figure 48). At the same US intervals the C-terminus of the wild type or the C_H2-S variant are already dissociated. Hence, the presence of the Lys101-helix dipole interaction results in an increase of the PMF and stabilizes the folded form significantly.

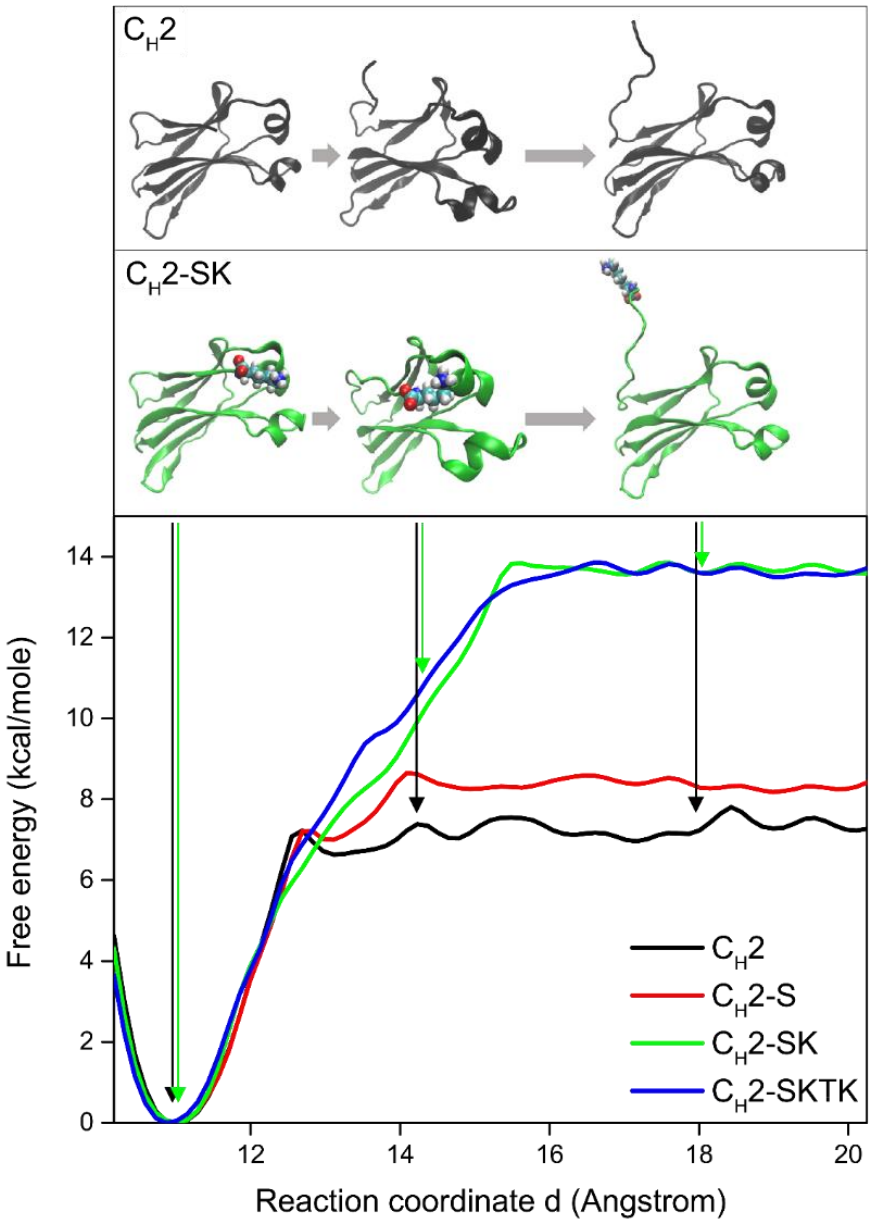

Figure 48. Calculated free energy change upon dissociation of the C-terminal β-strand from its placement in the folded structure to a solvated unfolded state. *The free energy change (potential-of-mean force) was calculated along a reaction coordinate that corresponds to the center of mass distance of backbone atoms of the C-terminal segment (delimited by residues 94 - 98) and the backbone atoms of the protein (delimited by residues 1 - 91). Representative snapshots taken in the folded, intermediate and dissociated states of the C-terminal segment are indicated color-code of the cartoon representation corresponds to the line color in the plot). The C-terminal Lys residue in C_H2-SK is indicated as van der Waals (vdW) sphere representation and still interacts with the helical segment (residues 72 - 81) in the intermediate state (at which the C-terminus of the C_H2 wild-type is already fully dissociated). María Daniela Pulido Cendales (group of Prof. Martin Zacharias, TUM).*

5.2. Discussion

In this study, the impact of C-terminal residues of a murine IgG1 C_H2 domain on its conformational stability and protein dynamics was determined. Importantly, the investigated terminal residues (Ser-Lys-Thr-Lys) are not part of any defined secondary structure elements as shown in the crystal structure (Figure 31, PDB: 3HKF). The addition of these naturally occurring residues resulted in a stability increase of ~ 15 °C. By assessing serially extended C_H2 variants from one to four residues, it was possible to attribute this strong improvement of conformational stability to a single residue: Lys101. All C_H2 extension variants are properly folded, revealing a slightly changed tertiary structure, which was attributed to a different microenvironment of the two Trp residues. H/DX experiments exhibited significantly decreased dynamics for the α-helix connecting β-strands E and F. This α-helix links both β-sheets, which make up the characteristic Ig fold β-barrel (Feige et al., 2010). This suggests a more constrictive and compact hydrophobic core resulting in an overall higher protein integrity as indicated by acrylamide quenching. Therefore, the interaction between Lys101 and the nearby α-helix seems to shield the core from the solvent by spatially blocking cavities. In this context, the NMR experiments revealed that the important interacting α-helix is most likely not formed in the absence of Lys101. Thus, the polar interaction between the polypeptide backbone of Leu77 and Lys101 is required to initiate and maintain the folding of the α-helix (Figure 49). Moreover, for C_H2-SK the secondary structure propensity analysis indicates the formation of a C-terminal β-strand in the presence of Lys101. Therefore, the two secondary structure elements show a reciprocal dependence on Lys101, resulting in an increase of the overall conformational stability. In agreement with the experimental observation, the MD simulations predict a higher solvent accessibility especially in the α-helix region in the absence of Lys101 in line with the experimental H/DX and NMR results. Furthermore, free energy simulations indicate that the Lys101 interaction with the nearby α-helix makes a significant free energy contribution in favor of the folded state. Feige et al. showed that the conserved α-helical structures fold rapidly in C_L. This orientates the two β-sheets in a favorable way, inducing a higher degree of compactness (Feige et al., 2008). In this context, the interaction of Lys101 with the α-helix initiates the native folding pathway by assuring the proper orientation of the residues forming the hydrophobic core and the β-sheets, specifically β-strands E and F. Since C_L and C_H2 share a similar folding pathway, the conserved polar interaction between Lys101 and the polypeptide backbone of Leu77 indicates

a general stabilizing effect of Lys101 on the Ig fold, mediated by the α-helix residue Leu77 between β-strands E and F (Feige et al., 2010). Thus, not only the intramolecular disulfide bond is an important connection of both β-sheets, but the α-helix between strands E and F, stabilized by Lys101, contributes to the overall structural stability.

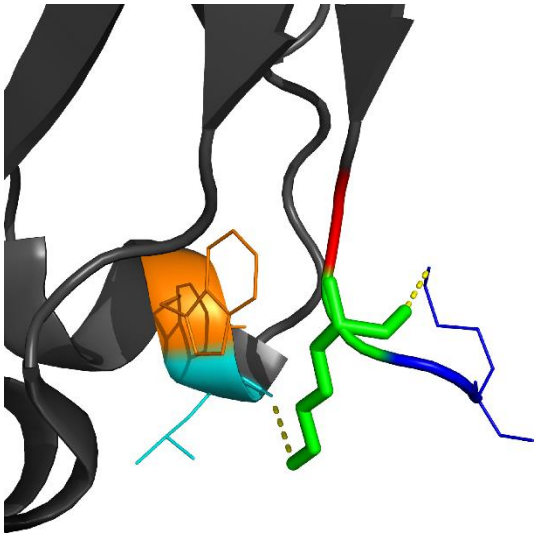

Figure 49. Polar interaction between Leu77 and Lys101. *As judged from the crystal structure (PDB: 3HKF, black), there are polar interactions (yellow dotted lines) between Lys101 (green) and the polypeptide backbone carbonyl of Leu77 (cyan). Important residues are highlighted: Trp76 (orange), Ser100 (red), Thr102 and Lys103 (blue).*

Sequential and structural alignment of other antibody domains showed that Lys101, or alternatively another positively charged residue like arginine at this position, and its interaction with the adjacent α-helix is conserved among all constant Ig domains (Figure 50), except C_H1 which is natively unfolded (Feige et al., 2009b). Interestingly, C-terminal charged residues of C_H1 interact with an α-helix of C_L when associated in the Fab fragment (PDB: 2QSC, 3OZ9). Thus, this structural feature seems to be evolutionarily favored to assure the proper initiation of folding and conformational stability. Traditionally, the formation of a hydrophobic core is seen as the main stabilizing interaction that determines the folding and stability of a globular protein. However, the conserved polar interaction between a positively charged side chain and an α-helix (dipole) at the surface of the C_H2 domain can also make a significant contribution to stability as shown by the present free energy simulations and stability studies. Consequently, the findings are most likely transferable to other constant Ig domains. A solvent-protected core, optimally oriented β-sheets and the presence of secondary structure elements like the interacting α-helix and a C-terminal β-strand contribute to a higher degree of integrity and explain the strong increase in conformational stability.

