TECHNISCHE UNIVERSITÄT MÜNCHEN

Department Chemie
Lehrstuhl für Biotechnologie

Molecular determinants and mechanisms of antibody light chain (AL) amyloidosis

Cardine NOKWE NKUMBE

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: 　　　　　　　　Univ.- Prof. Dr. Aymelt Itzen

Prüfer der Dissertation:

　　　　　1　Univ.- Prof. Dr. Johannes Buchner

　　　　　2　Univ.- Prof. Dr. Bernd Reif

Die Dissertation wurde am 16.10.2014 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 17.11.2014 angenommen.

"One is defined by principles and principles guide you through success"

Cardine N. NOKWE

Dedicated

to all the

concerned NOKWES and well-wishers

TABLE OF CONTENTS

ABSTRACT

Understanding the mechanisms underlying protein misfolding and aggregation has gained a lot of scientific attention in the last decades with the discovery that these processes are at the basis of several life-threatening human diseases with a huge social burden. Common examples of these diseases are neurodegenerative disorders like Alzheimer`s, Huntington`s and Parkinson`s disease, which belong to the localized amyloidosis, or the less known but highly fatal systemic antibody light chain (AL) amyloidosis. All these diseases are characterized by the extracellular and/or intracellular deposition of insoluble β-sheet rich amyloid fibrils. The fibrils are derived from the aggregation of misfolded, normally soluble proteins. In the case of antibodies, stringent cellular quality control mechanisms exist that ensure that only completely assembled antibodies, i.e. two identical heavy chains (HC) paired with two identical light chains (LC) are secreted. But sometimes, things go awry as in neoplastic plasma cells, resulting in the excess secretion of free LCs. Misfolding and aggregation of these free LCs or its N-terminal variable (V_L) domain are the basis of primary systemic AL amyloidosis. Analysis of proteins isolated from AL patient biopsies showed that the main constituent of amyloid deposits is the V_L domain of the LC. It is so far unclear which residues play a key role in governing these processes. In this work we show that the conserved residue 2 of V_L domains is crucial for controlling its thermodynamic stability and fibril formation. Hydrophobic side chains at position 2 stabilize the domain whereas charged residues destabilize and lead to amyloid fibril formation. NMR experiments identified several segments within the core of the V_L domain to be affected by changes in residue 2. Furthermore, molecular dynamic (MD) simulations showed that hydrophobic side chains at position 2 remain buried in a hydrophobic pocket while charged side chains show a high flexibility. This results in a predicted difference in the dissociation free energy of ~ 10 kJ mol^{-1}, which is in excellent agreement with our experimental values. Interestingly, this switch point is found only in V_L domains of the κ family and not in V_Lλ or in V_H domains, despite a highly similar domain architecture.

Both stabilizing and destabilizing interactions within the V_L domain can influence the kinetics of fibrillation. The underlying mechanisms behind this are not understood. In this thesis, through the characterization of naturally occurring amyloidogenic mutants, we show on the one hand that the destabilization of the native fold by an R61A mutation, which leads to the disruption of a key conserved salt bridge in the V_L domain can lead to the rapid population of intermediate and unfolded states that give rise to amyloid fibrils. On the other hand, we show that the acquisition of the S20N or D70N mutation does not destabilize or influence the folding of the V_L domain, however it predisposes the domain to fibrillar aggregation. Although the thermodynamic stability and the overall folding mechanisms are not affected by these amyloid forming mutants (S20N and D70N), our structural studies by NMR spectroscopy and MD simulations revealed that there are localized structural differences compared to the wild type protein, which may facilitate the conversion of the unfolded V_L protein into amyloid fibrils. Asn residues are known to have a high tendency for the formation and stabilization of β-sheet-rich structures by establishing native-like interactions. Therefore, the H-bond network established by the mutations and the side-chain interactions of the Asn20 or Asn70 are

suggested to be possible mechanisms through which the mutations induce and stabilize the fibrillar state of the V_L protein.

As V_Ls with different C-terminal lengths have been reported to show varying aggregation kinetics, it seems that some residues at the linker connecting the V_L and the LC constant (C-terminal; (C_L)) domain (V_L/C_L boundary residues) might be important in this respect. It was earlier shown that the truncation of boundary residues (N-terminus) of $C_L\lambda$ results in a decrease stability and a higher unfolding rate for the domain. Within this work, through the generation of various truncation variants, the effects of specific $V_L\kappa$/$C_L\kappa$-linker residues on the $V_L\kappa$ and the $C_L\kappa$ domain were analyzed. $V_L\kappa$ or $C_L\kappa$ variants lacking the Arg108 linker-residue were slightly but distinctively less stable than corresponding wild type proteins. While the $V_L\kappa$ wild type (wt) was resistant to fibril formation, $V_L\kappa$-ΔR108 was prone to fibrillation. Folding experiments showed that the C-terminal ΔR108 $V_L\kappa$ variant unfolds faster than the $V_L\kappa$-wt and that both variants refold with the same kinetics. These findings indicate that the C-terminal Arg108 is implicated in the rate limiting step of the unfolding kinetics but not with the rate limiting step of the refolding kinetics of the V_Lk domain. This agrees well with the notion that unfolding events precede fibrillation, which would explain why the faster unfolding $V_L\kappa$-ΔR108 variant is more prone to fibril formation. In the same line, folding experiments revealed that the refolding kinetics were slower for the C_Lk variants that lacked the N-terminal Arg108.

Taken together, the data presented in this work reveals novel insights into the architecture of antibody domains in general and the prerequisites for the formation of amyloid fibrils in particular. This might also contribute to the rational design and engineering of stable and functional antibody fragments. Additionally, the two mechanisms underlying the V_L fibrillation process proposed in this thesis show that both destabilizing and stabilizing interactions can result in amyloid fibril formation, whereby in both cases an unfolding event precedes the process of fibrillation. This might open new avenues in the investigation of the processes that govern fibril formation of other proteins, particularly those that are prone to amyloid formation. Finally, a better understanding of all the mechanisms determining fibril formation and the process of amyloidogenesis would lead to better design of therapeutic strategies.

ZUSAMMENFASSUNG

Die der Fehlfaltung von Proteinen und der Entstehung von Aggregation zugrundeliegenden Mechanismen zu verstehen, hat in den letzten Jahrzehnten aufgrund der Entdeckung, dass diese Phänomene direkt mit der Entwicklung zahlreicher lebensbedrohlicher Krankheiten zusammenhängen, stark an Bedeutung gewonnen. Beispiele für solche Krankheiten sind neurodegenerative Krankheiten wie Alzheimer, die Huntington-Krankheit und Parkinson, welche den lokalisierten Amyloidosen zugeordnet sind, aber auch die weniger bekannte, jedoch fatale systemische Antikörper-leichte-Kette- (AL)–Amyloidose. All diese Krankheiten sind durch eine intra- oder extrazelluläre Ablagerung von β-Faltblatt-reichen, unlöslichen Amyloidfibrillen charakterisiert. Diese Fibrillen entstehen durch die Aggregation fehlgefalteter, normalerweise löslicher Proteine. Vereinzelt kann es in den Antikörper- produzierenden neoplastischen Zellen zu übermäßiger Sekretion von freien leichten Ketten (LCs) kommen. Fehlfaltung und Aggregation dieser freien LCs oder ihrer variablen (V_L) Domänen verursachen primäre systemische AL – Amyloidose. Die Analyse von aus AL – Patienten isolierten Proteinen ergab, dass die N-terminale V_L Domäne der LCs den Hauptbestandteil der Amyloidablagerungen darstellt. Bisher ist jedoch unklar, welche Aminosäurereste eine Schlüsselrolle in der Entstehung dieser Ablagerungen spielen. In der vorliegenden Arbeit konnte gezeigt werden, dass die Aminosäure 2 von V_L Domänen eine bedeutende Rolle für deren thermodynamische Stabilität und die Bildung von Fibrillen spielt. Während hydrophobe Reste in dieser Position die Domäne stabilisieren, verringern geladene Reste die Stabilität und führen zur Bildung von Amyloidfibrillen. Mittels NMR konnten Regionen im Kern der V_L Domäne identifiziert werden, deren Konformation durch Variation des Restes 2 beeinflusst wird. Desweiteren, ergaben molekular-dynamische (MD) Simulationen, dass hydrophobe Reste an Position 2 mit einer hydrophoben Tasche interagieren, während geladene Reste an dieser Position eine hohe konformationelle Flexibilität auslösen. Auf Grundlage dieser Simulationen konnte eine Änderung der Dissoziationsenergie von ~ 10 kJ mol^{-1} berechnet werden, welche mit den experimentellen Ergebnissen sehr gut übereinstimmt. Trotz sehr ähnlicher Architektur, existiert dieser molekulare Schalter an Position 2 jedoch ausschließlich in V_L Domänen der κ Familie, nicht aber in V_Lλ oder V_H Domänen. Sowohl stabilisierende als auch destabilisierende Interaktionen in der V_L Domäne beeinflussen die Kinetik der Fibrillenbildung. Die dieser Beobachtung zugrundeliegenden Mechanismen sind bisher jedoch nicht verstanden.

Durch die Charakterisierung natürlich vorkommender amyloidogener Mutanten, konnte in dieser Arbeit gezeigt werden, dass durch die R61A Mutation wird eine konservierte Salzbrücke in der V_L Domäne gebrochen. Diese führt zur Destabilisierung der nativen Konformation und zu starker Anreicherung von Faltungsintermediaten und ungefalteten Zuständen, welche die Grundlage für die Entstehung von Amyloidfibrillen darstellen. Desweiteren konnte gezeigt werden, dass die S20N oder eine D70N Mutationen die V_L Domäne zwar nicht destabilisieren oder in ihrer Faltung beeinflussen, jedoch führen beide Mutationen jeweils zu einer Prädisposition der V_L Domäne für Fibrillenbildung. Auch wenn die thermodynamische Stabilität und der Faltungsmechanimus der V_L Domäne nicht durch eine dieser beiden Mutationen beeinflusst wird, ergaben NMR Messungen und MD Simulationen lokale

strukturelle Unterschiede dieser Varianten im Vergleich zum Wildtypprotein, welche für die Überführung des ungefalteten V_L Proteins in die Amyloidfibrille verantwortlich sein könnten. Generell zeigen Asparaginreste eine hohe Tendenz β-Faltblatt-reiche Strukturen auszubilden. Das durch die Mutationen induzierte Wasserstoffbrücken-Netzwerk und die Interaktionen der Seitenketten von Asn20 und Asn70, könnten möglicherweise einen fibrillären Zustand des V_L Proteins induzieren und/oder stabilisieren.

Da V_L Domänen mit variierenden Längen des C-terminus unterschiedliche Aggregationskinetiken aufweisen, scheinen die Reste des Linkers zwischen der V_L Domäne und der konstanten (C_L) Domäne hierbei eine wichtige Rolle zu spielen. Es ist bereits bekannt, dass eine N-terminale Verkürzung der $C_L\lambda$ zu erniedrigter Stabilität und erhöhter Entfaltung führt. In dieser Arbeit wurden mit Hilfe verkürzter Linker-Varianten der Einfluss von spezifischen $V_L\kappa/C_L\kappa$-Linker Resten auf die $V_L\kappa$ und $C_L\kappa$ Domänen untersucht. $V_L\kappa$ und $C_L\kappa$ Varianten ohne den Arg108-Linker Rest waren zwar weniger stabil als das entsprechende Wildtypprotein. Während der $V_L\kappa$ Wildtyp (wt) keine Fibrillen bildete, zeigte die $V_L\kappa$-ΔR108 Variante eine erhöhte Fibrillenbildung. Faltungsexperimente zeigten, dass die C-terminale ΔR108 $V_L\kappa$ Variante schneller als der Wildtyp entfaltet, die Rückfaltung jedoch mit ähnlicher Kinetik verläuft. Diese Beobachtungen legen den Schluss nahe, dass das C-terminale Arg108 in den geschwindigkeitsbestimmenden Schritt der Entfaltungskinetik, jedoch nicht in den geschwindigkeitsbestimmenden Schritt der Rückfaltungskinetik der $V_L\kappa$ Domäne eingreift. Dies ist konsistent mit der Beobachtung, dass Entfaltungsereignisse zur Bildung von Fibrillen beitragen können, was wiederum erklären würde, weshalb die schneller entfaltende $V_L\kappa$-ΔR108 Variante eine stärkere Fibrillenbildung aufweist. Diese These unterstützend konnte außerdem gezeigt werden, dass die Deletion des N-terminalen Arg108 die Entfaltung von $C_L\kappa$ verlangsamt.

Zusammengefasst liefern die Daten der vorliegenden Arbeit neue Erkenntnisse für das Verständnis der Architektur von Antikörpern im Allgemeinen und die Vorraussetzungen für die Bildung von Fibrillen im Besonderen. Die gewonnenen Ergebnisse könnten auch zur Optimierung eines rationalen Designs von stabilen und funktionellen Antikörperfragmenten beitragen. Zusätzlich zeigen die beiden auf Grundlage dieser Arbeit vorgeschlagenen Mechanismen, die der Fibrillenbildung der V_L Domänen zugrundeliegen, dass sowohl destabilisierende als auch stabilisierende Interaktionen zu verstärkter Fibrillenbildung führen können. Hierbei stellt die Ursache dafür jedoch in beiden Fällen ein Entfaltungsereignis dar. Auf Grundlage dieser Arbeit könnten sich neue Wege der Erforschung von Fibrillenbildungsprozessen eröffnen, gerade auch für die Untersuchung anderer Proteine, die für eine Fibrillenbildung anfällig sind. Das Verstehen aller Mechanismen und Faktoren, die für die Bildung von Fibrillen verantwortlich sind, sowie des Prozesses der Amyloidogenese, können auch zur Entwicklung effektiver therapeutischen Strategien beitragen.

ABBREVIATIONS

aa	amino acid(s)
AFM	Atomic Force Microscopy
AFS	Alternatively Folded State
AL	Amyloid (antibody) Light chain amyloidosis
AH	Amyloid (antibody) heavy chain amyloidosis
ANS	1-anilinonaphtalene-8-sulfonate
APS	Ammonium Persulfate
AUC	Analytical Ultracentrifugation
B cell	Bursal or Bone marrow derived lymphocyte
BCR	B Cell Receptor
bp	base pair
CD	Circular Dichroism
CDR	Complementary Determining Region(s)
C_L	Antibody light chain constant domain
dNTP	desoxyriboNucleoside TriPhospate
DTT	DiThioTreitol
E. coli	*Escherichia coli*
EDTA	EthyleneDiamineTetraacetic Acid
EM	Electron Microscopy
Fab	Fragment antigen binding
Fc	Fragment crystalline
FWR	FrameWorkRegion(s)
GdmCl	Guanidinium chloride
GSH	L-Glutathione reduced
GSSG	L-Glutathione oxidized
HC	Antibody heavy chan
IB	Inclusion bodies
Ig	Immunoglobulin(s)
IgBLAST	Immunoglobulin Basic Local Alignment Search Tool
IPTG	isopropyl-β-D-thiogalactopyranoside
IMGT	international ImMunoGeneTics information system
kb	kilo base
kDa	kilo dalton
LC	antibody Light Chain
LB	Luria Bertani
m	mass
M	Molar

MW	Molecular Weight
mAb(s)	monoclonal Antibody (-ies)
NCBI	National Center for Biotechnology Information
NMR	Nuclear Magnetic Resonance
OD	Optical Density
PBS	Phosphate Buffered Saline
PDB	protein data bank
PCR	Polymerase Chain Reaction
rpm	rounds per minute
RT	Room Temperature
SE	Sedimentation Equilibrium
SV	Sedimentation Velocity
SDS	Sodium DodecylSulfate-PolyAcrylamide
SDS-PAGE	Sodium DodecylSulfate-PolyAcrylamide Gel Electrophoresis
Taq polymerase	Thermus aquaticus *polymerase*
TEM	Transmission Electron Microscopy
TEMED	N,N,N',N'-Tetramethylethylendiamine
Tris	tris(hydroxymethyl)-aminomethane
TAE buffer	Tris-Acetate-EDTA buffer
TB	Terrific Broth
ThT	ThioFlavin T
Temp	Temperature
U	unit (s)
UV	UltraViolet
V	Variable
V_H	Heavy chain variable domain
V_L	Light chain variable domain
v	volume
v/v	volume per volume
w	weight
wt	wild type
w/v	weight per volume (mass concentration)

Single letter codes for nucleobases

A   Adenine

C   Cytosine

G   Guanine

T   thymine

Triple and single letter codes for amino acids

alanine	Ala	A	leucine	Leu	L
arginine	Arg	R	lysine	Lys	K
asparagine	Asn	N	methionine	Met	M
aspartic acid	Asp	D	phenylalanine	Phe	F
cysteine	Cys	C	proline	Pro	P
glutamine	Gln	Q	serine	Ser	S
glutamic acid	Glu	E	threonine	Thr	T
glycine	Gly	G	tryptophan	Trp	W
histidine	His	H	tyrosine	Tyr	Y
isoleucine	Ile	I	valine	Val	V

The genetic code

Nucleotide position in codon					
first	**second**				**third**
	T	C	A	G	
T	TTT - **Phe** TTC - **Phe** TTA - **Leu** TTG - **Leu**	TCT - **Ser** TCC - **Ser** TCA - **Ser** TCG - **Ser**	TAT - **Tyr** TAC - **Tyr** TAA - **STOP** TAG - **STOP**	TGT - **Cys** TGC - **Cys** TGA - **STOP** TGG - **Trp**	T
C	CTT - **Leu** CTC - **Leu** CTC - **Leu** CTG - **Leu**	CCT - **Pro** CCC - **Pro** CCA - **Pro** CCG - **Pro**	CAT - **His** CAC - **His** CAA - **Gln** CAG - **Gln**	CGT - **Arg** CGC - **Arg** CGA - **Arg** CGG - **Arg**	C
A	ATT - **Ile** ATC - **Ile** ATA - **Ile** ATG - **Met**	ACT - **Thr** ACC - **Thr** ACA - **Thr** ACG - **Thr**	AAT - **Asn** AAC - **Asn** AAA - **Lys** AAG - **Lys**	AGT - **Ser** AGC - **Ser** AGA - **Arg** AGG - **Arg**	A
G	GTT - **Val** GTC - **Val** GTA - **Val** GTG - **Val**	GCT - **Ala** GCC - **Ala** GCA - **Ala** GCG - **Ala**	GAT - **Asp** GAC - **Asp** GAA - **Glu** GAG - **Glu**	GGT - **Gly** GGC - **Gly** GGA - **Gly** GGG - **Gly**	G

CHAPTER 1 – Introduction

1.1   Background on proteins and protein folding

Proteins are organic macromolecules essential in nearly all biological processes such as; catalysis, immunity, signaling, transport and storage, movement, conduction of nerve impulses, growth and differentiation. Every protein starts as a polypeptide, translated from a mRNA at the ribosomes as a linear sequence of amino acids with no specific three-dimensional arrangement and function. Therefore, in order to fulfil their biological tasks, proteins most adopt an intricate three-dimensional structure during or after their biogenesis (1). The genetic code specifies 20 standard amino acids and every protein has a unique amino acid sequence that is encoded by the nucleotide sequence of the gene. Despite the unique nature of their amino acid sequence, several homologous proteins structures determined by X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy (2-5) have been identified and grouped into various families (6). Large and complex proteins are often organized into multi-domain architectures (7,8). These domains or folding motifs are contiguous stretches of polypeptide chains that can fold independently into compact globular structures and form the building blocks of proteins structures. Studies of protein structure and the relationship between sequence and structure have been under intense investigation for several decades (9,10). However, how a protein folds and why it adopts or sometimes fails to adopt a particular native structure remains largely unresolved (11-14). In the heavily crowded milieu of the cell, there is a high risk of protein misfolding (15). However, nature has evolved mechanisms to cope with this (16-18). Molecular chaperones play a key role in this process, as they correct mistakes during folding and prevent unfolded proteins or partially folded intermediates from non-productive interactions (17,19,20). In addition to this, protein disulfide isomerases (PDI) and peptidylprolyl *cis-trans* isomerases (PPIase) act as folding catalysts in promoting the correct formation of protein structures (21). Proteins that cannot be reconstituted into their functional state are removed from the cell through the ubiquitin-proteasome system or by autophagy (22,23), before any serious harm can ensue to the host organism.

Since they are indispensable for the proper functioning of biological systems, proteins must be folded efficiently while minimizing the occurrence of mistakes that might result in a nonfunctional and/or disease-state of the protein (15,24). According to Anfinsen's dogma and other investigations, under appropriate conditions, proteins can fold spontaneously into their functional three-dimensional structure dictated by information contained within their primary amino acid sequence (25-27). Each amino acid can adopt several possible conformations, which implies that a protein of several amino acids could take years to attain its native conformation by random search (28,29). However, the seconds or microsecond time scales in which the process normally occurs (30), indicates that folding does not proceed through random sampling of all the accessible configurations that every single amino-acid of the protein can adopt in the conformational space (*Levinthal paradox*) (28). Protein

folding must therefore be a directed process with pathways that reduce the search space leading to the most energetically stable conformation in the short time scales usually observed (28,31-33). Chaperones are often required to assist protein folding *in vivo*. However, they only do increase folding efficiency but not the specificity of the fold (19). Since a protein's conformation is dictated by its amino acid sequence, understanding the relationship between sequence and structure might enable prediction of folding outcomes including those involved in protein misfolding diseases (34). Significant progress has been made in this respect supported with the availability of databases correlating structural elements with sequence (13), but with still no definite solution for predicting structure from sequence (35). For example, why and how only some antibody domains readily misfold into pathogenic amyloid fibrils despite a huge similarity in their amino acid sequence remains an open question (36). Moreover, it is enigmatic why all amyloid fibrils share a similar structure of high β-sheets content irrespective of their precursor protein's primary amino acid sequence (37,38). On the other hand, prediction of structure based on sequence could lead to *de novo* design of new proteins (39,40) such as those used as biotherapeutics.

1.2   Conformational states of proteins

The native conformation of a protein is the most thermodynamically stable, essentially unique structure, characterized by its ability to fluctuate through a number of isoenergetic minima (conformational substates). This conformation of proteins depend both on the intrinsic properties of its amino acid sequence and on its environment. Native proteins are organized into defined secondary, tertiary and quaternary structures. The secondary structure elements are loops, turns, α-helices and β-strands, characterized by defined hydrogen bonds of the amide and carbonyl atoms of the polypeptide backbone. The tertiary structure is the three dimensional (3D) orientation of the secondary structure elements in space within a polypeptide chain and it controls the basic function of the protein. The quaternary structure is the 3D orientation of two or more polypeptide chains towards each other in oligomeric or multi-subunit proteins.

The dominant driving forces in protein folding and association are hydrophobic effects (41,42) and noncovalent contacts such as H-bonds and salt bridges (43). The hydrophobic effect arises from the fact that water molecules prefer to form H-bonds with each other or with other polar molecules which therefore hinders the exposure of hydrophobic surfaces to the solvent (43,44). In this process the water network squeezes the protein to maintain a native state in which non-polar groups such as alkyl side chains are buried in the core of the protein (burial of non-polar surface area) (45). This contributes to the favorable free energy which arises from the decrease of "ordered" water molecules surrounding non polar residues. It is therefore worth noting that the hydrophobic effect is not mainly due to the interactions between side chains of hydrophobic amino acids *per se* (mainly van der Waals forces), but rather to the increase in entropy, gained by the removal of hydrophobic surfaces from ordered solvating water (46,47). While on the one hand, hydrophobic interactions are crucial for the maintenance of the structural integrity of proteins as elucidated in this work for antibody variable domains (48), on the other hand, under destabilizing conditions, hydrophobic interactions can result in

protein aggregation (49), which implies that protein aggregation and in particular fibril formation is a common property of any polypeptide chain modulated by its surrounding milieu and by its amino acid sequence (50-52)**.** This is also a hint that the process of productive folding and aggregation might be governed by the same mechanisms involving non-specific interactions like hydrophobic interactions and H-bond formation. With the latter known to stabilize polar groups occurring within the hydrophobic core and NH and C=0 groups found in the backbone of polypeptide (53,54). Besides H-bonds and van der Waals interactions which play an important role in protein folding and stability, electrostatic interactions and disulfide bonds also play a crucial role (43). Electrostatic interactions of charged or ionized side chains can be involved in salt bridge formation, which are short-ranged well oriented interaction that stabilize the folded conformation through ion pairing (43,55). Conversely, electrostatic interactions that are long-ranged, directionless, non-specific coulombic interaction are also critical for protein folding, stability and function. Secondary structure formation is a result of specific pairwise interactions of side chains with the help of H-bonds and salt bridge formation (55). The covalent bond formation between two Cys residues (disulfide bonds) influences the folding and stability of proteins in general (43), and antibody domains in particular (36,56). All these forces are certainly also implicated in aggregation processes.

1.2.1   The protein folding landscape

The process which occurs as a protein folds from its unfolded state (U) to the native state (N), driven by the balance of entropy and enthalpy can be described using the concept of a "*folding funnel*" free energy landscape (Fig. 1.1) (1,57-60). In this multidimensional energy landscape, each point on the surface describes a specific conformational state with the minimum representing the folded and most stable state of the protein. In the simplest case, proteins fold in a two-state reaction to their native conformation. However, folding pathways are often characterized by multiple routes with often transient intermediates (61-64). These intermediates represent local minima on the energy landscape. They can either be on- or off-pathway to the native state and are characterized by a substantial fraction of native-like secondary structure elements (65).
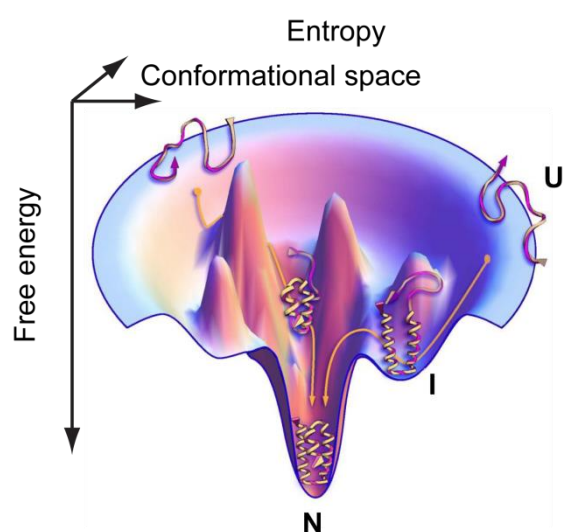
Figure 1.1. Representation of the funnel-shaped free energy landscape of protein folding. It is rugged with kinetic traps and folding occurs via alternative microscopic trajectories or pathways (orange arrows). The broad accessible conformational space corresponds to the high energy unfolded states (U) explored through the balance of entropy and enthalpy. The intermediates (I) are characterized by local energy minima whereas the lowest minimum with fewest accessible conformations and lowest energy characterizes the folded native structure (N). Adapted and modified from Dill and MacCallum (2012) (66).