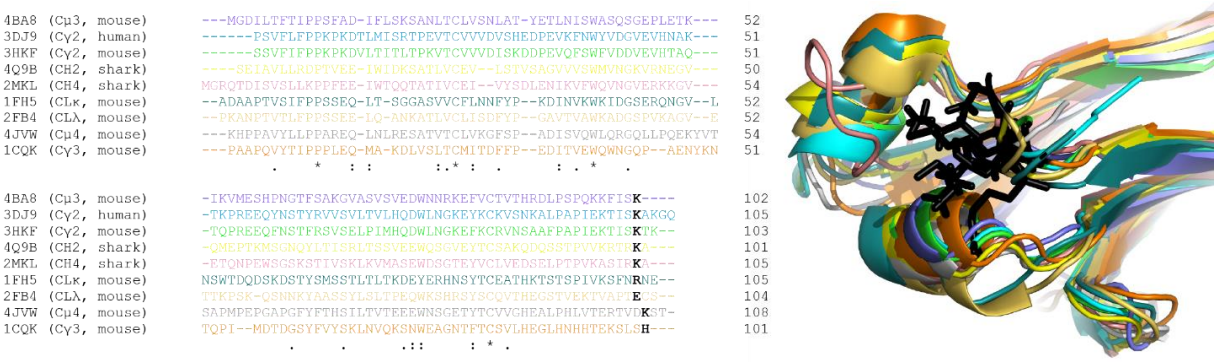

Figure 50. Sequential and structural alignment. *Alignment of nine different constant Ig domains derived from human, mouse and shark. Lysine or alternatively another large, charged residue is highly conserved at the C-terminus of constant Ig domains except C_H1. Black: C-terminal Lys in close proximity to the structurally important helix.*

The results raise the question how to determine immunoglobulin domain boundaries correctly. When the first Ig domain sequences were described, various individual domains were aligned, e.g. murine C_H1, C_H2, C_H3 and C_H4, derived from different Ig isoforms (IgM, IgD, IgG, etc.) (Kehry et al., 1979; Shimizu et al., 1981). This method revealed significant homology between the different domains such as conserved cysteines and tryptophanes along with other residues (Dayhoff, 1978). In parallel, it was found that every constant domain is encoded by an individual exon, which suggested a separation of the domains already on the genetic level (Sakano et al., 1979). Hence, the first and last residues of constant domains are determined by RNA splicing (Rabbitts et al., 1981). This strategy of defining Ig domain boundaries was supported by structural data (Poljak et al., 1973). In contrast to the constant domains, the structures of the V_H and V_L domains have been studied in detail to systematically assign framework regions (FR) and complementary determining regions (CDR), of which the latter ones bind antigens (Chothia et al., 1998; Kabat et al., 1984). As antibody folding and stability were primarily attributed to the hydrophobic core, the conserved disulfide bond and the characteristic β-strands, antibody fragment boundaries lost relevance (Bork et al., 1994; Feige et al., 2010; Goto and Hamaguchi, 1979, 1981; Liu and May, 2012). However, Nokwe et al. showed for a murine V_{LK} that the N- and C-terminal residues, i.e. Ile2 and Arg108 play a major role for the integrity of the domain (Nokwe et al., 2014, 2015). The use of antibody fragments as diagnostics and therapeutics together with the data of this study and other recent literature revives the question of domain boundaries (Gong et al., 2013b). During the early development of single chain variable fragments (scFv), the boundaries of V_L, V_H, Fc and C_H3 were randomly set by optimizing the position of the linker between both domains with

respect to the orientation of the two domains within the heterodimer (Alt et al., 1999; Bird et al., 1988). Biophysically optimized rather than genetically determined N- and C-termini might serve as a complementary method to increase the conformational stability of antibody fragments resulting in lower aggregation propensity, pharmacokinetic half-life and longer shelf life. This approach might contribute to more stable and reliable therapeutics and diagnostics. Therefore, for defining Ig domain boundaries, the impact of N- and C-terminal linker residues needs to be considered. Particularly large, charged residues are prone to facilitate ionic interactions. Thus, the Ig fold is not only maintained by intramolecular disulfides and the packing of the hydrophobic core. Ionic interactions of surface-exposed residues play an important role to facilitate and maintain proper folding of the conserved α-helix and the adjacent β-sheets as well as other secondary structure elements like the C-terminal β-strand, thereby protecting the hydrophobic core and assuring conformational stability.

6. Conclusions and perspectives

This study focused on the structural integrity of different antibody domains, relevant for gaining deeper understanding of the antibody domain architecture and expanding our knowledge of the life-threatening disease AL amyloidosis. The results might contribute to the development of stable and aggregation-resistant antibody therapeutics and diagnostics.

The first study (Chapter 3) employing the MAK33 full-length LC framework, yielded valuable information on the importance of the LC linker for the conformational stability of the entire protein as well as for the regulation of the relative orientation of the two constituent domains, V_L and C_L. It was found that the C_L domain is able to protect the V_L domain from amyloid fibril formation in the context of the full-length LC. As already shown for the individual C_L and V_L domains (Nokwe et al., 2015), the LC linker residue Arg108 is key to maintain conformational stability and integrity by facilitating salt bridges with C_L residues as well as managing the inter-domain contact between C_L and V_L. The latter interactions are responsible for the relative orientation of the two LC domains to each other, making the LC linker in the presence of Arg108 likely less accessible for proteolytic cleavage. Further investigation is needed to determine the role of the LC linker in the proteolytic cleavage of the LC in AL amyloidosis patients. Moreover, *in vivo* and *in cellulo* experiments to quantify the toxicity of amyloidogenic full-length LCs and their truncated fragments would be informative to determine the relative importance.

The study was extended to investigate known amyloidogenic mutations, namely I2E, S20N and R61A, in the context of the full-length LC framework (Nokwe et al., 2014, 2016), summarized in chapter 4. It was validated that C_L is able to suppress fibril formation of some full-length LCs. In contrast to that, two LC variants, LC I2E and LC R61A, revealed accelerated fibril formation compared to the corresponding V_L proteins. Thus, the presence of C_L can either inhibit or accelerate amyloid fibril formation of LC variants in a mutation-dependent manner. The latter one was is likely due to inter-domain interactions between C_L and V_L leading to a further destabilized V_L domain. Particularly the residues Ser40, Thr80, Ser168 and Lys169 are involved in these interactions, however their substitution did not entirely recover the lower amyloidogenic potential of V_L R61A. Most likely, other polar residues located nearby the aforementioned amino acids maintain the unfavorable inter-domain interactions. To gain deeper structural insight in the underlying mechanism and mode of interaction, further NMR

experiments employing the LC R61A double mutation variants are necessary. Moreover, MD simulation might reveal the role of the mentioned residues in this context. LC R61A variants containing more than one additional mutation might additionally provide valuable information. Differential scanning calorimetry (DSC) is an applicable technique to determine the stability of the individual domains within the full-length LC.

The third study focused on the constant heavy chain domain 2 (C_H2) and its substantial stabilization by the presence of naturally occurring C-terminal residues (Chapter 5). It was found that Lys101 is of major importance for the conformational stability of the domain. Interestingly, lysine residues or alternatively other large, charged residues are highly conserved at the C-terminus of antibody domains. The effect of the presence of Lys101 is the facilitation of two secondary structure elements, a C-terminal β-strand and a nearby α-helix, by introducing a dipole interaction with the α-helix. Both secondary structures are highly conserved in the Ig fold and affect the stability of the entire Ig domain. Thus, the results are most likely generally applicable for constant Ig domains and other proteins of the Ig superfamily, which all share a common fold. The results provide new insight in the architecture of Ig domains and support the importance of correctly defined domain borders as well as the key effect of terminal residues.

Taken together this work provides new mechanistical understanding of the immunoglobulin framework, particularly its stability and integrity. This might contribute to the future development of antibody-based therapeutics and diagnostics as well as defining properties of anti-amyloidosis drugs.

7. Abbreviations

For amino acids, the one and three letter code was used

AFS	Alternatively folded state
AL	antibody light chain
Aβ	Amyloid-beta peptide
CD	Circular dichroism
CDR	Complementarity determining region
C_H1	constant heavy chain domain 1
C_H2	constant heavy chain domain 2
C_H3	constant heavy chain domain 3
C_L	constant light chain
Cryo-EM	Cryo electron microscopy
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide
E.coli	*Escherichia Coli*
ER	Endoplasmic reticulum
Fab	fragment antigen binding
Fc	fragment crystallizable
FUV	Far-UV
Fv	Fragment variable
GdmCl	Guanidinium chloride
GOI	Gene of interest
H/DX	Hydrogen/deuterium exchange
HC	Heavy chain
HPLC	High pressure liquid chromatography
IDP	Intrinsically disordered proteins
Ig	Immunoglobulin
LB_0	Luria-Bertani medium
LC	Light chain
M&M	Metadynamic Metainference
MAK33	Monoklonaler Antikörper 33 (engl.: monoclonal antibody 33)
MALS	Multi angle light scattering
MD	molecular dynamics
MGUS	Monoclonal gammopathy of undetermined significance
MS	Mass spectrometry
NMR	Nuclear magnetic resonance
NUV	Near-UV
OD_{600}	Optical density at 600 nm wavelength
PBS	phosphate buffered saline
PDB	The protein database
PDB	The protein database
PPIase	Peptidylprolyl *cis-trans* isomerases
RDC	Residual dipolar couplings
RMSD	Root-mean-square deviation
rpm	Rounds per minute
SEC	Size exclusion column
SLIC	Sequence- and ligation-independent cloning

t_{50}	Time point when 50% of the maximal ThT signal was reached (d)
TEM	Transmission electron microscopy
TEM	Transmission electron microscopy
ThT	Thioflavin T
T_m	Melting temperature (°C)
US	Umbrella sampling
UV-Vis	Ultraviolet and visible light wavelengths
V_H	variable heavy chain
V_L	variable light chain
wt	wild type

8. References

Abascal, J.L.F., and Vega, C. (2005). A general purpose model for the condensed phases of water: TIP4P/2005. J. Chem. Phys. *123*.

Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., and Lindah, E. (2015). Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX *1–2*, 19–25.

Ahn, J., Skilton, S.J., and Yu, K. (2013). Hydrogen-deuterium exchange mass spectrometry: An emerging biophysical tool for probing protein behavior and higher-order structure. LCGC North Am. *31*, 464–471.

Alt, M., Müller, R., and Kontermann, R.E. (1999). Novel tetravalent and bispecific IgG-like antibody molecules combining single-chain diabodies with the immunoglobulin γ1 Fc or CH3 region. FEBS Lett. *454*, 90–94.

Anfinsen, C.B. (1973). Principles that govern the folding of protein chains. Science (80-.). *181*, 223–230.

Arnold, J.N., Wormald, M.R., Sim, R.B., Rudd, P.M., and Dwek, R.A. (2007). The Impact of Glycosylation on the Biological Function and Structure of Human Immunoglobulins. Annu. Rev. Immunol. *25*, 21–50.

Augustine, J.G., de la Calle, A., Knarr, G., Buchner, J., and Frederick, C.A. (2001). The Crystal Structure of the Fab Fragment of the Monoclonal Antibody MAK33. J. Biol. Chem. *276*, 3287–3294.

Baden, E.M., Sikkink, L.A., and Ramirez-Alvarado, M. (2009). Light chain amyloidosis - current findings and future prospects. Curr. Protein Pept. Sci. *10*, 500–508.

Baldwin, R.L. (1995). The nature of protein folding pathways: The classical versus the new view. J. Biomol. NMR *5*, 103–109.

Baldwin, A.J., Knowles, T.P.J., Tartaglia, G.G., Fitzpatrick, A.W., Devlin, G.L., Shammas, S.L., Waudby, C.A., Mossuto, M.F., Meehan, S., Gras, S.L., et al. (2011). Metastability of native proteins and the phenomenon of amyloid formation. J. Am. Chem. Soc. *133*, 14160–14163.

Bartlett, A.I., and Radford, S.E. (2009a). An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms. Nat. Struct. Mol. Biol. *16*, 582–588.

Bartlett, A.I., and Radford, S.E. (2009b). An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms. Nat. Struct. Mol. Biol.

Bateman, A., Eddy, S.R., and Chothia, C. (1996). Members of the immunoglobulin superfamily in bacteria. Protein Sci. *5*, 1939–1941.

Bence, N.F., Sampat, R.M., and Kopito, R.R. (2001). Impairment of the ubiquitin-proteasome system by protein aggregation. Science (80-.). *292*, 1552–1555.

Bengtén, E., Wilson, M., Miller, N., Clem, L.W., Pilström, L., and Warr, G.W. (2000). Immunoglobulin Isotypes: Structure, Function, and Genetics. In Origin and Evolution of the Vertebrate Immune System, (Berlin, Heidelberg: Springer, Berlin, Heidelberg), pp. 189–219.

van den Berg, B., Ellis, R.J., and Dobson, C.M. (1999). Effects of macromolecular crowding on

protein folding and aggregation. EMBO J. *18*, 6927–6933.

Best, R.B., and Mittal, J. (2010). Protein simulations with an optimized water model: Cooperative helix formation and temperature-induced unfolded state collapse. J. Phys. Chem. B *114*, 14916–14923.

Bird, R., Hardman, K., Jacobson, J., Johnson, S., Kaufman, B., Lee, S., Lee, T., Pope, S., Riordan, G., and Whitlow, M. (1988). Single-chain antigen-binding proteins. Science (80-.). *242*, 423–426.

Blancas-Mejía, L.M., and Ramirez-Alvarado, M. (2016). Recruitment of Light Chains by Homologous and Heterologous Fibrils Shows Distinctive Kinetic and Conformational Specificity. Biochemistry *55*, 2967–2978.

Blancas-Mejía, L.M., Tischer, A., Thompson, J.R., Tai, J., Wang, L., Auton, M., and Ramirez-Alvarado, M. (2014). Kinetic control in protein folding for light chain amyloidosis and the differential effects of somatic mutations. J. Mol. Biol. *426*, 347–361.

Blancas-Mejía, L.M., Horn, T.J., Marin-Argany, M., Auton, M., Tischer, A., and Ramirez-Alvarado, M. (2015). Thermodynamic and fibril formation studies of full length immunoglobulin light chain AL-09 and its germline protein using scan rate dependent thermal unfolding. Biophys. Chem. *207*, 13–20.

Bonomi, M., and Camilloni, C. (2017). Integrative structural and dynamical biology with PLUMED-ISDB. Bioinformatics *33*, 3999–4000.

Bonomi, M., Camilloni, C., and Vendruscolo, M. (2016a). Metadynamic metainference: Enhanced sampling of the metainference ensemble using metadynamics. Sci. Rep. *6*.

Bonomi, M., Camilloni, C., Cavalli, A., and Vendruscolo, M. (2016b). Metainference: A Bayesian inference method for heterogeneous systems. Sci. Adv. *2*.

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

Boxer, D.H., Zhang, H., Gourley, D.G., Hunter, W.N., Kelly, S.M., and Price, N.C. (2004). Sensing of remote oxyanion binding at the DNA binding domain of the molybdate-dependent transcriptional regulator, ModE. Org. Biomol. Chem. *2*, 2829–2837.

Branduardi, D., Bussi, G., and Parrinello, M. (2012). Metadynamics with adaptive gaussians. J. Chem. Theory Comput. *8*, 2247–2254.

Bryngelson, J.D., Onuchic, J.N., Socci, N.D., and Wolynes, P.G. (1995). Funnels, pathways, and the energy landscape of protein folding: A synthesis. Proteins Struct. Funct. Bioinforma. *21*, 167–195.

Buchner, J., Renner, M., Lilie, H., Hinz, H.J., Jaenicke, R., Kiefhaber, T., and Rudolph, R. (1991). Alternatively folded states of an immunoglobulin. Biochemistry *30*, 6922–6929.

Buckel, P., Hübner-Parajsz, C., Mattes, R., Lenz, H., Haug, H., and Beaucamp, K. (1987). Cloning and nucleotide sequence of heavy- and light-chain cDNAs from a creatine-kinase-specific monoclonal antibody. Gene *51*, 13–19.

Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. J. Chem. Phys. *126*.

Buxbaum, J. (1986). Aberrant immunoglobulin synthesis in light chain amyloidosis. Free light chain and light chain fragment production by human bone marrow cells in short-term tissue culture. J. Clin. Invest. *78*, 798–806.

Buxbaum, J.N., Chuba, J. V., Hellman, G.C., Solomon, A., and Gallo, G.R. (1990). Monoclonal immunoglobulin deposition disease: Light chain and light and heavy chain deposition diseases and their relation to light chain amyloidosis. Clinical features, immunopathology, and molecular analysis. Ann. Intern. Med. *112*, 455–464.

Camilloni, C., and Vendruscolo, M. (2015). A Tensor-Free Method for the Structural and Dynamical Refinement of Proteins using Residual Dipolar Couplings. J. Phys. Chem. B *119*, 653–661.

Camilloni, C., Sala, B.M., Sormanni, P., Porcari, R., Corazza, A., De Rosa, M., Zanini, S., Barbiroli, A., Esposito, G., Bolognesi, M., et al. (2016). Rational design of mutations that change the aggregation rate of a protein while maintaining its native structure and stability. Sci. Rep. *6*.

Campos, L.A., Cuesta-López, S., López-Llano, J., Falo, F., and Sancho, J. (2005). A double-deletion method to quantifying incremental binding energies in proteins from experiment: Example of a destabilizing hydrogen bonding pair. Biophys. J. *88*, 1311–1321.

Chen, K., and Tjandra, N. (2012). The use of residual dipolar coupling in studying proteins by NMR. Top. Curr. Chem. *326*, 47–67.

Chen, S., Ferrone, F.A., and Wetzel, R. (2002). Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation. Proc. Natl. Acad. Sci. *99*, 11884–11889.

Chi, E.Y., Krishnan, S., Randolph, T.W., and Carpenter, J.F. (2003). Physical stability of proteins in aqueous solution: Mechanism and driving forces in nonnative protein aggregation. Pharm. Res. *20*, 1325–1336.

Chiti, F. (2000). Mutational analysis of the propensity for amyloid formation by a globular protein. EMBO J. *19*, 1441–1449.

Chiti, F., and Dobson, C.M. (2009). Amyloid formation by globular proteins under native conditions. Nat. Chem. Biol. *5*, 15–22.

Chiti, F., and Dobson, C.M. (2017). Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. Annu. Rev. Biochem. *86*, 27–68.

Chiti, F., Webster, P., Taddei, N., Clark, A., Stefani, M., Ramponi, G., and Dobson, C.M. (1999). Designing conditions for in vitro formation of amyloid protofilaments and fibrils. Proc. Natl. Acad. Sci. *96*, 3590–3594.

Chothia, C., Gelfand, I., and Kister, A. (1998). Structural determinants in the sequences of immunoglobulin variable domain. J. Mol. Biol. *278*, 457–479.