A complex rugged structure best describes the surface of the funnel. The smoother the surface is, the more rapid the folding process (64). As amino acid residues in the unfolded or partially folded polypeptide samples different conformations, different native as well as non-native contacts are formed. Native interactions between residues are assumed to be energetically more favored (stable) over non-native contacts and as these contacts are formed, the number of conformations reduces thereby driving the polypeptide chain towards the deep minimum of the funnel. Nevertheless, local energetic minima can be encountered, which are kinetic traps that slow down the folding process. These local minima represent intermediates, partially folded conformations that might be obligatory for a polypeptide to attain its native state. Some intermediates can be intrinsically prone to aggregation and when significantly populated, they can specifically interact with each other to form aggregates (67,68). One well described intermediate is the molten globule state, characterized by significant amount of native-like secondary structure (compact state; globule) with only few tertiary interactions (molten tertiary interactions) and a marginal stability (69,70). Another intermediate state is the so called "alternatively folded state" (AFS) formed at low pH, AFS shares some similarity with the molten globule as it also shows a high secondary structure content, but nevertheless adopts a clearly distinct structure to that of the native state as the name implies (71-74). The AFS usually possesses a defined quaternary structure, usually highly stable as shown for some immunoglobulin domains (74). Under a given set of conditions, the surface of the energy landscape is unique for every polypeptide chain and this is determined by its thermodynamic and kinetic properties (31,75,76). Since common mechanisms like the search for a set of critical contacts are thought to govern protein folding and aggregation, both processes compete with each other (1).

1.2.2   **Protein stability**

For every protein, an activation energy barrier separates the native state (N) from the unfolded state (U) and this has to be surmounted for the protein to switch between the two states as shown in the free energy diagram (Fig 1.2). The barrier is influenced by the contributions of the enthalpic and entropic terms. Protein stability depends on free energy changes (ΔG), best described by the Gibbs free energy of activation ($\Delta G = \Delta H - T\Delta S$, where ΔH is the enthalpic and ΔS is the entropic term) (77). While the enthalpic term is accounted for by electrostatic interactions, hydrogen bonding and van der Waals interactions that contribute to desolvation effect (78), hydrophobic interactions account for the entropic term, which decreases with increasing order of the protein backbone (structural complexity) due to restriction of chain configuration. The smaller the activation energy barrier, the faster the protein will fold (79).
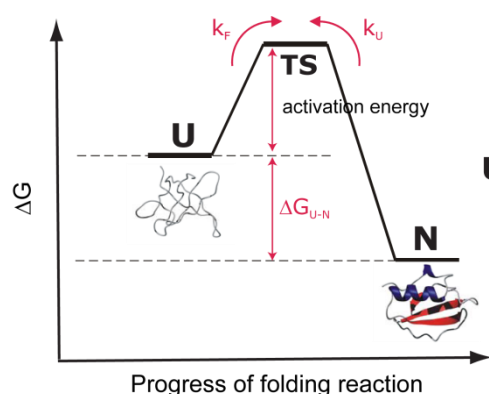
Figure 1.2. Energy diagram of a two-state protein with states populated during folding. The protein in its unfolded (or denatured) state (U) folds via the transition state TS to the folded, native state (N). The barrier to folding in each direction has a rate constant K_F or K_U. The overall change in free energy between the folded and unfolded state is given by ΔG_{U-N}. Adapted and modified from Simpson, 2006 (80).

ΔG_{U-N} is the difference in energy between the unfolded (U) and native (N) states assuming two state transitions obeyed by many proteins under equilibrium conditions (Fig 1.2). The free energy difference between these states is therefore given by ΔG_{U-N}:

$$\Delta G_{U-N} = -RT \ln K_{U-N} \qquad [1.1]$$

Where *R* is the universal gas constant and *T* is temperature. K_{U-N} is the equilibrium constant for the folded and unfolded states and is defined by the rate constants of the forward and backward reactions:

$$K_{U-N} = \frac{K_F}{K_U} \qquad [1.2]$$

Similarly, for unfolding, i.e from the native state to the unfolded state, the free energy difference is given by ΔG_{N-U}:

$$\Delta G_{N-U} = -RT \ln K_{N-U} \qquad [1.3]$$

$$K_{N-U} = \frac{K_U}{K_F} \qquad [1.4]$$

The equilibrium lies mostly to the native state (ΔG_{U-N} is negative) for most proteins in water. Therefore to study unfolding and to assess protein stability, the protein must be unfolded by an external perturbant such as heat, pH or a chemical denaturant. The unfolding and refolding of a protein can be measured and only if the folding process is reversible can the ΔG value be determined.

To study protein folding or stability, proteins are often denatured using a chemical denaturant such as chaotropic agents like urea or guanidinium chloride (GdmCl) (81) or by thermal denaturation (82). In addition, pH and pressure dependent denaturation are also possible (82). In very high concentrations of GdmCl and urea, proteins are considered to be in a 'denatured state' (closely approximated to the unfolded state) (81), which is a highly open and solvent-exposed conformation with little or no residual structure (83). However, the mechanism of denaturation by these chemicals is not fully understood (84). Generally, it is thought that they weaken hydrophobic interactions and stabilize the denatured state by solvating non-polar groups better than water (83,84). Alternatively, it is assumed that the denaturant binds to the polypeptide backbone that has more binding sites in the unfolded state than in

the native (85,86). The degree of denaturation is dependent on the concentration of the chaotropic agent and according to both models, there is a linear dependency between the empirically determined free energy of unfolding and the denaturant concentration within the region of the unfolding transition (83,86). Temperature or extreme pHs also mediate their denaturation effects by disrupting noncovalent interactions which define native conformations (82). The extent of un- or re-folding can be determined by using one of a variety of techniques used to monitor protein structural changes, such as fluorescence, absorbance, NMR or CD spectroscopy (see materials and method section).

1.3   Protein misfolding and aggregation diseases

Understanding the mechanisms underlying protein misfolding and aggregation has gained a lot of scientific attention in the last decades with the discovery that these processes are at the basis of several life-threatening human diseases with a huge social burden. Common examples of these diseases are neurodegenerative disorders likes Alzheimer`s, Huntington`s and Parkinson`s disease (87), spongiform encephalopathies such as Creutzfeldt-Jakob disease (88) and amyotrophic lateral sclerosis (89). Other examples are the less known but highly fatal antibody light chain amyloidosis or dialysis-related amyloidosis (90,91). All these diseases are characterized by the extracellular (and/or intracellular) deposition of insoluble amyloid fibrils. The fibrils are derived from the aggregation of misfolded, normally soluble proteins (Fig 1.3) (1).

Protein folding or misfolding is as a result of conformational changes, which occur at a relatively low energy cost. In single polypeptides, these conformational switching is due to changes in main-chain torsional angles and side chain orientations. The changes may be localized, with reorientation of some residues and if this occurs in critical regions of the polypeptide it might lead to reorganization of the whole protein structure. Such reorganizations in protein structure might lead to aggregation by affecting the thermodynamic drive towards this diseased-associated configuration (92-95). Specifically, this leads to a weakening of tertiary interactions while simultaneously favoring secondary interactions like H-bonding. This might also result in the exposure of normally buried hydrophobic residues to the solvent and thus cause the protein to misfold into a molten-globule like structure. As the aggregation process is often favored by usually only a partial unfolding of the protein, only a small decrease in the stability is required. Therefore, conditions that destabilize the native state often lead to aggregation. Specific mutations in the polypeptide sequence that lead to conformational changes and/or a loss in stability appear to play a pivotal role in these pathological protein aggregation as shown in this work (48) and as reported by others (96-98). While the transformations induced within the protein through conformational changes are thought to be the main molecular driving force behind protein aggregation, cellular events such as the failure or dysregulation of the proteolytic process in Alzheimer's disease (99), or the overproduction of a precursor protein as in AL amyloidosis (100), are the main cause for the accumulation of aggregates.
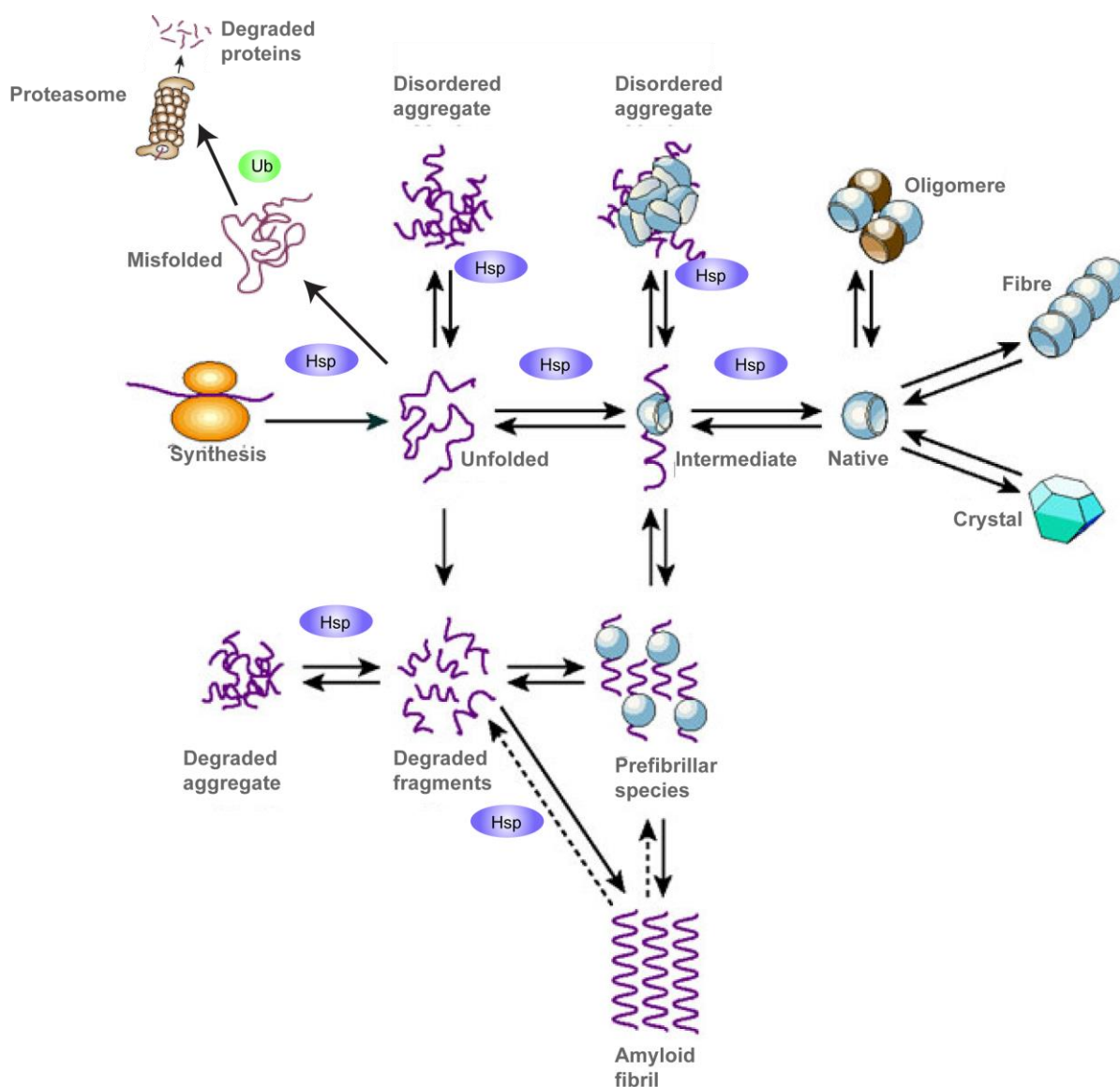
Figure 1.3. Schematic representation of structures that can be formed by polypeptide chains. The populations and interconversions of the various states are determined by their relative thermodynamic and kinetic stabilities under given conditions. In living systems, however, transitions between the different states are highly regulated by the environment and by the presence of molecular chaperones, proteolytic enzymes and other factors. Failure of such regulatory mechanisms is likely to be a major factor in the onset and development of misfolding diseases. Adapted and modified from Dobson, 2003 (1).

Table 1.1 presents some common examples of amyloid diseases. Clinically two forms of amyloid deposition diseases can be distinguished; (1) Local amyloidosis, here the amyloid is localized or restricted to a particular organ or tissue like in neurodegenerative disorders such as in Alzheimer's disease with a restricted localization of amyloid fibrils (plaques) to the brain (89), and (2) systemic amyloidosis, where the amyloid fibrils can be deposited in multiple organs or tissues, the viscera, blood vessel walls and connective tissues like in the case of systemic antibody light chain (AL) amyloidosis (90,100,101). Investigation of the deposits seen in patients with these diseases have shown them in many cases to be highly structured, consisting of fibrils of repeating β-strands stacked parallel to each other (37,38). Results are indicating that amyloid structures are an inherent part of the protein folding landscape (1). However, for the majority of proteins this state is not significantly populated.

7

Table 1.1 Common human amyloid diseases and their precursor proteins.