Collins, S.R., Douglass, A., Vale, R.D., and Weissman, J.S. (2004). Mechanism of prion propagation: Amyloid growth occurs by monomer addition. PLoS Biol. *2*.

Connors, L.H., Jiang, Y., Budnik, M., Théberge, R., Prokaeva, T., Bodi, K.L., Seldin, D.C., Costello, C.E., and Skinner, M. (2007). Heterogeneity in primary structure, post-translational modifications, and germline gene usage of nine full-length amyloidogenic κ1 immunoglobulin light chains. Biochemistry *46*, 14259–14271.

Dasari, S., Theis, J.D., Vrana, J.A., Meureta, O.M., Quint, P.S., Muppa, P., Zenka, R.M., Tschumper, R.C., Jelinek, D.F., Davila, J.I., et al. (2015). Proteomic detection of immunoglobulin light chain variable region peptides from amyloidosis patient biopsies. J. Proteome Res. *14*, 1957–1967.

Dayhoff, M. (1978). Atlas of Protein Sequence and Structure (National Biomedical Research Foundation, Silver Spring, MD).

Demuro, A., Mina, E., Kayed, R., Milton, S.C., Parker, I., and Glabe, C.G. (2005). Calcium dysregulation and membrane disruption as a ubiquitous neurotoxic mechanism of soluble amyloid oligomers. J. Biol. Chem. *280*, 17294–17300.

Dill, K.A. (1990). Dominant Forces in Protein Folding. Biochemistry.

Dill, K.A., and Chan, H.S. (1997a). From levinthal to pathways to funnels. Nat. Struct. Biol. *4*, 10–19.

Dill, K.A., and Chan, H.S. (1997b). From levinthal to pathways to funnels. Nat. Struct. Biol.

Dill, K.A., and MacCallum, J.L. (2012). The protein-folding problem, 50 years on. Science (80-.).

Do, T.N., Carloni, P., Varani, G., and Bussi, G. (2013). RNA/peptide binding driven by electrostatics - Insight from bidirectional pulling simulations. J. Chem. Theory Comput. *9*, 1720–1730.

Dobson, C.M. (2003). Protein folding and misfolding. Nature *426*, 884–890.

Drozdetskiy, A., Cole, C., Procter, J., and Barton, G.J. (2015). JPred4: A protein secondary structure prediction server. Nucleic Acids Res. *43*, W389–W394.

Dueholm, M.S., Petersen, S. V., Sønderkær, M., Larsen, P., Christiansen, G., Hein, K.L., Enghild, J.J., Nielsen, J.L., Nielsen, K.L., Nielsen, P.H., et al. (2010). Functional amyloid in pseudomonas. Mol. Microbiol. *77*, 1009–1020.

Ellgaard, L., and Helenius, a. (2003). Quality control in the endoplasmic reticulum. Nat. Rev. Mol. Cell Biol. *4*, 181–191.

Engen, J.R., Wales, T.E., and Shi, X. (2011). Hydrogen Exchange Mass Spectrometry for Conformational Analysis of Proteins. In Encyclopedia of Analytical Chemistry, R.A. Meyers, ed. (John Wiley & Sons), p.

Englander, S.W., and Mayne, L. (2014). The nature of protein folding pathways. Proc. Natl. Acad. Sci. *111*, 15873–15880.

Feige, M.J., Walter, S., and Buchner, J. (2004). Folding mechanism of the C H2 antibody domain. J. Mol. Biol. *344*, 107–118.

Feige, M.J., Hagn, F., Esser, J., Kessler, H., and Buchner, J. (2007). Influence of the internal disulfide bridge on the folding pathway of the CL antibody domain. J. Mol. Biol. *365*, 1232–1244.

Feige, M.J., Groscurth, S., Marcinowski, M., Yew, Z.T., Truffault, V., Paci, E., Kessler, H., and Buchner, J. (2008). The structure of a folding intermediate provides insight into differences in immunoglobulin amyloidogenicity. Proc. Natl. Acad. Sci. *105*, 13373–13378.

Feige, M.J., Nath, S., Catharino, S.R., Weinfurtner, D., Steinbacher, S., and Buchner, J. (2009a). Structure of the Murine Unglycosylated IgG1 Fc Fragment. J. Mol. Biol. *391*, 599–608.

Feige, M.J., Groscurth, S., Marcinowski, M., Shimizu, Y., Kessler, H., Hendershot, L.M., and Buchner, J. (2009b). An Unfolded CH1 Domain Controls the Assembly and Secretion of IgG Antibodies. Mol. Cell *34*, 569–579.

Feige, M.J., Hendershot, L.M., and Buchner, J. (2010). How antibodies fold. Trends Biochem. Sci. *35*, 189–198.

Ferreira, S.T., Vieira, M.N.N., and De Felice, F.G. (2007). Soluble protein oligomers as emerging toxins in Alzheimer's and other amyloid diseases. In IUBMB Life, pp. 332–345.

Fersht, A.R. (1995). Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. Proc. Natl. Acad. Sci. *92*, 10869–10873.

Fersht, A.R. (1997). Nucleation mechanisms in protein folding. Curr. Opin. Struct. Biol. *7*, 3–9.

Fink, A.L. (1998). Protein aggregation: Folding aggregates, inclusion bodies and amyloid. Fold. Des. *3*.

Fitzpatrick, A.W.P., Debelouchina, G.T., Bayro, M.J., Clare, D.K., Caporini, M.A., Bajaj, V.S., Jaroniec, C.P., Wang, L., Ladizhansky, V., Muller, S.A., et al. (2013). Atomic structure and hierarchical assembly of a cross- amyloid fibril. Proc. Natl. Acad. Sci. *110*, 5468–5473.

Fitzpatrick, A.W.P., Falcon, B., He, S., Murzin, A.G., Murshudov, G., Garringer, H.J., Crowther, R.A., Ghetti, B., Goedert, M., and Scheres, S.H.W. (2017). Cryo-EM structures of tau filaments from Alzheimer's disease. Nature *547*, 185–190.

Folta-Stogniew, E. (2005a). Oligomeric States of Proteins Determined by Size-Exclusion Chromatography Coupled With Light Scattering, Absorbance, and Refractive Index Detectors. New Emerg. Proteomic Tech. 97–112.

Folta-Stogniew, E. (2005b). Oligomeric States of Proteins Determined by Size-Exclusion Chromatography Coupled With Light Scattering, Absorbance, and Refractive Index Detectors. New Emerg. Proteomic Tech. 97–112.

Fowler, D.M., Koulov, A. V., Balch, W.E., and Kelly, J.W. (2007). Functional amyloid - from bacteria to humans. Trends Biochem. Sci. *32*, 217–224.

French, D., Laskov, R., and Scharff, M. (1989). The role of somatic hypermutation in the generation of antibody diversity. Science (80-.). *244*, 1152–1157.

Gade Malmos, K., Blancas-Mejia, L.M., Weber, B., Buchner, J., Ramirez-Alvarado, M., Naiki, H., and Otzen, D. (2017). ThT 101: a primer on the use of thioflavin T to investigate amyloid formation. Amyloid *24*, 1–16.

Gehlsen, K.R., Gong, R., Bramhill, D., Wiersma, D.A., Kirkpatrick, S.A., Wang, Y., Feng, Y., and Dimitrov, D.S. (2012). Pharmacokinetics of engineered human monomeric and dimeric CH2 domains. MAbs *4*, 466–474.

Gertz, M.A. (2018). Immunoglobulin light chain amyloidosis diagnosis and treatment algorithm 2018. Blood Cancer J. *8*.

Gething, M.J., and Sambrook, J. (1992). Protein folding in the cell. Nature *355*, 33–45.

Ghisaidoobe, A., and Chung, S. (2014). Intrinsic Tryptophan Fluorescence in the Detection and Analysis of Proteins: A Focus on Förster Resonance Energy Transfer Techniques. Int. J. Mol. Sci. *15*, 22518–22538.

Gillmore, J.D., Lovat, L.B., Persey, M.R., Pepys, M.B., and Hawkins, P.N. (2001). Amyloid load and clinical outcome in AA amyloidosis in relation to circulating concentration of serum amyloid A protein. Lancet *358*, 24–29.

Glenner, G.G., Harbaugh, J., Ohms, J.I., Harada, M., and Cuatrecasas, P. (1970). An amyloid protein: The amino-terminal variable fragment of an immunoglobulin light chain. Biochem. Biophys. Res. Commun. *41*, 1287–1289.

Glenner, G.G., Ein, D., Eanes, E.D., Bladen, H.A., Terry, W., and Page, D.L. (1971). Creation of "amyloid" fibrils from Bence Jones proteins in vitro. Science *174*, 712–714.

Glickman, M.H., and Ciechanover, A. (2002). The Ubiquitin-Proteasome Proteolytic Pathway: Destruction for the Sake of Construction. Physiol. Rev. *82*, 373–428.

Glockshuber, R., and Schmidt, T. (1992). The Disulfide Bonds in Antibody Variable Domains : Effects on Stability, Folding in Vitro, and Functional Expression in Escherichia Coli. Biochemistry *31*, 1270–1279.

Gong, R., Vu, B.K., Feng, Y., Prieto, D.A., Dyba, M.A., Walsh, J.D., Prabakaran, P., Veenstra, T.D., Tarasov, S.G., Ishima, R., et al. (2009). Engineered Human Antibody Constant Domains with Increased Stability. J. Biol. Chem. *284*, 14203–14210.

Gong, R., Wang, Y., Ying, T., Feng, Y., Streaker, E., Prabakaran, P., and Dimitrov, D.S. (2013a). N-terminal truncation of an isolated human IgG1 CH2 domain significantly increases its stability and aggregation resistance. Mol. Pharm. *10*, 2642–2652.