Disease	Precursor protein	Organ(s) affected	Systemic or local
Light chain amyloidosis (AL)	Antibody λ or κ light chain	Kidney, heart, liver, etc	Systemic
Heavy chain (AH)	Antibody heavy chain	Kidney, liver, etc	Systemic
Dialysis-related amyloidosis (Aβ$_2$m)	β2- microglobulin	Joints, heart, lung, etc	Systemic
Senile systemic amyloidosis (SSA) and Familial amyloidotic polyneuropathy (FAP), both (ATTR)	Transthyretin	Heart, kidney, vessels, etc	Systemic
Alzheimer`s disease	Amyloid β-peptide tau	Brain	Local
Parkinson`s disease	α-synuclein	Brain	Local
Creutzfeldt-Jakob disease	Prion protein	Brain	Local
Huntington`s disease	Huntingtin	Brain	Local
Amyotrophic lateral sclerosis (ALS)	Superoxide dismutase	Spinal motor neurons and motor cortex	Local

Taken from (90,102,103)

1.3.1   Principles of Amyloid fibril formation and fibril structure

In vitro experiments suggest that any protein can convert to amyloid fibril, thus indicating that the formation of such stable structures occur on an alternative folding pathway to that upon which the native state is formed (Figure 1.4) (104,105). It has been suggested that a protein trapped in a non-native intermediate upon folding becomes the precursor to amyloid formation (96). Alternatively, intermediates populated on the native folding pathway might be the amyloid precursor (Figure 1.4) (67). Nevertheless, the final conformation adopted often depends on the relative thermodynamic stabilities of the various accessible conformations and on the kinetics of their interconversion (1). Factors that favor accumulation of intermediates greatly predispose to amyloidosis (97,106). Variations in environmental conditions like pH, temperature, ionic strength, addition of denaturant or the presence of specific mutations play an important role in this respect. Therefore, identifying the factors that might lead to accumulation of fibril-prone intermediate and the dissecting of the folding landscape are both necessary for a thorough understanding of the mechanisms of amyloid formation.

Unfolded state   Intermediate state   Native state

Aggregates   fibril core structure   Mature fibrils
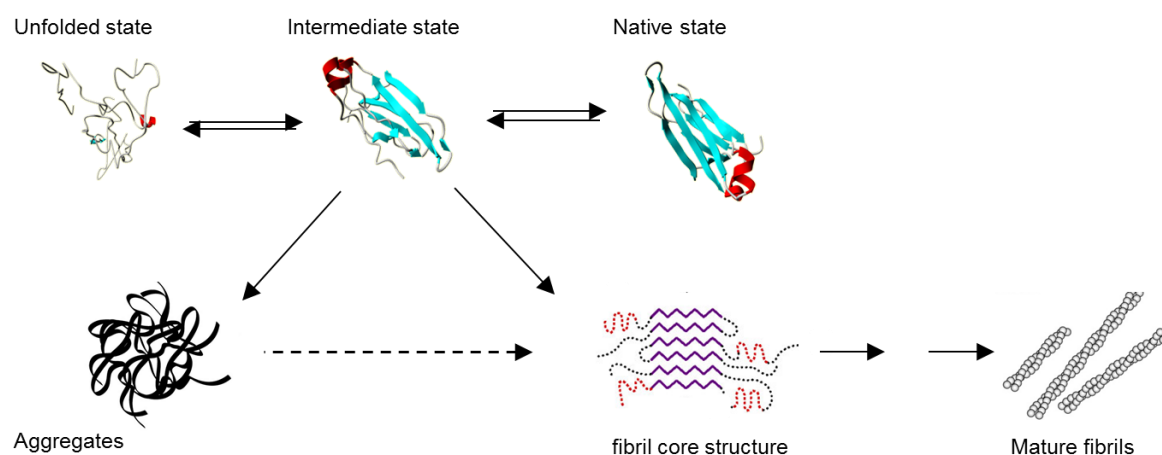
Figure 1.4. Proposed model for protein folding, misfolding and fibril formation. Several intermediates (both native-like and unfolded) can be populated *en route* to the native conformation (here just one is shown for simplicity).These intermediates are postulated to be the species that give rise to amorphous aggregates and/or amyloid fibrils. Factors that favor accumulation of intermediates predispose to amyloidosis. Adapted and modified from Baden et al., (2009) (106).

Based on experimental evidence, a nucleation mechanism is often reported to be at the basis of the growth of amyloid fibrils. The process begins with the transient accumulation of β-sheet-enriched oligomers to form a critical nucleus during the lag phase, an event usually difficult to trigger due to the high free-energy barrier (107). This high energy oligomer acts as bottle-neck that limits the rate of fibril formation. In other instances, self-association may result in the formation of amorphous aggregates that might facilitate nucleation (Fig. 1.4) (106). Once the nucleus is formed, the reaction proceeds to the elongation phase with a rapid growth of the fibrils through the addition of monomers (Fig. 1.5a) (107,108). A model proposed by Ionescu-Zanetti and colleagues suggested that two filaments combine to form a protofibril (2-5 nm diameter) and two protofibrils intertwine to form a type 1 fibril (~ 20 nm in diameter) (Fig. 1.5d) (109). Generally, the whole assembly pathway involves conformational transitions during and after nucleation (96,110-112)**.** Fibrillar aggregates are characterized by a common highly organized H-bonded structure that might give them their unique kinetic stability. All amyloid fibrils share a common core filament structure irrespective of the nature of their precursor proteins. X-ray and electron diffraction studies of amyloid fibrils have confirmed the general presence of a cross-β helical structure with β-strands that make up β-sheets. The β-sheets have their strands running perpendicular to the long axis of the fibrils and hydrogen bonded along the axis of the fibrils (37,38,113,114). Filaments are composed of two or more β-sheets that are stacked normal to the helical axis and extend along it (Fig. 1.5d-iii). Two or more filaments make up a fibril. The rate of fibril formation largely depends on the amount of pre-existing fibrils and seeding experiments have revealed that the presence of fibrillar seeds of a precursor protein leads to its rapid conversion to amyloid fibrils (115). Typical for all amyloid fibrils is the binding of the dye Congo red with a characteristic green birefringence under polarized light (116,117) and the binding of the benzothiazole dye Thioflavin T that results in a shift and an increase in fluorescence intensity (118,119). The fact that even proteins not known to be associated with any misfolding disease can be induced to form amyloid structures, along with the generic nature of their fold, implies that the ability to form this structure may be an inherent property of proteins (32,97).
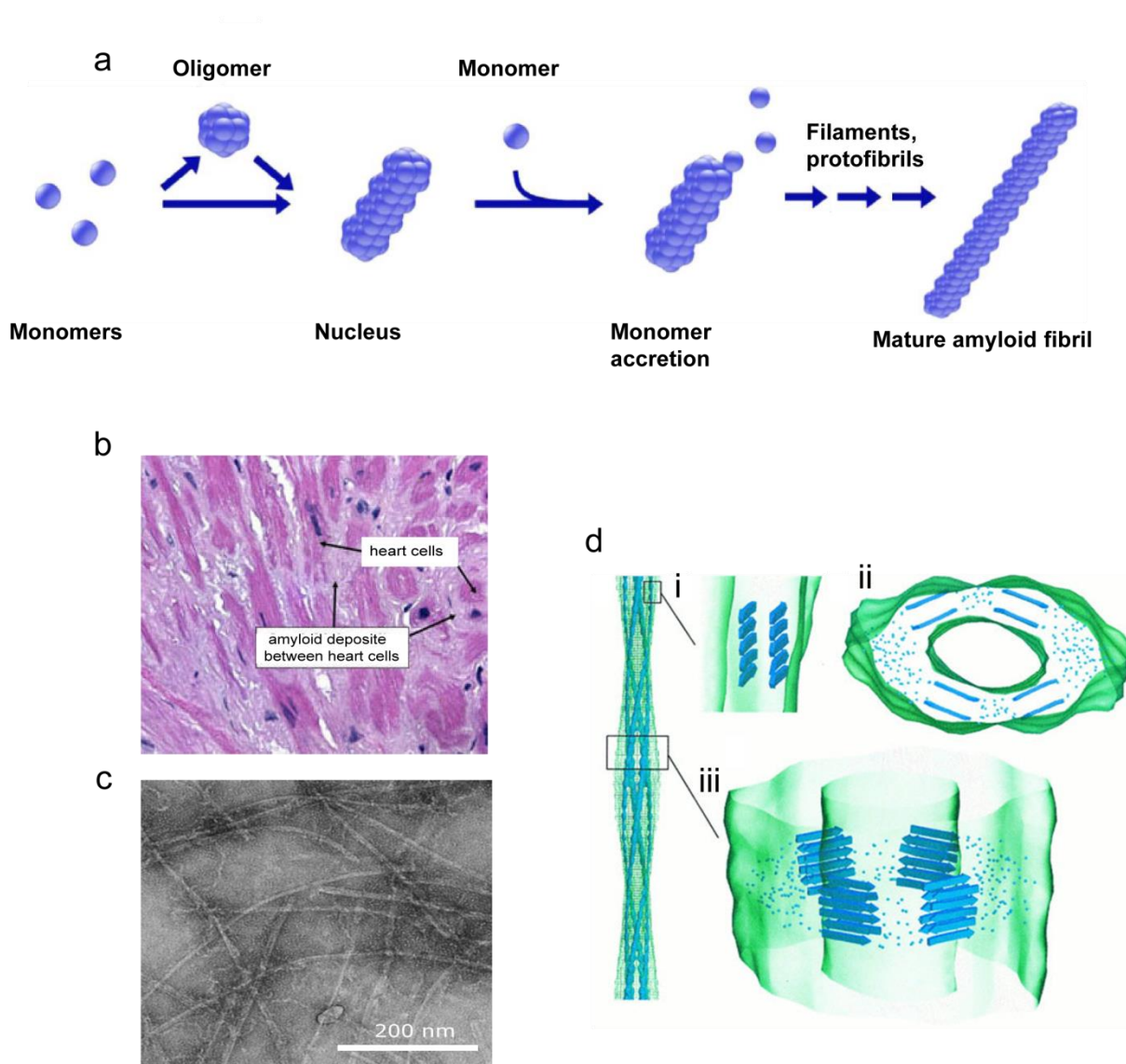
Figure 1.5. (a) Model of nucleation-dependent fibrillation, adapted from Bhak et al., (2009) (120). (b) Heart biopsies from patients with AL stained for amyloid fibrils (121). Amyloid fibrils are stained light pink as shown by arrow heads. (c) Electron microscopic micrographs of amyloid fibrils produced in vitro from an antibody V_L domain. (d) Modelling the polypeptide fold in the fibril density from an SH3 domain here reported as a representative example of amyloid structures based on cryo-EM. Outer surface is shown as green wire mesh and the protofilaments as solid blue surfaces. The protofilaments are composed of beta-sheets containing a mixture of parallel and anti-parallel strands while the remaining regions of the polypeptide sequence are shown as disconnected dots. (d-i) Side view of a protofilament. (d-ii) Cross-section and (d-iii) slightly tilted side view of the fibril. Adapted and modified from Jimenez et al., (1999) (114).

1.4   Antibodies

Antibodies, generally also referred to as immunoglobulins, belong to the immunoglobulin (Ig) superfamily of proteins that also include T cell receptors and major histocompatibility complexes (122,123). These molecules are produced by B cells either as surface-bound antigen receptors or as secretory products (antibodies) by differentiated B cells (plasma cells) and are the main components of the humoral immunity, which is essential for adaptive immune defense mechanisms (124). Immunoglobulins or antibodies provide protection through neutralization, complement activation, opsonization and subsequent clearance by cellular immune effectors (125).

10

Antibodies are multi-domain glycoprotein complexes that consist of two identical heavy chains (HCs) and two identical light chains (LCs) that are linked by disulfide bonds (Fig. 1.6a). In higher vertebrates, Igs exist in five different isotypes defined by the primary structure of their HC; Ig mu (M), Ig delta (D), Ig gamma (G), Ig alpha (A), and Ig epsilon (E). For the LC, two classes have been defined; Ig-kappa (κ) and Ig-lambda (λ), which are mutually exclusively utilized. IgG is the predominant antibody isotype with the longest serum half-life. Four IgG subclasses (IgG1-IgG4) are distinguished according to differences in structure, antigens and effector function (126) of which IgG1 is the most frequent. Despite differences in amino acid sequences, the domain structure, the Ig fold is highly conserved. It is a beta-barrel of approximately 110 amino acids in which the two halves of the barrel are linked by a disulfide bond, which is also highly conserved (127). This basic tertiary structure (Fig. 1.6b-d) is shared by all members of the Ig superfamily. The N-terminal domains (V_H and V_L) of both chains lend antibodies their antigen-binding specifics while the C-terminal (constant) domains mediate effector functions. Engineering these domains to improve their function has been a major effort in several academic institutions and biotech companies (128-132).