Gong, R., Wang, Y., Ying, T., Feng, Y., Streaker, E., Prabakaran, P., and Dimitrov, D.S. (2013b). N-Terminal Truncation of an Isolated Human IgG1 CH2 Domain Significantly Increases Its Stability and Aggregation Resistance. Mol. Pharm. *10*, 2642–2652.

Goto, Y., and Hamaguchi, K. (1979). The role of the intrachain disulfide bond in the conformation and stability of the constant fragment of the immunoglobulin light chain. J. Biochem. *86*, 1433–1441.

Goto, Y., and Hamaguchi, K. (1981). Formation of the intrachain disulfide bond in the constant fragment of the immunoglobulin light chain. J. Mol. Biol. *146*, 321–340.

Goto, Y., and Hamaguchi, K. (1982). Unfolding and refolding of the constant fragment of the immunoglobulin light chain. J. Mol. Biol. *156*, 891–910.

Goto, Y., and Nishikiori, S. (1991). Role of electrostatic repulsion in the acidic molten globule of cytochrome c. J. Mol. Biol. *222*, 679–686.

Götz, A.W., Williamson, M.J., Xu, D., Poole, D., Le Grand, S., and Walker, R.C. (2012). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. generalized born. J. Chem. Theory Comput. *8*, 1542–1555.

Greenfield, N.J. (2006). Using circular dichroism spectra to estimate protein secondary structure. Nat. Protoc. *1*, 2876–2890.

Gremer, L., Schölzel, D., Schenk, C., Reinartz, E., Labahn, J., Ravelli, R.B.G., Tusche, M., Lopez-

Iglesias, C., Hoyer, W., Heise, H., et al. (2017). Fibril structure of amyloid-β(1–42) by cryo–electron microscopy. Science (80-.). *358*, 116–119.

Grogan, M., DIspenzieri, A., and Gertz, M.A. (2017). Light-chain cardiac amyloidosis: Strategies to promote early diagnosis and cardiac response. Heart *103*, 1065–1072.

Grzesiek, S., and Bax, A. (1992). Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. J. Magn. Reson. *96*, 432–440.

Grzesiek, S., and Bax, A. (1993). Amino acid type determination in the sequential assignment procedure of uniformly13C/15N-enriched proteins. J. Biomol. NMR *3*, 185–204.

Hartl, F.U., and Hayer-Hartl, M. (2002). Protein folding. Molecular chaperones in the cytosol: From nascent chain to folded protein. Science (80-.). *295*, 1852–1858.

Helms, L.R., and Wetzel, R. (1995). Destabilizing loop swaps in the CDRs of an immunoglobulin VL domain. Protein Sci. *4*, 2073–2081.

Helms, L.R., and Wetzel, R. (1996). Specificity of abnormal assembly in immunoglobulin light chain deposition disease and amyloidosis. J. Mol. Biol. *257*, 77–86.

Hora, M., Sarkar, R., Morris, V., Xue, K., Prade, E., Harding, E., Buchner, J., and Reif, B. (2017). MAK33 antibody light chain amyloid fibrils are similar to oligomeric precursors. PLoS One *12*.

Hurle, M.R., Helms, L.R., Li, L., Chan, W., and Wetzel, R. (1994). A role for destabilizing amino acid replacements in light-chain amyloidosis. Proc. Natl. Acad. Sci. U. S. A. *91*, 5446–5450.

Joh, N.H.J., Min, A., Faham, S., Whitelegge, J.P., Yang, D., Woods, V.L., and Bowie, J.U. (2008). Modest stabilization by most hydrogen-bonded side-chain interactions in membrane proteins. Nature *453*, 1266–1270.

John, C. (2017). Analysis of different antibody domain mutations affecting the VH /VL interface, CH 1 folding and CH 2 stability. Technische Universität München.

Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. *79*, 926–935.

Jucker, M., and Walker, L.C. (2011). Pathogenic protein seeding in Alzheimer disease and other neurodegenerative disorders. Ann. Neurol. *70*, 532–540.

Jung, S., and Pluckthun, A. (1997). Improving in vivo folding and stability of a single-chain Fv antibody fragment by loop grafting. Protein Eng. Des. Sel. *10*, 959–966.

Kabat, E.A., Wu, T.T., Perry, H.M., Gottesman, K.S., and Koeler, C. (1984). Sequences of proteins of immunological interest. Anal. Biochem. *138*, 265.

Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers.

Karplus, M., and Weaver, D.L. (1994). Protein folding dynamics: The diffusion-collision model and experimental data. Protein Sci. *3*, 650–668.

Kay, L.E., Xu, G.Y., and Yamazaki, T. (1994). Enhanced-Sensitivity Triple-Resonance Spectroscopy with Minimal H2O Saturation. J. Magn. Reson. Ser. A *109*, 129–133.

Kayed, R., Head, E., Thompson, J.L., McIntire, T.M., Milton, S.C., Cotman, C.W., and Glabe, C.G. (2003). Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. Science *300*, 486–489.

Kehry, M., Sibley, C., Fuhrman, J., Schilling, J., and Hood, L.E. (1979). Amino acid sequence of a mouse immunoglobulin μ chain. Proc. Natl. Acad. Sci. U. S. A. *76*, 2932–2936.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. *10*, 845–858.

Kelly, J.W. (1998). The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. Curr. Opin. Struct. Biol. *8*, 101–106.

Kelly, S.M., Jess, T.J., and Price, N.C. (2005). How to study proteins by circular dichroism. Biochim. Biophys. Acta - Proteins Proteomics *1751*, 119–139.

Kindt, T.J., Goldsby, R.A., Osborne, B.A., and Kuby, J. (2003). Immunology.

Kleizen, B., and Braakman, I. (2004). Protein folding and quality control in the endoplasmic reticulum. Curr. Opin. Cell Biol. *16*, 343–349.

Klimtchuk, E.S., Gursky, O., Patel, R.S., Laporte, K.L., Connors, L.H., Skinner, M., and Seldin, D.C. (2010). The Critical Role of the Constant Region in Thermal Stability and Aggregation of Amyloidogenic Immunoglobulin Light Chain. Biochemistry *49*, 9848–9857.

Knowles, T.P.J., Waudby, C.A., Devlin, G.L., Cohen, S.I.A., Aguzzi, A., Vendruscolo, M., Terentjev, E.M., Welland, M.E., and Dobson, C.M. (2009). An analytical solution to the kinetics of breakable filament assembly. Science (80-.). *326*, 1533–1537.

Knowles, T.P.J., Vendruscolo, M., and Dobson, C.M. (2014). The amyloid state and its association with protein misfolding diseases. Nat. Rev. Mol. Cell Biol. *15*, 384–396.

Krebs, M.R.H., Morozova-Roche, L.A., Daniel, K., Robinson, C. V., and Dobson, C.M. (2004). Observation of sequence specificity in the seeding of protein amyloid fibrils. Protein Sci. *13*, 1933–1938.

Krivov, G.G., Shapovalov, M. V., and Dunbrack, R.L. (2009). Improved prediction of protein side-chain conformations with SCWRL4. Proteins Struct. Funct. Bioinforma. *77*, 778–795.

Kumar, S., and Udgaonkar, J.B. (2009). Structurally distinct amyloid protofibrils form on separate pathways of aggregation of a small protein. Biochemistry *48*, 6441–6449.

Kumar, S., Rosenberg, J.M., Bouzida, D., Swendsen, R.H., and Kollman, P.A. (1992). The weighted histogram analysis method for free energy calculations on biomolecules. I. The method. J. Comput. Chem. *13*, 1011–1021.

Kurchan, E., Roder, H., and Bowler, B.E. (2005). Kinetics of loop formation and breakage in the denatured state of iso-1-cytochrome c. J. Mol. Biol. *353*, 730–743.

Kuwajima, K. (1989). The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. Proteins Struct. Funct. Bioinforma. *6*, 87–103.

Kyle, R.A., Linos, A., Beard, C.M., Linke, R.P., Gertz, M.A., O'Fallon, W.M., and Kurland, L.T. (1992). Incidence and natural history of primary systemic amyloidosis in Olmsted County, Minnesota, 1950 through 1989. Blood *79*, 1817–1822.

Laio, A., and Parrinello, M. (2002). Escaping free-energy minima. Proc. Natl. Acad. Sci. *99*, 12562–12566.

Last, N.B., and Miranker, A.D. (2013). Common mechanism unites membrane poration by amyloid and antimicrobial peptides. Proc. Natl. Acad. Sci. *110*, 6382–6387.

Lattman, E.E., and Rose, G.D. (1993). Protein folding--what's the question? Proc. Natl. Acad. Sci. *90*, 439–441.

Lavatelli, F., Perlman, D.H., Spencer, B., Prokaeva, T., McComb, M.E., Théberge, R., Connors, L.H., Bellotti, V., Seldin, D.C., Merlini, G., et al. (2008). Amyloidogenic and Associated Proteins in Systemic Amyloidosis Proteome of Adipose Tissue. Mol. Cell. Proteomics *7*, 1570–1583.

Lee, J., Culyba, E.K., Powers, E.T., and Kelly, J.W. (2011). Amyloid-β forms fibrils by nucleated conformational conversion of oligomers. Nat. Chem. Biol. *7*, 602–609.

Lesk, A.M., and Chothia, C. (1982). Evolution of proteins formed by β-sheets. II. The core of the immunoglobulin domains. J. Mol. Biol. *160*, 325–342.

Levinthal, C. (1968). Are there pathways for protein folding? J. Chim. Phys. *65*, 44–45.

Levinthal, C. (1969). How to fold graciously. Mössbauer Spectrosc. Biol. Syst. Proc. *24*, 22–24.

Li, M.Z., and Elledge, S.J. (2012). SLIC: a method for sequence- and ligation-independent cloning. Methods Mol. Biol. *852*, 51–59.

Li, D., Gong, R., Zheng, J., Chen, X., Dimitrov, D.S., and Zhao, Q. (2017). Engineered antibody CH2 domains binding to nucleolin: Isolation, characterization and improvement of aggregation. Biochem. Biophys. Res. Commun. *485*, 446–453.

Lilie, H., and Buchner, J. (1995). Domain interactions stabilize the alternatively folded state of an antibody Fab fragment. FEBS Lett. *362*, 43–46.

Linke, R.P., Tischendorf, F.W., Zucker-Franklin, D., and Franklin, E.C. (1973). The Formation of Amyloid-Like Fibrils in Vitro from Bence Jones Proteins of the VλI Subclass. J Immunol *111*, 24–26.

Liu, H., and May, K. (2012). Disulfide bond structures of IgG molecules. MAbs *4*, 17–23.

Löhr, T., Jussupow, A., and Camilloni, C. (2017). Metadynamic metainference: Convergence towards force field independent structural ensembles of a disordered peptide. J. Chem. Phys. *146*.

Lu, J., Yu, Y., Zhu, I., Cheng, Y., and Sun, P.D. (2014). Structural mechanism of serum amyloid A-mediated inflammatory amyloidosis. Proc. Natl. Acad. Sci. *111*, 5189–5194.

Lundmark, K., Westermark, G.T., Olsen, A., and Westermark, P. (2005). Protein fibrils in nature can enhance amyloid protein A amyloidosis in mice: Cross-seeding as a disease mechanism. Proc. Natl. Acad. Sci. *102*, 6098–6102.

Luo, Y., Ma, B., Nussinov, R., and Wei, G. (2014). Structural insight into tau protein's paradox of intrinsically disordered behavior, self-acetylation activity, and aggregation. J. Phys. Chem. Lett. *5*, 3026–3031.

M H Tao, S M Canfield, S.L.M. (1991). The differential ability of human IgG1 and IgG4 to activate complement is determined by the COOH-terminal sequence of the CH2 domain. J.

Exp. Med. *173*, 1025–1028.

Makhatadze, G.I., and Privalov, P.L. (1992). Protein interactions with urea and guanidinium chloride. A calorimetric study. J. Mol. Biol. *226*, 491–505.

Makhatadze, G.I., and Privalov, P.L. (1993a). Contribution of Hydration to Protein Folding Thermodynamics. I. The Enthalpy of Hydration. J. Mol. Biol.

Makhatadze, G.I., and Privalov, P.L. (1993b). Contribution of Hydration to Protein Folding Thermodynamics. I. The Enthalpy of Hydration. J. Mol. Biol. *232*, 639–659.

Marin-Argany, M., Güell-Bosch, J., Blancas-Mejía, L.M., Villegas, S., and Ramirez-Alvarado, M. (2015). Mutations can cause light chains to be too stable or too unstable to form amyloid fibrils. Protein Sci. *24*, 1829–1840.

McGuffin, L.J., Bryson, K., and Jones, D.T. (2000). The PSIPRED protein structure prediction server. Bioinformatics *16*, 404–405.

McNaught, K.S.P., Olanow, C.W., Halliwell, B., Isacson, O., and Jenner, P. (2001). Failure of the ubiquitin proteasome system in Parkinson's disease. Nat. Rev. Neurosci. *2*, 589–594.

Merlini, G., Seldin, D.C., and Gertz, M.A. (2011). Amyloidosis: Pathogenesis and new therapeutic options. J. Clin. Oncol. *29*, 1924–1933.

Mimura, Y., Church, S., Ghirlando, R., Ashton, P.R., Dong, S., Goodall, M., Lund, J., and Jefferis, R. (2001). The influence of glycosylation on the thermal stability and effector function expression of human IgG1-Fc: Properties of a series of truncated glycoforms. Mol. Immunol. *37*, 697–706.

Misra, P., Kodali, R., Chemuru, S., Kar, K., and Wetzel, R. (2016). Rapid α-oligomer formation mediated by the Aβ C terminus initiates an amyloid assembly pathway. Nat. Commun. *7*.

Morales, R., Moreno-Gonzalez, I., and Soto, C. (2013). Cross-Seeding of Misfolded Proteins: Implications for Etiology and Pathogenesis of Protein Misfolding Diseases. PLoS Pathog. *9*.

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

Morgan, A., Jones, N.D., Nesbitt, A.M., Chaplin, L., Bodmer, M.W., and Emtage, J.S. (1995). The N-terminal end of the CH2 domain of chimeric human IgG1 anti-HLA-DR is necessary for C1q, Fc gamma RI and Fc gamma RIII binding. Immunology *86*, 319–324.

Morgan, G.J., Usher, G.A., and Kelly, J.W. (2017). Incomplete Refolding of Antibody Light Chains to Non-Native, Protease-Sensitive Conformations Leads to Aggregation: A Mechanism of Amyloidogenesis in Patients? Biochemistry *56*, 6597–6614.

Morris, A.M., Watzky, M.A., and Finke, R.G. (2009). Protein aggregation kinetics, mechanism, and curve-fitting: a review of the literature. Biochim. Biophys. Acta *1794*, 375–397.

Muhandiram, D.R., and Kay, L.E. (1994a). Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity. J. Magn. Reson. Ser. B *103*, 203–216.

Muhandiram, D.R., and Kay, L.E. (1994b). Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity. J. Magn. Reson. Ser. B.

Munishkina, L. a, Cooper, E.M., Uversky, V.N., and Fink, A.L. (2004). The effect of macromolecular crowding on protein aggregation and amyloid fibril formation. J. Mol. Recognit.

Myers, J.K., Nick Pace, C., and Martin Scholtz, J. (1995). Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. Protein Sci. *4*, 2138–2148.

Neira, J.L., and Fersht, A.R. (1999). Exploring the folding funnel of a polypeptide chain by biophysical studies on protein fragments. J. Mol. Biol. *285*, 1309–1333.

Nokwe, C.N., Zacharias, M., Yagi, H., Hora, M., Reif, B., Goto, Y., and Buchner, J. (2014). A Residue-specific Shift in Stability and Amyloidogenicity of Antibody Variable Domains. J. Biol. Chem. *289*, 26829–26846.

Nokwe, C.N., Hora, M., Zacharias, M., Yagi, H., John, C., Reif, B., Goto, Y., and Buchner, J. (2015). The Antibody Light-Chain Linker Is Important for Domain Stability and Amyloid Formation. J. Mol. Biol. *427*, 3572–3586.

Nokwe, C.N., Hora, M., Zacharias, M., Yagi, H., Peschek, J., Reif, B., Goto, Y., and Buchner, J. (2016). A Stable Mutant Predisposes Antibody Domains to Amyloid Formation through Specific Non-Native Interactions. J. Mol. Biol. *428*, 1315–1332.

Oberti, L., Rognoni, P., Barbiroli, A., Lavatelli, F., Russo, R., Maritan, M., Palladini, G., Bolognesi, M., Merlini, G., and Ricagno, S. (2017). Concurrent structural and biophysical traits link with immunoglobulin light chains amyloid propensity. Sci. Rep. *7*, 16809.

Ohgushi, M., and Wada, A. (1983). "Molten-globule state": a compact form of globular proteins with mobile side-chains. FEBS Lett. *164*, 21–24.

Ono, K., Takahashi, R., Ikeda, T., and Yamada, M. (2012). Cross-seeding effects of amyloid β-protein and α-synuclein. J. Neurochem. *122*, 883–890.

Orengo, C.A., Todd, A.E., and Thornton, J.M. (1999). From protein structure to function. Curr. Opin. Struct. Biol.

Pace, C.N. (1986). [14] Determination and Analysis of Urea and Guanidine Hydrochloride Denaturation Curves. Methods Enzymol. *131*, 266–280.

Pace, C.N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995). How to measure and predict the molar absorption coefficient of a protein. Protein Sci. *4*, 2411–2423.

Pace, C.N., Fu, H., Fryar, K.L., Landua, J., Trevino, S.R., Shirley, B.A., Hendricks, M.M.N., Iimura, S., Gajiwala, K., Scholtz, J.M., et al. (2011). Contribution of hydrophobic interactions to protein stability. J. Mol. Biol. *408*, 514–528.

Parrinello, M., and Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. J. Appl. Phys. *52*, 7182–7190.

Pepys, M.B. (2006). Amyloidosis. Annu. Rev. Med. *57*, 223–241.

Pfaendtner, J., and Bonomi, M. (2015). Efficient Sampling of High-Dimensional Free-Energy Landscapes with Parallel Bias Metadynamics. J. Chem. Theory Comput. *11*, 5062–5067.

Phillips, S.R., Wilson, L.J., and Borkman, R.F. (1986). Acrylamide and iodide fluorescence quenching as a structural probe of tryptophan microenvironment in bovine lens crystallins.

Curr. Eye Res. *5*, 611–619.

Piehl, D.W., Blancas-Mejía, L.M., Wall, J.S., Kennel, S.J., Ramirez-Alvarado, M., and Rienstra, C.M. (2017). Immunoglobulin Light Chains Form an Extensive and Highly Ordered Fibril Involving the N- and C-Termini. ACS Omega *2*, 712–720.

Pinney, J.H., Smith, C.J., Taube, J.B., Lachmann, H.J., Venner, C.P., Gibbs, S.D.J., Dungu, J., Banyperasad, S.M., Wechalekar, A.D., Whelan, C.J., et al. (2013). Systemic Amyloidosis in England: An epidemiological study. Br. J. Haematol. *161*, 525–532.