The V domains belong to a beta-barrel structure subclass different from that of the constant domains. Their anti-parallel beta-pleated sheets comprise nine β-strands in a 4 + 5 orientation (ABED/CC'C''FG) (Fig. 1.6c) (133), while those of the constant domains are made up of seven β-strands in a 4 + 3 orientation (ABED/CFG) (Fig 1.6d). Another major difference between the constant and the variable domains is the presence of complementarity determining regions (CDR1-3) in the latter. These are diverse stretches of amino acids (hypervariable loop regions) interspersed between a highly conserved framework. CDRs link β-strands within the variable domain and are responsible for the antigen specificities. Despite the small number of germ line-encoded immunoglobulin variable gene segments, B cells are able to produce antibodies with variable domains of very high diversity that can recognize an almost infinite number of antigens. This is made possible by the stoichastic recombination of variable (V), diversity (D) and joining (J) gene segments – V(D)J recombination (134,135). Diversification is further increased in antibodies derived from B cells that have undergone the germinal center reaction comprised of somatic hypermutations, which leads to introduction of mutations to the variable region (136). However, these mutations can sometime lead to deleterious effects such as in autoimmune diseases (137,138) and in antibody LC aggregation diseases (139).
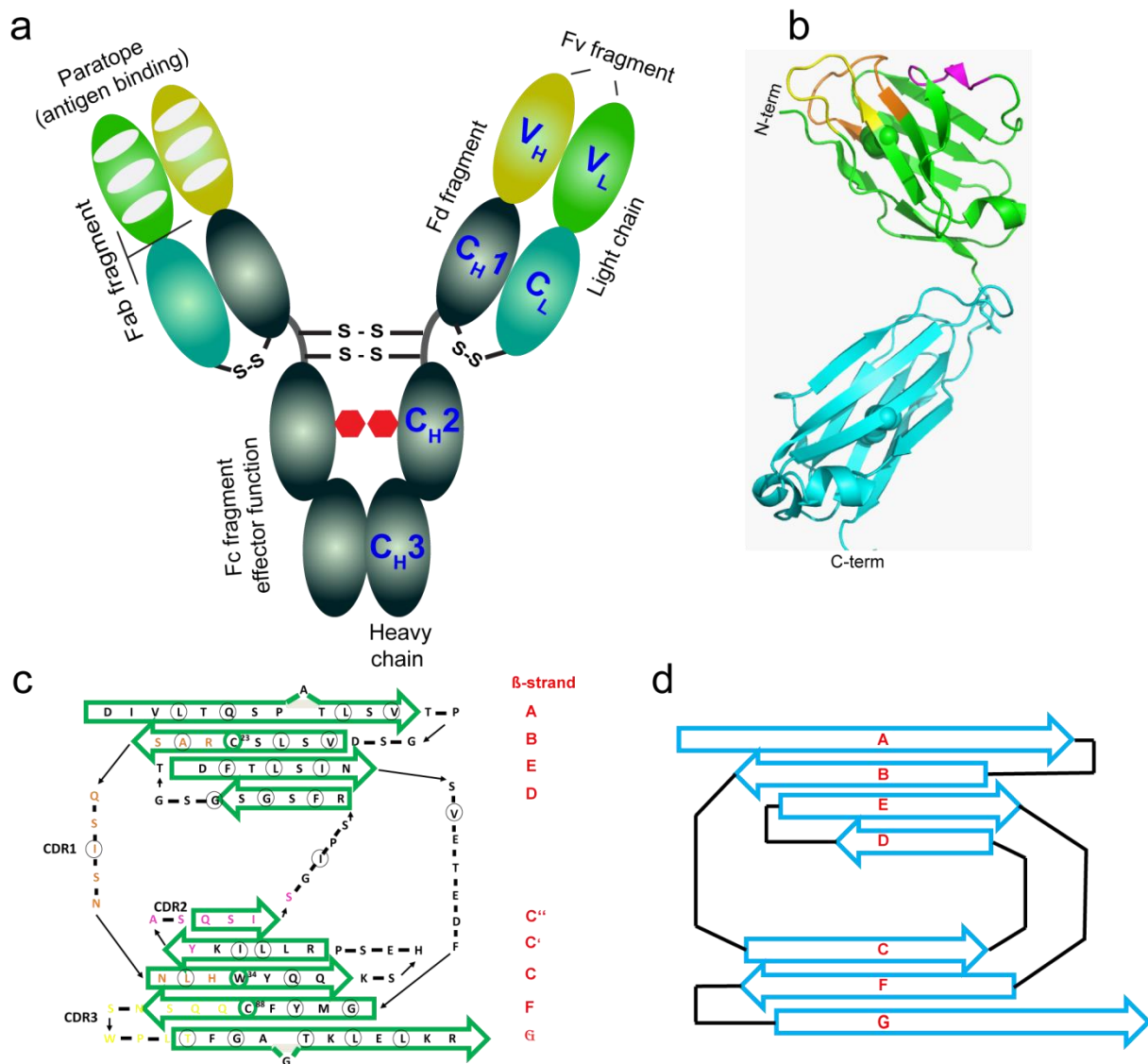
Figure 1.6. (a) Domain organization in an IgG antibody molecule. S-S indicates a disulfide bridge and red hexagons represent glycans. Functional elements are shown; the F_V (variable domain fragment), Fab (fragment antigen binding, a heterodimer of V_H/C_H1 (Fd) and the LC) and the FC (fragment crystallizable, a heterodimer of C_H2 and C_H3). (b) Structure of an LC. CDRs within the N-terminal V_L domain are highlighted in orange (CDR1), magenta (CDR2) and yellow (CDR3). The conserved disulfide bridge is shown in van der Waals spheres. The structure is based on MAK33 IgG1/κ (PDB: 1FH5). (c) Topology diagram of variable domains, modelled here for MAK33 V_L, encircled residues point towards the core (including the two Cys that form the disulfide bridge, encircled green), and (d) constant domains, in a 2 dimensional representation. This Ig fold topology is shared by all members of the Ig superfamily. Adapted and modified from (133,139,140).

Stringent cellular quality control mechanisms exist that ensure that only completely assembled antibodies can be secreted by B cells (Fig 1.6a) (141,142) (143). But sometime, things go awry as in neoplastic plasma cells, resulting in particular to the excess secretion of free LCs (Bence Jones proteins) (Fig 1.6b). Misfolding and aggregation of these free LCs are the basis of primary systemic LC amyloidosis (100,144).

1.5   Systemic antibody light chain (AL) amyloidosis

Amyloidosis is a pathological condition caused by the extracellular and/or intracellular deposition of insoluble abnormal fibrils in organs and tissues. These fibrils are derived from the aggregation of misfolded, normally soluble proteins (1,145). Systemic antibody (amyloid) light chain (AL) amyloidosis is a fatal protein misfolding disease (90,146). Despite being a rare disease, it is the most common and most fatal systemic amyloidosis, causing about one per thousand deaths in industrialized nations (145,147). It is usually secondary to an otherwise benign low-grade monoclonal gammopathy and sometimes complicates multiple myeloma or other clonal B cell dyscrasias. These aberrant clonal B cells overproduce unpaired antibody light chains (LCs), which are secreted into the bloodstream and taken up by multiple organs and tissues where they aggregate, severely compromising the functions of the affected organs and ultimately leading to death (91,100).

Amyloid fibrils of antibody LCs can be deposited in any part of the body except the brain, with the most affected organs being the heart, kidney and the tongue. Analysis of proteins isolated from AL patient biopsies show that the main constituent of amyloid deposits is the variable domain of the light chain (V_L) (101,145). In another antibody aggregation disease called light chain deposition disease (LCDD), characterized by punctuate amorphous deposits in the basement membrane of glomerulus and other tissues, the entire LC is usually the main constituent (101,148,149).

1.5.1   Underlying cause of fibrillation in AL amyloidosis

Due to the sequence variability in their variable domains, LCs comprise structurally the most diverse family of proteins involved in amyloidosis (150). In AL amyloidosis, each patient has a unique monoclonal LC that misfolds into amyloid fibrils (139,151,152). Numerous evidence suggests that AL proteins are less stable than their non-amyloidogenic counterparts (153-155). Thus, many antibody LCs incorporate structural features that impair their stability and solubility, leading to their assembly into fibrils and their subsequent pathological deposition (150). On the one hand, intrinsic factors like the acquisition of destabilizing mutations (139) (98,156) and the proteolytic cleavage of the light chain constant (C_L) domain (152), have been linked with fibrillation of the V_L domain. *N*-linked glycans that may be incorporated as a result of the acquisition of *N*-linked glycosylation sites (Asn-Xaa-Ser/Thr) through somatic hypermutations (150), are more frequent in AL proteins (157,158). Despite this evidence, the precise role of glycosylation in AL proteins is yet to be determined. Disulfide-linked dimerization, S-cysteinylation, fragmentation, S-sulfonation, 3-chlorotyrosine formation and conversion of aspartic acid to pyruvate (159) have also been implicated with AL protein aggregation. However, the exact relevance of these modifications for AL pathogenesis is unknown (106).

On the other hand, extrinsic factors such as environmental conditions also influence AL protein aggregation and their subsequent pathological deposition; for example, a given LC protein may be amyloidogenic in one individual but not in another or in a given tissue or organ of an individual and not in another due to differences in the surrounding environment (160). Components of the extracellular matrix including heparan sulfate, dermatan sulfate proteoglycans and glycoaminoglycan chains (161)

have been found extensively in amyloid deposits. Some of these strongly interact with fibrils (162-164). These molecules may contribute to fibrillogenesis or lead to stabilization of existing amyloid structures (165,166). Moreover, all amyloid deposits contain serum amyloid P component (SAP), a circulating plasma protein of the pentraxin family (167). SAP reversibly binds with high affinity in a calcium-dependent manner all types of amyloid fibrils. The binding of SAP which is itself highly resistant to protease digestion, stabilizes amyloid fibrils and protects them from degradation by proteases and phagocytic cells (168). In addition to these, high protein concentration (169,170) and the presence of "seeds" of preformed fibrils (171) have been reported to accelerate the rate of fibril formation.

In addition to these natural factors that might render an LC protein amyloidogenic, several *in vitro* studies have also shown that destabilizing conditions such as high temperatures an extreme pH (172,173), pressure and agitation (174) , ultrasonic irradiation (107,175), denaturants (172,173,176-178), ionic strength (179) and detergents (180-183), favor the accumulation of partially states that are prone to aggregation or to fibril formation, by decreasing the high-energy barrier to nucleation (107). Despite all these evidence, most LC proteins are non-amyloidogenic, which leaves the question of why some LC form fibrils unresolved.

1.5.2   State-of-the-art in LC amyloid formation

Since a majority of LCs are resistant to amyloid formation, several amyloidogenic and non-amyloidogenic V_L domains have been compared, with the view to decipher the determinants (139,150,153,154,159,173,177,184-186). Most of these studies have identified sequence and structural features that can predispose to amyloid formation, but these are often limited to given V_L subfamilies (e.g. κ1 or λ6) with hardly one feature pertaining to all or an entire variable domain family (187-189). It would be interesting to add the consensus sequences or features of these subfamilies. A multitude of studies focused on factors that enhance or inhibit fibril formation (97,106), the contribution of germline gene usage (159,187,189,190), and the effects of somatic mutations on amyloidogenicity and on the structure of AL proteins (98,191,192). Comparative studies still continue to yield intriguing correlations between the identity, location or modification of certain residues and amyloidogenicity. Prediction of the amyloidogenic propensity of a given antibody LC based on the listed predisposing factors should be possible in the future and this might help in the design or selection of candidate therapeutic monoclonal antibodies.

Other studies investigating the relationship between the stability of the domain and its amyloid potential have been undertaken (98,156). However, the determinants and mechanism of antibody fibril formation remain unresolved. In recent years, progress has been made in understanding the folding mechanism of antibodies *in vitro* and *in vivo* (summarised in (36)). These studies not only allowed the stability and function of IgG domains to be rationally influenced (193-195), but also helped to clarify fundamental principles underlying the folding of immunoglobulin domains in general (193,196-198). To date, a key question in amyloid research remains the molecular mechanisms by which amyloid fibrils

or the process of amyloidogenesis cause cellular toxicity and organ damage (94,104,145,199-207). The elucidation of the amyloid forming pathway of AL proteins with a clear identification of species (intermediates) *en route* to the fibrils is yet to be determined. Moreover, structural features that predispose AL proteins to fibril formation are not completely known. The use of biochemical techniques and high-end biophysical and computational methods will not only shed light on the structural basis of the disease, but will also lead to deeper insights into the nature of the toxic species involved in AL amylodosis and hence the design of therapeutic options that directly target such species.

1.6   Objectives of the thesis

The objectives of this thesis are in three folds:

i.   To analyze the role of conserved framework residues for the integrity and amyloidogenicity of variable domains.

Despite the high similarity between the V_L and V_H domains, V_H is less susceptible to amyloid fibril formation (208). Furthermore, why some V_Ls form amyloids and others not, still remains largely enigmatic. In general, it is therefore not clear which residues are of special importance for the structural integrity of the V_L domain. Highly conserved consensus sequences within the framework region are potentially important elements in this respect. A striking feature of variable domains is the conserved N-terminus of β-strand A (FWR1). However, the role of the N-terminal residues of the LC on domain structure, stability and amyloidogenic potential are still unresolved. The V_Lκ domain from MAK33 IgG1/κ, a well characterized antibody (56,71,74,193,209), is known to be resistant to fibril formation (193). For MAK33 Fab, a crystal structure is available (PDB: 1FH5) (210). Therefore, by comparing MAK33 V_Lκ with other variable (V_Lκ, V_Lλ and V_H) domains, a detailed analysis of the importance of the conserved N-terminal residues for the structural integrity, the conformational stability and amyloidogenicity will be performed. This will be important for understanding the principles of antibody variable domain structure and amyloidogenicity.

ii.   To identify underlying mechanisms for V_L fibrillation.