Pirovano, W., and Heringa, J. (2010). Protein secondary structure prediction. Methods Mol. Biol. *609*, 327–348.

Plaxco, K.W., Simons, K.T., Ruczinski, I., and Baker, D. (2000). Topology, stability, sequence, and length: Defining the determinants of two-state protein folding kinetics. Biochemistry *39*, 11177–11183.

Politi, R., and Harries, D. (2010). Enthalpically driven peptide stabilization by protective osmolytes. Chem. Commun. *46*, 6449–6451.

Poljak, R.J., Amzel, L.M., Avey, H.P., Chen, B.L., Phizackerley, R.P., and Saul, F. (1973). Three-dimensional structure of the Fab' fragment of a human immunoglobulin at 2,8-A resolution. Proc. Natl. Acad. Sci. U. S. A. *70*, 3305–3310.

Pornillos, O., Ganser-Pornillos, B.K., Banumathi, S., Hua, Y., and Yeager, M. (2010). Disulfide Bond Stabilization of the Hexameric Capsomer of Human Immunodeficiency Virus. J. Mol. Biol. *401*, 985–995.

Privalov, P.L., and Makhatadze, G.I. (1993a). Contribution of Hydration to Protein Folding Thermodynamics. II. The Entropy and Gibbs Energy of Hydration. J. Mol. Biol.

Privalov, P.L., and Makhatadze, G.I. (1993b). Contribution of Hydration to Protein Folding Thermodynamics. II. The Entropy and Gibbs Energy of Hydration. J. Mol. Biol. *232*, 660–679.

Qin, Z., Hu, D., Zhu, M., and Fink, A.L. (2007). Structural characterization of the partially folded intermediates of an immunoglobulin light chain leading to amyloid fibrillation and amorphous aggregation. Biochemistry *46*, 3521–3531.

Rabbitts, T.H., Forster, A., and Milstein, C.P. (1981). Human immunoglobulin heavy chain genes: evolutionary comparisons of C mu, C delta and C gamma genes and associated switch sequences. Nucleic Acids Res. *9*, 4509–4524.

Ramirez-Alvarado, M., Merkel, J.S., and Regan, L. (2000). A systematic exploration of the influence of the protein stability on amyloid fibril formation in vitro. Proc. Natl. Acad. Sci. *97*, 8979–8984.

Ravetch, J. V., and Bolland, S. (2001). IgG Fc Receptors. Annu. Rev. Immunol. *19*, 275–290.

Richardson, J.S. (1981). The anatomy and taxonomy of protein structure. Adv. Protein Chem.

Rudolph, R., and Lilie, H. (1996). In vitro folding of inlusion body proteins. FASEB J.

Rus, H., Cudrici, C., and Niculescu, F. (2005). The role of the complement system in innate immunity. Immunol. Res. *33*, 103–112.

Saelices, L., Johnson, L.M., Liang, W.Y., Sawaya, M.R., Cascio, D., Ruchala, P., Whitelegge, J.,

Jiang, L., Riek, R., and Eisenberg, D.S. (2015). Uncovering the mechanism of aggregation of human transthyretin. J. Biol. Chem. *290*, 28932–28943.

Safar, J., Roller, P.P., Gajdusek, D.C., and Gibbs, C.J. (1993). Thermal stability and conformational transitions of scrapie amyloid (prion) protein correlate with infectivity. Protein Sci. *2*, 2206–2216.

Sakano, H., Rogers, J.H., Hüppi, K., Brack, C., Traunecker, A., Maki, R., Wall, R., and Tonegawa, S. (1979). Domains and the hinge region of an immunoglobulin heavy chain are encoded in separate DNA segments. Nature *277*, 627–633.

Sambrook, J., Fritsch, E., and Maniatis, T. (1989). Molecular cloning: a laboratory manual.

Sanchorawala, V. (2006). Light-chain (AL) amyloidosis: diagnosis and treatment. Clin. J. Am. Soc. Nephrol. *1*, 1331–1341.

Santoro, M.M., and Bolen, D.W. (1988). Unfolding Free Energy Changes Determined by the Linear Extrapolation Method. 1. Unfolding of Phenylmethanesulfonyl a-Chymotrypsin Using Different Denaturants. Biochemistry *27*, 8063–8068.

Schleucher, J., Sattler, M., and Griesinger, C. (1993). Coherence Selection by Gradients without Signal Attenuation: Application to the Three-Dimensional HNCO Experiment. Angew. Chemie Int. Ed. English *32*, 1489–1491.

Schmid, F. (2005). Spectroscopic Techniques to Study Protein Folding and Stability. In The Protein Folding Handbook, J. Buchner, and T. Kiefhaber, eds. (Wiley-VCH Verlag Gmbh & Co. KGaA Weinheim), pp. 22–44.

Schmidt, A., Annamalai, K., Schmidt, M., Grigorieff, N., and Fändrich, M. (2016). Cryo-EM reveals the steric zipper structure of a light chain-derived amyloid fibril. Proc. Natl. Acad. Sci. *113*, 6200–6205.

Schroeder, H.W., and Cavacini, L. (2010). Structure and function of immunoglobulins. J. Allergy Clin. Immunol. *125*, S41-52.

Sharkey, R.M., and Goldenberg, D.M. (2005). Perspectives on cancer therapy with radiolabeled monoclonal antibodies. J. Nucl. Med. *46 Suppl 1*, 115S–27S.

Shimaoka, M., Lu, C., Salas, A., Xiao, T., Takagi, J., and Springer, T.A. (2002). Stabilizing the integrin M inserted domain in alternative conformations with a range of engineered disulfide bonds. Proc. Natl. Acad. Sci. *99*, 16737–16741.

Shimizu, A., Takahashi, N., Yamawaki-Kataoka, Y., Nishida, Y., Kataoka, T., and Honjo, T. (1981). Ordering of mouse immunoglobulin heavy chain genes by molecular cloning. Nature *289*.

Shoichet, B.K., Baase, W.A., Kuroki, R., and Matthews, B.W. (1995). A relationship between protein stability and protein function. Proc. Natl. Acad. Sci. *92*, 452–456.

Simpson, E.R., Herold, E.M., and Buchner, J. (2009). The Folding Pathway of the Antibody VL Domain. J. Mol. Biol. *392*, 1326–1338.

Soscia, S.J., Kirby, J.E., Washicosky, K.J., Tucker, S.M., Ingelsson, M., Hyman, B., Burton, M.A., Goldstein, L.E., Duong, S., Tanzi, R.E., et al. (2010). The Alzheimer's disease-associated amyloid β-protein is an antimicrobial peptide. PLoS One *5*.

Spera, S., and Bax, A. (1991). Empirical Correlation between Protein Backbone Conformation

and Cα and Cβ 13C Nuclear Magnetic Resonance Chemical Shifts. J. Am. Chem. Soc. *113*, 5490–5492.

Stevens, F.J., Myatt, E.A., Chang, C.-H., Westholm, F.A., Eulitz, M., Weiss, D.T., Murphy, C., Solomon, A., and Schiffer, M. (1995). A Molecular Model for Self-Assembly of Amyloid Fibrils: Immunoglobulin Light Chains. Biochemistry *34*, 10697–10702.

Stoscheck, C.M. (1990). Quantitation of protein. Methods Enzymol. *182*, 50–68.

Swindells, M.B., Porter, C.T., Couch, M., Hurst, J., Abhinandan, K.R., Nielsen, J.H., Macindoe, G., Hetherington, J., and Martin, A.C.R. (2017). abYsis: Integrated Antibody Sequence and Structure—Management, Analysis, and Prediction. J. Mol. Biol.

Takano, K., Scholtz, J.M., Sacchettini, J.C., and Pace, C.N. (2003). The contribution of polar group burial to protein stability is strongly context-dependent. J. Biol. Chem. *278*, 31790–31795.

Tallmadge, D.H., Huebner, J.S., and Borkman, R.F. (1989). Acrylamide quenching of tryptophan photochemistry and photophysics. Photochem. Photobiol. *49*, 381–386.

Thies, M.J.., and Pirkl, F. (2000). Chromatographic purification of the CH2 domain of the monoclonal antibody MAK33. J. Chromatogr. B Biomed. Sci. Appl. *737*, 63–69.

Tokuriki, N., Stricher, F., Serrano, L., and Tawfik, D.S. (2008). How protein stability and new functions trade off. PLoS Comput. Biol. *4*.

Tribello, G.A., Bonomi, M., Branduardi, D., Camilloni, C., and Bussi, G. (2014). PLUMED 2: New feathers for an old bird. Comput. Phys. Commun. *185*, 604–613.

Tych, K.M., Batchelor, M., Hoffmann, T., Wilson, M.C., Hughes, M.L., Paci, E., Brockwell, D.J., and Dougan, L. (2016). Differential Effects of Hydrophobic Core Packing Residues for Thermodynamic and Mechanical Stability of a Hyperthermophilic Protein. Langmuir *32*, 7392–7402.

Tyedmers, J., Mogk, A., and Bukau, B. (2010). Cellular strategies for controlling protein aggregation. Nat. Rev. Mol. Cell Biol. *11*, 777–788.

Uversky, V.N. (2003). A protein-chameleon: Conformational plasticity of α-synuclein, a disordered protein involved in neurodegenerative disorders. J. Biomol. Struct. Dyn. *21*, 211–234.

Uversky, V.N., and Fink, A.L. (2004). Conformational constraints for amyloid fibrillation: The importance of being unfolded. Biochim. Biophys. Acta - Proteins Proteomics *1698*, 131–153.