The loss or gain of certain residues through somatic mutations may have a global destabilizing effect on AL proteins and as a consequence these proteins require less free energy to unfold (98,211). However, despite evidence for the contribution of low thermodynamic stability to the amyloidogenicity of AL proteins, studies have also shown that the propensity to form fibrils for some V_L domains is inversely correlated to their free energy of unfolding, suggesting that both stabilizing and destabilizing interactions within the V_L domain can influence the kinetics of fibrillation (154,155,212). The underlying mechanisms behind this are not understood. In this work, naturally occurring amyloidogenic mutations will be introduced into the MAK33 V_Lκ domain and the different V_L variants will be characterized with a view to define the reasons for amyloidogenic predisposition.

iii. To analyze the effects of the V_L/C_L-linker (boundary) residues on the V_Lκ and C_Lκ domain.

The identification of mostly the V_L fragments within aggregated deposits in AL patients (148) is an indication for the proteolytic cleavage of the C_L domain (152), which contributes to the stability of the LC and might protect it from fibrillation. In addition, since the C_L domain is not involved in the deposits, it seems that structural elements exist that protect it from misfolding (209), whereas some factors within the V_L domain might predispose it to the disease state. As V_Ls with different C-terminal lengths have been reported to show varying aggregation kinetics, it seems that some residues at the linker connecting the V_L and the C_L domain of the LC (V_L/C_L boundary residues) might be important in this respect. In the study of Goto and Hamagushi, (1987) (213), they showed that the truncation of boundary residues (N-terminus) of C_Lλ results in a decrease stability and a higher unfolding rate for

the domain. However, these effects could not be assigned to a specific residue. In the context of the V_L, the effects of these residues at the C-terminus of the domain have not been described. λLC and κLC belong to two different families and despites similarities in their basic domain architecture, the properties responsible for the preservation of their structural integrity might be different. For example, λ occurs more frequently in AL disease than κ (146). Identifying features of the connecting-loop that are common to the two families will help understand why mostly the V_L appears in amyloid deposits. In addition, studying the effects of these connecting-loop residues in the context of the κLC might lead to the identification of common features that help maintain the structural integrity of antibody domains. Therefore within this thesis, through the generation of various truncation variants, the effects of specific $V_L\kappa$/$C_L\kappa$-linker residues on the MAK33 $V_L\kappa$ and the $C_L\kappa$ domain will be analyzed in terms of stability, folding and fibril formation.

CHAPTER 2 – Experimental section

2 Materials and methods

2.1 Materials

2.1.1 Equipment and Consumables

Equipment name and type	Supplier
Autoclave Varioclav EP-Z	H+P
Cell Disruption Apparatus Basic Z	Constant Systems
Membrane vacuum pump	Sartorius
pH meter	WTW
Power amplifiers EPS 3500, 3501 and 1001	GE Healthcare
MARKII (refractometer)	Leica
Thermal cycler Primus 25	MWG
Thermal cycler MJ Mini 48 well	Biorad
Ultra filtration cell 8050	Amicon
Vortex MS2	IKA
Water bath F6-K	Haake
Ultraflex II MALDI ToF/ToF	Bruker Daltonics
SX18-MVstopped-flow apparatus with pdp spectra kinetic monochromator and fluorescence detector	Applied photophysics
Centrifuges	
Avanti J25 and J26 XP	Beckman Coulter
Optima XL-A (equipped with FDS)	Beckman Coulter (Aviv)
Optima XL-I	Beckman Coulter
Rotina 46R	Hettich
Rotina 420R	Hettich
Universal 320R	Hettich
Tabletop centrifuge 5418	Eppendorf
Tabletop centrifuge Mikro R200	Hettich
Chromatography systems	
ÄKTA Prime	GE Healthcare
FPLC + REC 112	GE Amersham
Frac-900/950 fraction collectors	GE Healthcare
Superloops (various volumes)	GE Healthcare

Circular dichroism spectropolarimeters

J710 (with PFD-350S Peltier device)	Jasco
J715 (with PTC 348 WI Peltier device)	Jasco

Thermoblocks

Digital heat block	VWR
Eppendorf-Thermomixer	Eppendorf
TB1 Thermoblock	Biometra

Spectrofluorimeters

Jasco Spectrofluorimeter	Jasco
FluoroMax-4	Horiba Jobin Yvon

Gel documentation systems

Biodoc II	Biometra
ImageQuant 300	GE Healthcare
Gel electrophoresis and blotting devices	Hoefer
Homogeniser Ultra Turrax DIAX900	Heidolph
HPLC systems	Jasco
Ice maker	Zieger
Incubator	New Brunswick Scientific

Magnetic stirrers

MR2000	Heidolph
MR3001	Heidolph
MR80	Heidolph

Microscopes

Atomic Force Microscope (AFM)	Veeco
Transmission electron microscope JEM100CX	Joel

Scale balance

BP 121 S	Sartorius
BL 310	Sartorius

UV-Vis spectrophotometers

Helios γ	Thermo Fisher
UltroSpec 1100 pro	Amersham Biosciences
Nanodrop	Peqlab

Other equipment used not mentioned directly by name and supplier were glassware like beakers and bottles of different dimensions.

Consumables, name and type	Supplier
Amicon Ultra-15 Centrifugal Filter Units	Millipore
Amicon Ultra-4 Centrifugal Filter Units	Millipore
Cuvettes, plastic, 1 mL	Brand

Dialysis membranes Spectra/Por (various MWCOs)	Spectrum Laboratories
Immobilon-P membrane (PVDF)	Roth
Membrane discs	Sartorius
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
Zip Tip	Millipore

2.1.2   Chemicals

Chemicals	Supplier
$^{15}NH_4Cl$ (Ammonium chloride, nitrogen 15 isotope)	Cambridge Isotope Laboratories
13C-glucose	Cambridge Isotope Laboratories
2-Mercaptoethanol	Sigma
8-Anilino-1-naphthalenesulfonic acid (ANS)	Sigma
Acetic acid	Roth
Acrylamid/Bisacrylamide solution 38:2 (40% w:v)	Serva Electrophoresis
Agar Agar	Serva Electrophoresis
Agarose	Serva Electrophoresis
Ammonium persulfate (APS)($(NH_4)_2S_2O_8$)	Roth
Ampicillin sodium salt	Roth
Bacto-pepton	BD Biosciences
Bacto-trypton	BD Biosciences
Bromphenol blue	Serva Electrophoresis
Coomassie Blue R	Serva Electrophoresis
Coomassie Brilliant Blue R-250	Serva Electrophoresis
Deoxynucleoside triphosphates (dNTPs)	Roche
Dimethyl sulfoxide (DMSO) (C_2H_6OS)	Sigma
Dithiothreitol (DTT)	Roth
EDTA	Merck
Ethanol	Merck
Ethidium bromide ($C_{21}H_{20}BrN_3$)	Sigma
Glucose	Merck
Glutathione, oxidized (GSSG)	Sigma
Glutathione, reduced (GSH)	Sigma
Glycerol ($C_3H_8O_3$)	Roth

Glycine	Roth
Guanidinium chloride (GdmCl)	Sigma
Hydrochloric acid (HCl) 32%	Merck
Isopropyl β-d-1-thiogalaktopyranoside (IPTG)	Serva Electrophoresis
Isopropanol	Sigma
Kanamycin sulphate	Roth
KCl	Carl Roth
KH_2PO_4	Merck
L-Arginine	Sigma-Aldrich
LB medium	Serva Electrophoresis
$Na_2HPO_4.2H_2O$	Merck
$NaH_2PO4.H2O$	Merck
Sodium acetate ($C_2H_3NaO_2.3H_2O$)	Merck
Phosphuric acid (H_3PO_4)	Roth
Protease inhibitor Mix G	Serva Electrophoresis
Sodium chloride (NaCl)	Merck
Sodium azide (NaN_3)	Sigma
Sodium dodecylsulphate (SDS)	Serva Electrophoresis
Tetraethylethylendiamin (TEMED)	Roth
Thoflavin T (ThT)	Sigma
Tris ($C_4H_{11}NO_3$)	Roth
Triton X-100	Merck
Urea	Merck

Deionised Water (H_2O) was produced by Millipore Direct-Q5 system.

2.1.3   Buffers and Solutions

Type	Composition
Protein Purification	
Inclusion Body (IB) preparation (resuspension) buffer	50 mM Tris / HCl pH 7.5
	5 mM EDTA
	5 mM NaCl
	Add Protease Inhibitor before use
Low salt buffer	25 mM Tris/HCl pH 7.8-8.0
	5 mM EDTA
	5 M Urea
High salt buffer	25 mM Tris/HCl pH 7.8-8.0
	5 mM EDTA
	5 M Urea
	1 M NaCl

IB dissolving buffer	25 mM Tris/HCl pH 7.8-8.0
	5 mM EDTA
	5 mM EDTA
	8 M Urea
	0.1% (v/v) β-mercaptoethanol
Refolding buffer V_L, C_L, β2m	250 mM Tris/HCl pH 8.0
	5 mM EDTA
	100 mM L-Arginine
	0.5 mM GSH
	1 mM GSSG
Refolding buffer VH	250 mM Tris/HCl pH 8.0
	5 mM EDTA
	400 mM L-Arginine
	0.5 mM GSH
	1 mM GSSG
Phosphate buffered saline (1x PBS)	137 mM NaCl
	2.7 mM KCl
	10 mM $Na_2HPO_4.2H_20$
	1.76 mM KH_2PO_4
	pH 7.4

Buffers for SDS-PAGE

Fairbanks A (Coomassie staining)	25% (v/v) 2-Propanol
	10% (v/v) Acetic acid
	0.05% (w/v) Coomassie Blue R
Fairbanks D (destaining solution)	10% (v/v) Acetic acid
Laemmli sample buffer (5x)	0.3 M Tris/HCl, pH 6.8
	10% (w/v) SDS
	50% (v/v) Glycerol
	5% (v/v) β-Mercaptoethanol
	0.05% (w/v) Bromophenol blue
SDS running buffer (10x)	0.25 M Tris/HCl, pH 8
	2 M Glycine
	1% (w/v) SDS
Separation gel buffer (4x)	1.5 M Tris/HCl, pH8.8
	0.8% (w/v) SDS
Stacking gel buffer (2x)	0.25 M Tris/HCl, pH 6.8
	0.4% (w/v) SDS

SDS-PAGE solutions

50 ml 18 % separation gel	22.5 ml of Acrylamid/Bisacrylamide solution 38:2 (40% w:v)

	12.5 ml Separation gel buffer (4x)
	15 ml ddH$_2$O
50 ml 5% stacking gel	6.25 ml of Acrylamid/Bisacrylamide solution 38:2 (40% w:v)
	25.0 ml Stacking gel buffer (2x)
	18.75 ml ddH$_2$O
10% APS	10 mg APS, fill up to 100 µl with ddH$_2$O (use in a 1:100 dilution)
Temed	Use in a 1:1000 dilution

Buffers for molecular biology

50x TAE buffer	Tris/Acetate pH 8.0
	50 mM EDTA
Loading buffer (10x)	50% (v/v) Glycerol
	10 mM Tris/HCl pH 8.0
	0.2% (w/v) Orange G
	0.2% (w/v) Xylencyanol

Buffers for amyloid induction assay

Buffer at neutral pH	1x PBS, pH 7.4
	0.05% NaN$_3$
Buffer at neutral pH with SDS	1x PBS
	0.5 mM SDS
	0.05% NaN$_3$
	pH 7.4
Buffer with pH 2.5	Equal volumes of 1x PBS pH 7.4 and sodium acetate/sodium phosphate buffer pH2 (25 mM NaAc 25 mM NaH$_2$P0$_4$.H$_2$O adjust pH with H$_3$PO$_4$
ThT stock solution	500 µM ThT in 20 mM Tris/HCl pH 7.4

2.1.4   Chromatography resins and Columns

Type	Supplier
Superdex 75 Prep Grade	GE Healthcare
Q Sepharose Fast Flow	GE Healthcare
SP Sepharose Fast Flow	GE Healthcare

2.1.5   Enzymes, Standards and Commercial kits

Kit name	Supplier
Pfu DNA polymerase	Promega

23

PWO polymerase	Roche
GoTaq polymerase	Promega
Restriction enzymes	NEB and Promega
Dpn1	NEB
T4 DNA Ligase	Promega
peqGold 1 kb ladder Orange G	Peqlab
SDS-PAGE Standard Low Range	Biorad
Wizard Plus SV Minipreps DNA Purification System Protocol	Promega
Wizard SV Gel and PCR clean-up system	Promega

2.1.6   Bacteria strains and Plasmids

Strain	Genotype	origin
E. coli BL21 star (DE3)	F-*ompT hsd*SB(rB-mB-) *gal dcm rne131* (DE3)	Invitrogen
E. coli BL21 codon plus (DE3)-RIL	F- ompT hsdS(rB- mB-) dcm+ Tetr gal λ endA Hte [argU ileY leuW Camr]	Stratagene
E. coli K12 *JM109*	e14-(McrA-) recA1 endA1 gyrA96 thi-1 hsdR17(rK- mK+) supE44 relA1 Δ(lac-proAB) [F′ traD36 proAB lacI^{q}ZΔM15]	Promega
E. coli Mach1	ΔrecA1398 endA1 tonA Φ80ΔlacM15 ΔlacX74 hsdR (rk- mk+)	Invitrogen

All constructs for expression in *Escherichia coli* were cloned into pET28-b (Novagen).

2.1.7   Media for expression in Bacteria

Media	Composition
Luria Bertani (LB$_0$)	20 g/l LB medium
	15 g/l Agar Agar (to LB medium for agar plates)
Solutions for chemical competent cells	1 M MgCl$_2$ (used MgCl$_2$.6H$_2$O
	Solution A (1000 ml)
	3 M NaoAc (pH 5.5); 13 ml
	1 M CaCl2; 100 ml
	2.8 mM MnCl2; 25 ml
	Add 862 ml ddH$_2$O
	Solution A + Glycerine (87%) (400 ml)
	Glycerine (87%); 69 ml
	Solution A; 331 ml
Super Optimal Broth (SOB)	0.5% Yeast extract
	2% Tryptone
	10 mM NaCl

	2.5 mM KCl
	10 Mm MgCl2
	10 mM MgSO4
Terrific Broth (TB) solution (for	10 mM HEPES
supercompetent cells)	15 mM CaCl2
	250 mM KCl
	pH 6.7 with KOH or HCl
	55 mM MnCl2
M9	6 g/l Na2HPO4
	3 g/l KH2PO4
	0.5 g/l NaCl
	pH 7.5 (NaOH)
	add before use:
	2 g/l D-(+)-Glucose (^{13}C or ^{12}C) (20 %); 10 ml/l
	0.5 g/L ^{15}NH4Cl (10 %); 5 ml/l
	1 M MgSO4; 1 ml/l
	1 M CaCl2; 0.3 ml/l
	1 mg/ml Thiamine HCl; 1.5 ml/l
	0.1 mg/ml Biotin; 15 ml/l
	10 ml trace elements (100x)
Antibiotics (1000x stocks)	200 µg/l (in ddH2O) Ampicillin
	50 µg/l (in ddH2O) Kanamycin

1l of 100x trace elements contain: 5 g EDTA, 0.83 g $FeCl_2.6H_2O$, 84 mg $ZnCl_2$, 13 mg $CuCl_2.2H_2O$, 10 mg $CoCl_2.6H2O$, 10 mg H_3BO_3, and 1.6 mg $MnCl_2.4H2O$. pH 7.5 with NaOH. Solution was sterile filtered before use. Both LB, M9 media and solutions for competent cell preparation were sterilized using an autoclave at 121°C for 20 min. Antibiotic stocks, 1 mg/ml; Thiamine HCl, 0.1 mg/ml biotin, labeled ^{13}C-glucose (20 %) and ^{15}NH$_4$Cl (10 %) were passed through a sterile filter (0.22 µm). Antibiotic were stored at -20°C.

2.1.8   Oligonucleotides for quick change PCR

Template	name	5'-3' – sequence
1OPG V$_L$κ	E2I_fw	aggagatataccatggacatactgctgacccagagtcc
	E2I_rv	ggactctgggtcagcagtatgtccatggtatatctcct
	E2A_fw	gagatataccatggacgcactgctgacccagagtc
	E2A_rv	gactctgggtcagcagtgcgtccatggtatatctc
	E2L_fw	aggagatataccatggacttactgctgacccagagtcc
	E2L_rv	ggactctgggtcagcagtaagtccatggtatatctcct
	E2Q_fw	ggagatataccatggaccagctgctgacccagagtcc
	E2Q_rv	ggactctgggtcagcagctggtccatggtatatctcc
	E2D_fw	gatataccatggacgatctgctgacccagagtc
	E2D_rv	gactctgggtcagcagatcgtccatggtatatc
	E2K_fw	ggagatataccatggacaaactgctgacccagagt
	E2K_rv	actctgggtcagcagtttgtccatggtatatctcc
	L3V_fw	aggagatataccatggacgaagtgctgacccag
	L3V_rv	ctgggtcagcacttcgtccatggtatatctcct
	G100A_fw	cctctgacctttggtgctggtagcaaactggaa
	G100A_rv	ttccagtttgctaccagcaccaaaggtcagagg
	S102_fw	ctctgacctttggtggtggtaccaaactggaaattaaataata
	S102_rv	tattatttaatttccagtttggtaccaccaccaaaggtcagag
	I106L_fw	ggtggtggtagcaaactggaacttaaataataaaagcttgcgg
	I106L_rv	ccgcaagcttttattatttaagttccagtttgctaccaccacc
1OPG V$_L$-E2I	E2I/L3V_(IV)_fw	gaaggagatataccatggacatagtgctgacccaga
	E2I/L3V_(IV)_rv	tctgggtcagcactatgtccatggtatatctccttc
	E2I/G100A_(IA)_fw	cctctgacctttggtgctggtagcaaactggaa
	E2I/G100A_(IA)_rv	ttccagtttgctaccagcaccaaaggtcagagg
	E2I/S102T_(IT)_fw	ctctgacctttggtggtggtaccaaactggaaattaaataata
	E2I/S102T_(IT)_rv	tattatttaatttccagtttggtaccaccaccaaaggtcagag
	E2I/I106L_(IL)_fw	ggtggtggtagcaaactggaacttaaataataaaagcttgcgg
	E2I/I106L_(IL)_rv	ccgcaagcttttattatttaagttccagtttgctaccaccacc
1OPG V$_L$-IA	E2I/G100A/S102T _(IAT)_fw	ctctgacctttggtgctggtaccaaactggaaattaaataata
	E2I/G100A/S102T _(IAT)_rv	tattatttaatttccagtttggtaccagcaccaaaggtcagag
MAK33 V$_L$κ	MK-V$_L$ΔR108_fw	ccaagctggagctgaaataatagaagcttgcggc
	MK-V$_L$ΔR108_rv	gccgcaagcttctattatttcagctccagcttgg
	I2E_fw	ctttaagaaggagatataccatggatgaggtgctaactcagtctccagc
	I2E_rv	gctggagactgagttagcacctcatccatggtatatctccttcttaaag
	I2A_fw	ctttaagaaggagatataccatggatgctgtgctaactcagtctcca
	I2A_rv	tggagactgagttagcacagcatccatggtatatctccttcttaaag
	I2L_fw	ttaagaaggagatataccatggatcttgtgctaactcagtctc
	I2L_rv	gagactgagttagcacaagatccatggtatatctccttcttaa

	I2Q_fw	ctttaagaaggagatataccatggatcaggtgctaactcagtctccagc
	I2Q_rv	gctggagactgagttagcacctgatccatggtatatctccttcttaaag
	I2D_fw	ctttaagaaggagatataccatggatgatgtgctaactcagtctcca
	I2D_rv	tggagactgagttagcacatcatccatggtatatctccttcttaaag
	I2K_fw	ttaagaaggagatataccatggataaggtgctaactcagtctccag
	I2K_rv	ctggagactgagttagcaccttatccatggtatatctccttcttaa
	D17A_fw	tctgtgactccaggagctagcgtcagtctttcc
	D17A_rv	ggaaagactgacgctagctcctggagtcacaga
	S20N_fw	ctccaggagatagcgtcaatctttcctgcagg
	S20N_rv	cctgcaggaaagattgacgctatctcctggag
	N31D_fw	gccagccaaagtattagcgacaacctacactggtatc
	N31D_rv	gataccagtgtaggttgtcgctaatactttggctggc
	D70N_fw	gcagtggatcagggacaaatttcactctcagtatc
	D70N_rv	gatactgagagtgaaatttgtccctgatccactgc
MK-V$_L$ΔR108	I2EΔR108_fw	ctttaagaaggagatataccatggatgaggtgctaactcagtctccagc
	I2EΔR108_rv	gctggagactgagttagcacctcatccatggtatatctccttcttaaag
	T97VΔR108_fw	aacagctggcctctcgtgttcggtgctgggac
	T97VΔR108_rv	gtcccagcaccgaacacgagaggccagctgtt
	T102SΔR108_fw	ttcggtgctgggagcaagctggagctg
	T102ΔR108_rv	cagctccagcttgctcccagcaccgaa
	MK-VLΔD1_fw	taagaaggagatataccatgattgtgctaactcagtctcc
	MK-VLΔD1_rv	ggagactgagttagcacaatcatggtatatctccttctta
	MK-VLΔD1/I2_fw	gaaggagatataccatggtgctaactcagtctcc
	MK-VLΔD1/I2_rv	ggagactgagttagcaccatggtatatctccttc
MK-T102SΔR108	I2E/T102SΔR108_fw	ctttaagaaggagatataccatggatgaggtgctaactcagtctccagc
	I2E/T102ΔR108_rv	gctggagactgagttagcacctcatccatggtatatctccttcttaaag
1VGE V$_L$κ	L2E_fw	aagaaggagatataccatggaagaggttatgacccagagc
	L2E_rv	gctctgggtcataacctcttccatggtatatctccttctt
1AQK V$_L$λ	N2E_fw	aggagatataccatggaagaggttctgacccagcctccg
	N2E_rv	cggaggctgggtcagaacctcttccatggtatatctcct
	N2A_fw	agaaggagatataccatggaagctgttctgacccagcctc
	N2A_rv	gaggctgggtcagaacagcttccatggtatatctccttct
	N2S_fw	ctttaagaaggagatataccatggaaagtgttctgacccagc
	N2S_rv	gctgggtcagaacactttccatggtatatctccttcttaaag
1VGE V$_L$λ	V2E_fw	agatataccatggcacaggagaaactgctggaacagagcg
	V2E_rv	cgctctgttccagcagtttctcctgtgccatggtatatct
	V2A_fw	gatataccatggcacaggctaaactgctggaacagag
	V2A_rv	ctctgttccagcagtttagcctgtgccatggtatatc
MK-C$_L$κ*	C$_L$k^{MGRAD}_fw	TTTTCCATGGGTCGAGCAGA**C**GCAGCACCCACTGTATCCATC
	C$_L$k^{MAD}_(ΔR108)_fw	TTTTCCATGGCAGACGCAGCACCCACTGTATCCATC
	C$_L$k$^{MGRAD/MAD}$_rv	TTTTAAGCTTCTATTATTCATTCCTGTTGAA
MK_C$_L$k^{MGRAD}	C$_L$k^{MRAD} (wt)_fw	Aaggagatataccatgcgagcagacgcagcac

	C$_L$k^{MRAD} (wt)_rv	`gtgctgcgtctgctcgcatggtatatctcctt`
C$_L$k^{MAD}_(ΔR108)	C$_L$k^{MD}(ΔR108/Ala109)fw	`aaggagatataccatggacgcagcacccactg`
	C$_L$k$^{MD(}$(ΔR108/Ala109)rv	`cagtgggtgctgcgtccatggtatatctcctt`

Variants were all cloned in the pET28b vector. All QuickChange mutagenesis primers were designed with the help of Agilent Technologies QuickChange primer design tool https://www.genomics.agilent.com/primerDesignProgram.jsp. To ease primer design of variants carrying mutations at the N-terminus and/or C-terminus, vector backbone nucleotide sequences were also included. 33 nucleotides of the vector backbone preceding the sequence of interest were submitted for generation of N-terminal variant primers, whereas for the C-terminal variants, 36 nucleotides of the vector backbone immediately after the sequence of interest were considered. Primers were purchased from Eurofins MWG Operon. The MK-C$_L$κ* template contains residues A111-E214 of the MAK33 κ light chain and four additional N-terminal residues (GlySerHisMet) derived from the pET28a(+) vector after Thrombin cleavage of the His-tag, as used in the study of Feige et al., (2007)(56). That is, this template lacks the CLκ N-terminal conserved consensus start sequence Arg108/Ala109/Asp10, which is introduced in this study by sub cloning, while deleting the additional four residues.

2.2   Software, databases and web-based tools

Name	Provider
Software	
Adobe Illustrator CS4	Adobe Inc.
Adobe Photoshop CS4	Adobe Inc.
BioEdit	Ibis Biosciences
DynaFit	Biokin Ltd.
Endnote X7	Thomson Scientific
Microsoft Office	Microsoft
Mmass V2.4.0	Open Source
OriginPro 8.6	OriginLab Corp.
Pymol V1.3	Schrödinger
SedFit	Peter Schuck
Databases	**URL**
Abysis	http://www.bioinf.org.uk/abs/
PDB	www.rcsb.org/pdb
IMGT	http://www.imgt.org
NCBI/ PubMed	www.ncbi.nlm.nih.gov/pubmed
UniProtKB/Swiss Prot/Expasy	http://web.expasy.org/docs/swiss-prot_guideline.html
Web-based Tools	
Agilent Technologies	https://www.genomics.agilent.com/primerDesignProgram.jsp

QuickChange

ClustalW Multiple Sequence Alignement	http://www.genome.jp/tools/clustalw/
Denaturant Calculator	http://sosnick.uchicago.edu/gdmcl.html
Kabat Numbering	http://www.bioinf.org.uk/abs/
NCBI IgBlast	http://www.ncbi.nlm.nih.gov/igblast/
ExPasy ProtParam	http://web.expasy.org/protparam/
EXPasy Translation tool	http://web.expasy.org/translate/
T_m calculation for oligo	http://www.promega.com/techserv/tools/biomath/calc11.htm
Reverse Complement	http://www.bioinformatics.org/sms/rev_comp.html or http://bioinfx.net/

All trademarks are a property of their respective holder.

2.3   Molecular biology methods

All solutions were autoclaved or sterile filtered. If not otherwise stated, all experiments were performed at room temperature. Polymerase chain reaction (PCR), DNA electrophoreses, restriction digests, transformation and cultivation of *E. coli* were conducted according to standard protocols described elsewhere, see (214)   The most frequently used methods are described in the following sections.