Vendruscolo, M., Knowles, T.P.J., and Dobson, C.M. (2011). Protein solubility and protein homeostasis: A generic view of protein misfolding disorders. Cold Spring Harb. Perspect. Biol. *3*.

Vivian, J.T., and Callis, P.R. (2001). Mechanisms of Tryptophan Fluorescence Shifts in Proteins. Biophys. J. *80*, 2093–2109.

Vogt, G., Woell, S., and Argos, P. (1997). Protein thermal stability, hydrogen bonds, and ion pairs. J. Mol. Biol. *269*, 631–643.

Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D. (2005). The CCPN data model for NMR spectroscopy:

Development of a software pipeline. Proteins Struct. Funct. Genet. *59*, 687–696.

Wang, J.H. (2013). The sequence signature of an Ig-fold. Protein Cell *4*, 569–572.

Wang, W. (2005). Protein aggregation and its inhibition in biopharmaceutics. Int. J. Pharm. *289*, 1–30.

Weber, B., Hora, M., Kazman, P., Göbl, C., Camilloni, C., Reif, B., and Buchner, J. (2018a). The Antibody Light-Chain Linker Regulates Domain Orientation and Amyloidogenicity. J. Mol. Biol.

Weber, B., Brandl, M.J., Pulido Cendales, M.D., Berner, C., Pradhan, T., Feind, G.M., Zacharias, M., Reif, B., and Buchner, J. (2018b). A single residue switch reveals principles of antibody domain integrity. J. Biol. Chem.

Wechalekar, A.D., Gillmore, J.D., and Hawkins, P.N. (2016). Systemic amyloidosis. Lancet *387*, 2641–2654.

de Weers, M., Tai, Y.-T., van der Veer, M.S., Bakker, J.M., Vink, T., Jacobs, D.C.H., Oomen, L.A., Peipp, M., Valerius, T., Slootstra, J.W., et al. (2011). Daratumumab, a Novel Therapeutic Human CD38 Monoclonal Antibody, Induces Killing of Multiple Myeloma and Other Hematological Tumors. J. Immunol. *186*, 1840–1848.

Weiner, L.M., Surana, R., and Wang, S. (2010). Monoclonal antibodies: Versatile platforms for cancer immunotherapy. Nat. Rev. Immunol. *10*, 317–327.

Weiser, J., Shenkin, P.S., and Still, W.C. (1999). Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). J. Comput. Chem. *20*, 217–230.

Wen, J., Arakawa, T., and Philo, J.S. (1996). Size-exclusion chromatography with on-line light-scattering, absorbance, and refractive index detectors for studying proteins and their interactions. Anal. Biochem. *240*, 155–166.

Wetlaufer, D.B., Malik, S.K., Stoller, L., and Coffin, R.L. (1964). Nonpolar Group Participation in the Denaturation of Proteins by Urea and Guanidinium Salts. Model Compound Studies. J. Am. Chem. Soc. *86*, 508–514.

Williams, A.F., and Barclay, A.N. (1988). The Immunoglobulin Superfamily—Domains for Cell Surface Recognition. Annu. Rev. Immunol. *6*, 381–405.

Williamson, M.P. (2013). Using chemical shift perturbation to characterise ligand binding. Prog. Nucl. Magn. Reson. Spectrosc. *73*, 1–16.

Wishart, D., and Sykes, B. (1994). The 13C Chemical-Shift Index: A simple method for the identification of protein secondary structure using 13C chemical-shift data. J. Biomol. NMR *4*.

Wittekind, M., and Mueller, L. (1993). HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha- and Beta-Carbon Resonances in Proteins. J. Magn. Reson. Ser. B *101*, 201–205.

Wright, P.E., and Dyson, H.J. (1999). Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. J. Mol. Biol. *293*, 321–331.

Xia, Z., Yang, Z., Huynh, T., King, J.A., and Zhou, R. (2013). UV-radiation induced disruption of dry-cavities in human γd-crystallin results in decreased stability and faster unfolding. Sci. Rep. *3*.

Zhang, A., Hu, P., MacGregor, P., Xue, Y., Fan, H., Suchecki, P., Olszewski, L., and Liu, A. (2014). Understanding the conformational impact of chemical modifications on monoclonal antibodies with diverse sequence variation using hydrogen/deuterium exchange mass spectrometry and structural modeling. Anal. Chem. *86*, 3468–3475.

Zwanzig, R., Szabo, A., and Bagchi, B. (1992). Levinthal's paradox. Proc. Natl. Acad. Sci. *89*, 20–22.

9. Publications and presentations

Weber, B., Hora, M., Kazman, P., Göbl, C., Camilloni, C., Reif, B., & Buchner, J. (2018). The Antibody Light Chain Linker Regulates Domain Orientation and Amyloidogenicity. Journal of molecular biology, 430(24), 4925-4940.

Weber, B., Brandl, M. J., Cendales, M. D. P., Berner, C., Pradhan, T., Feind, G. M., Zacharias, M., Reif, B., Buchner, J. (2018). A single residue switch reveals principles of antibody domain integrity. Journal of Biological Chemistry, 293(44), 17107-17118.

Pasalic, D., Weber, B., Giannone, C., Anelli, T., Müller, R., Fagioli, C., Felkl, M., John, C., Mossuto, M.F., Becker, C.F.W., Sitia, R., Buchner, J. (2017). A peptide extension dictates IgM assembly. Proceedings of the National Academy of Sciences, 114(41), E8575-E8584.

Herold, E.M., John, C., Weber, B., Kremser, S., Eras, J., Berner, C., Deubler, S., Zacharias, M., Buchner, J. (2017). Determinants of the assembly and function of antibody variable domains. Scientific Reports, 7.

Gade Malmos, K., Blancas-Mejia, L.M., Weber, B., Buchner, J., Ramirez-Alvarado, M., Naiki, H., Otzen, D. (2017). ThT 101: a primer on the use of thioflavin T to investigate amyloid formation. Amyloid, 24(1), 1–16. (Review article)

Hora, M., Carballo Pacheco, M., Weber, B., Buchner, J., Strodel, B., & Reif, B. (2017). Solid- and solution-state nuclear magnetic resonance spectroscopic studies on antibody light chain amyloid formation and interactions with epigallocatechin gallate. Amyloid, 24, 10.

Hora, M., Carballo-Pacheco, M., Weber, B., Morris, V. K., Wittkopf, A., Buchner, J., Strodel, B., Reif, B. (2017). Epigallocatechin-3-gallate preferentially induces aggregation of amyloidogenic immunoglobulin light chains. Scientific Reports, 7.

Oral presentations

IgM – How to assemble a hexameric antibody, *3rd Sino-German Symposium on Protein Folding, Redox Regulation and Quality Control, Beijing, China, October 18th, 2017*

IgM – How to assemble a hexameric antibody, *13th ER & Redox Club Meeting, Homburg, Germany, April 27th, 2017*

Domain interactions of immunoglobulin light chains, *27th Faltertage, Halle (Saale), Germany, October 23rd, 2016*

10. Acknowledgements

Zuerst möchte ich meinem Doktorvater Johannes Buchner danken, für die Unterstützung und Freiheit bei allen Fragen und vor allem dafür, dass er das Experiment mit einem forschungsunerfahrenen Apotheker eingegangen ist.

Dem ganzen Lehrstuhl Biotechnologie, den aktuellen und ehemaligen Kollegen, möchte ich herzlich danken für die schönen, lehrreichen, anstrengenden und lustigen Zeiten und Erfahrungen. Ein besonderer Dank geht an Christina Nickels, die mich mit viel Geduld in die Welt der Proteinbiochemie eingewiesen hat. Dankeschön an Pamina Kazman für die Durchführung der TEM Messungen. Vielen Dank an Frau Rubinstein, die sich immer mit Herz, Seele und Tatkraft um die Anliegen aller kümmert. Ich möchte mich sehr bei Bettina Richter, Anja Osterauer, Gina Feind, Florian Rührnößl und Ruby Khan für die tolle Unterstützung im Labor danken.

Großer Dank an alle Kooperationspartner innerhalb und außerhalb der TUM: Manuel Hora, Matthias Brandl und Tejaswini Pradhan, sowie Prof. Bernd Reif für die NMR Experimente, den guten Austausch und die gemeinsame Planung einiger Projekte. Danke auch an Christoph Göbl und Prof. Carlo Camilloni, die ebenfalls zum Gelingen dieser Projekte beitragen haben. Danke an Daniela Pulido Cendales und Prof. Martin Zacharias für die Durchführung der MD Simulationen. Mille grazie an Chiara Giannone und Prof. Roberto Sitia für die gute Zusammenarbeit und den stetigen Austausch. Vielen Dank auch an Astrid König, Marie-Theres Vielberg und Prof. Michael Groll für die Möglichkeit und Durchführung der Proteinkristallisationen. Besten Dank an Prof. Matthias Feige, der mit offenem Ohr und gutem Rat zur Seite stand.

Vielen Dank an Heike und Thomas Zimmer, sowie Johannes und Lisa Zimmer für die so herzliche und stetige Unterstützung.

Danke Mama und Papa für tatsächlich alles.

Chrisi, danke auch Dir für alles.

11. Declaration

I, Benedikt Weber, hereby declare that this thesis was prepared by me independently and using only references and resources stated here. The work has so far not been submitted to any audit commission. Parts of this work will be published in scientific journals.

Hiermit erkläre ich, Benedikt Weber, dass die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Die Arbeit wurde bisher keiner Prüfungskommission vorgelegt. Teile dieser Arbeit werden in wissenschaftlichen Journalen veröffentlicht.

_____

Benedikt Weber