2.3.1   Polymerase Chain Reaction (PCR)

Site-directed mutagenesis and target gene amplification were done by PCR. *Pfu* polymerase (Promega), PWO polymerase (Roche), or GoTaq polymerase (Promega) was used depending on the template. Primers for site-directed mutagenesis were designed with the help of Agilent QuickChange mutagenesis primer design tool while primers for target gene amplification and subcloning were designed by hand and their melting temperatures (T_m for oligos) control with T_m calculation for oligos tool. Primers were ordered from Eurofins MWG Operon. For *Pfu* Polymerase, approximately two min/1000 bp are recommended, for GoTaq polymerase 1 kbp strand can be replicated in less than

10 s.

The following pipetting scheme was used for QuickChange PCR reactions.

5 µl 5x *Pfu*-Polymerase buffer
2 µl dNTPs (10 mM each dNTP)
1 µl forward primer (10 pmol/µl)
1 µl reverse primer (10 pmol/µl)
1 µl template DNA (100 ng/µl)
1 µl *Pfu*-Polymerase
39 µl nuclease free ddH2O
Σ = 50 µl

Primers stocks were diluted 1:10 to yield ~ 10 pmol/µl used for the PCR.

Table 2.1 below depicts the cycler program for QuickChange site-directed mutagenesis.

Table 2.1. PCR program for QuickChange site-directed mutagenesis

Temperature	Duration	process
94 °C	2 min	Activation of Polymerase
94 °C	30 s	Denaturation of DNA
53 °C	1 min	Annealing of primers
72 °C	10 min	Elongation
4 °C	∞	hold

(18 cycles for the denaturation, annealing and elongation steps)

The following pipetting scheme was used for target gene amplification.

20 µl 5x GoTaq Polymerase Buffer
2 µl dNTPs (10 mM each dNTP)
1 µl forward primer (100 pmol/µl)
1 µl reverse primer (100 pmol/µl)
2 µl template DNA (100 ng/µl)
1 µl GoTaq Polymerase
73 µL nuclease free ddH2O
Σ = 100 µl

The cycler program for target gene amplification was as follows:

Table 2.2. PCR program for amplification of target gene

Temperature	Duration	process
95 °C	2 min	Activation of Polymerase
95 °C	30 s	Denaturation of DNA
53 °C	1 min	Annealing of primers
72 °C	1 min	Elongation
72 °C	10 min	Final elongation
4 °C	∞	hold

(35 cycles for the denaturation, annealing and elongation steps)

2.3.2 Agarose gel electrophoresis

This allowed isolation and purification of PCR products and digested fragments by gel extraction. Depending on the fragment size, agrose gels of between 0.7 and 1.5 % agarose in TAE buffer were prepared. For size determination, a 1 kb DNA ladder (Peqlab) was loaded alongside. The gels were run at 120 mV voltage for 30 min. For visualization under UV light, the DNA was stained with either ethidium bromide or Stain G (Serva Electrophoresis).

2.3.3 Preparation of chemical competent and supercompetent *E. coli* cells

Transfomation-competent *E. coli* were used to propagate the recombinant expression vectors and for overexpression of antibody domains as inclusion bodies. Chemically competent bacteria were prepared by growing one fresh colony of *E. coli* first in a pre-culture of 4 ml LB medium at 37 °C

overnight. 2 ml of the overnight pre-culture was then transferred to 100 ml LB medium at 37 °C under agitation and left to grow to an OD_{600} of 0.5 – 1 (~ 2 h), were most bacteria are still in the growth phase. 2 ml of 1M $MgCl_2$ was added to the 100 ml culture, mixed and incubated for another 10 min at 37 °C while shaking before incubating on ice for 60 min. After centrifugation for 5 min at 4500 rpm/4 °C, the pellet was resuspended in 20 ml of ice-cold solution A and then incubated on ice again for 60 min before an additional round of centrifugation (as before) was performed. The bacteria pellets were now resuspended in 2 ml of iced-cold solution A + glycerol (87%).

For preparation of supercompetent cells, 10-12 large colonies were transferred from a fresh overnight culture plate into 250 ml SOB medium and incubated at 19 °C with vigorous shaking to an OD_{600} of 0.5 – 1 (~ 24 h). After a cold shock of 10 minutes on ice, cells were spun down at 4500 rpm for 5 min at 4°C. Cells were gently resuspended in 80 ml ice-cold TB solution and incubated on ice for 10 min before a last round of centrifugation as before. The pellet was finally resuspended in 20 ml ice-cold TB solution containing 1.4 ml DMSO. 80 µl aliquots were immediately snap-frozen in liquid nitrogen and stored at -80 °C.

2.3.4   Cultivation of *E. coli* cells

Bacteria were cultured on LB-agar plates or in LB-broth at 37 °C. Selection was by means of antibiotics, added depending on the type of antibiotic resistance gene encoded by the transformed plasmid. Liquid cultures were inoculated 1:40 from fresh overnight cultures or with a single colony from plates. Growth was controlled by visible spectroscopy at 600 nm. For cryo-stocks 1 ml of fresh overnight culture was mixed with 500 µl undiluted glycerol, snap-frozen and stored at -80 °C.

2.3.5   Cloning strategies

2.3.5.1 QuickChange site-directed mutagenesis

Single, double and triple point substitutions, variants with deletion or insertion of single residues were generated with QuickChange mutagenesis PCR with forward and reverse primers carrying the target mutation. Inserts of the desired template in pET28b vector were used as PCR templates for the generation of the respective variants. The PCR was followed by digestion of template DNA with DpnI enzyme (NEB) for 1 h at 37 °C. The PCR product was then transformed through a heat shock dependent transformation into competent *E. coli* Mach1 cells. Positive clones were selected on kanamycin LB plates overnight at 37 °C. Plasmid DNA from a single colony was isolated with the Wizard Plus SV Miniprep kit according to the manufacturer's instructions and sequenced using the T7 forward or pET-RP sequencing primer by Eurofins MWG Operon to verify the desired mutation.

2.3.5.2 Standard PCR and subcloning

Standard PCR was used for the introduction of new restriction sites and sequence information at the 5`or 3`ends. Primers carrying the sequence of interest and the desired restriction sites (mostly N-term Nco*I* and C-term Hind*III*) were designed based on the target gene. The PCR was performed with the

target gene (e.g. MK-C_Lκ*) as template. A negative control PCR without the template was also included. The formation of amplicon was controlled by agarose gel electrophoresis and the correct product was excised and purified with Wizard DNA gel purification kit as instructed by the manufacturer. The insert and vector both carrying the same restriction sites were digested with the appropriate restriction enzymes for at least 3 h at 37 °C and then purified as before. By means of T4 DNA ligase, the clean insert and vector were ligated in a ratio of 2:1 or 3:1 in a total volume of 10 μl, overnight at 4°C or at room temperature for 25 min (Quick Ligation). The whole ligation reaction was transformed into competent Mach 1 *E. coli* cells and positive colonies were selected on kanamycin LB plates. Sequences were verified by sequencing. This approach was successful in the generation of different N-terminal C_Lκ variants lacking the N-term 6His-tag.

2.4   Protein chemistry methods

2.4.1   Recombinant protein expression and purification

All V_L, V_H and C_L variants were expressed and purified as already described in (74,193). In brief, the plasmid was transformed into *E. coli* BL21 (DE3)-star cells (for V_L variants) or in *E. coli* JM109 cells (for V_H and C_L variants) as starter culture in 100 ml at 37 °C ahead of expression.

Overexpression was performed in 2 l LB broth with 35 μg/ml kanamycin in a 5 l cornical flask at 37 °C. This was inoculated with a 1:40 diluted started culture and at an OD_{600} of 0.6–0.8, overexpression was induced using 1 mM isopropyl-β-d thiogalactopyranoside (IPTG, final concentration). Cells were harvested after overnight growth, and inclusion bodies were prepared as previously described (56,215). Briefly, harvested cells were homogenized and disrupted in 150 μl IB resuspension buffer containing Protease inhibitor Mix G (SERVA). After addition of the endonuclease DNAse I, the lysate was treated with 1% of detergent Triton X-100 and incubated at 4°C for at least 1 h for solubilization of lipids and cell membranes. Inclusion bodies were harvested at 20.000 rpm, 20 min (JA 25.50 rotor, Beckman Coulter) and washed twice in IB resuspension buffer.

The IB pellet was unfolded and solubilized in IB dissolving buffer at room temperature for at least 2 h. After centrifugation at 20.000 rpm, 30 min, JA 25.50 rotor, the soluble fraction was then injected onto a Q-Sepharose FF (ion exchange) column equilibrated in low salt buffer. The flow-through and eluted fractions were both collected. The fraction (flow-through) containing the protein determined by SDS-PAGE were diluted two times before being refolded by dialysis into the appropriate refolding buffer at 4 °C overnight. Refolded proteins were concentrated to a volume of 10 ml.

To remove misfolded aggregates and remaining impurities, the protein was cleaned using a Superdex 75 gel-filtration column equilibrated in PBS buffer. The recovery and purity of intact protein was verified by SDS-PAGE and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI TOF MS).

For NMR experiments, expression of ^{15}N- and ^{13}C--labelled V_L and V_H variants was performed in *E. coli* BL21 codon plus (DE3)-RIL cells in M9 minimal medium containing $^{15}NH_4Cl$ and ^{13}C-glucose. Before inoculation of the main culture, the overnight starter-culture in LB was harvested and washed twice in PBS to remove any traces of LB medium. Purification was done the same as for unlabeled variants.

2.5   MALDI-TOF MS analysis

Purity check and molecular weight analysis of purified proteins was performed on a Bruker Ultraflex-2 MALDI-TOF/TOF mass spectrometer. To remove potassium ions from PBS buffer, a Zip Tip preparation was performed for all samples according to the manufacturer's instructions (Millipore). The HCCA matrix was dissolved in 0.1% TCA, 80% acetonitrile and 20% ddH$_2$O. The full-length proteins were spotted in the matrix solution and analyzed after the evaporation of the solvent. MS spectra were analyzed with the mMass V2.4.0 software.

2.6   Spectroscopy

2.6.1   UV-Vis spectroscopy

The presence of a chromophore (system of conjugated pi-bonds with excitable electrons) in a molecule allows the absorption of light with a specific wavelength depending on the chromophore. When a light photon in the ultraviolet (UV)-visible (Vis) range of the spectrum hits a chromophore in a molecule, it would be absorbed by exciting an electron from its ground state to an excited state. In proteins, aromatic amino acids (mainly tryptophanes and tyrosines and to a lesser extent phenylalanines and disulphide bonds) absorb in the near UV range (280, 275 and 257 nm, respectively). The indole groups of Trp are largely responsible for the UV absorbance. This makes possible the determination of protein concentration at 280 nm. The calculation of the protein concentration from UV absorbance was carried out according to the Lambert-Beer law (Equation 2 1)

$$A = \varepsilon \cdot c \cdot d \rightarrow c = \frac{A}{\varepsilon \cdot d}$$ [2.1]

Where; A = Absorbance, ε = molar extinction coefficient (M^{-1} cm^{-1}), c = molar protein concentration (M), d = path length (cm).

The theoretical molar extinction coefficients were determined using the Expasy ProtParam tool. All UV spectra were acquired with a Helios γ or Ultrospec 1100 UV/Vis spectrophotometer at room temperature and baseline corrected for buffer absorbance. Measurements were done in UV cuvette quartz cuvette of 1 cm pathlength. Alternatively, a nanodrop spectrometer was used. After a blank run with buffer, the protein's UV spectra were recorded with the pathlength defined to be 1 cm.

2.6.2   Circular Dichroism (CD) Spectroscopy

CD spectroscopy is an averaging technique which gives information about the conformational state of an entire ensemble of molecules in a given sample. CD spectroscopy measures differential absorption of circularly polarized light by chiral centers in a molecule as a function of wavelength. Proteins are chemically asymmetric due to the nature of the polypeptide chain and the presence of a chiral α-

carbon atom in constituent amino acids (except for glycine). The α-carbon is a stereocenter which together with the asymmetry of the polypeptide chain give proteins their optical properties. The optical activities of proteins are in two-fold: (1) the difference in refractive index and (2) the difference in the absorption of right- and left-handed circularly polarized light, circular dichroism which is measured by CD spectroscopy. Lambert-Beer law also applies to CD spectroscopy for both right- and left-handed circularly polarized light with each exhibiting a different extinction coefficient:

$$\Delta A(\lambda) = A_L(\lambda) - A_R(\lambda) = [\varepsilon_L(\lambda) - \varepsilon_R(\lambda)]lc = \Delta\varepsilon lc \qquad [2.2]$$

Equation 2.2: Absorption between left and right handed circularly polarized light
A = absorption, λ = wavelength, ε = extinction coefficient, l = path length, c = concentration, R and L subscripts describe left- and right-handed circular polarized light.

Far UV CD spectra (typically between 180 – 260 nm) are used to determine secondary structure of proteins due to the effect of the dihedral angle of peptide bonds on the absorption of the chiral center at the α-carbon atoms. Distinct signals appear predominantly due to the differential absorption of backbone amide groups in asymmetric secondary structures. Typical secondary structure motifs like α-helices or β-sheets are identified according to their characteristic shape. Alpha-helical structures exhibit a negative band at about 222 nm and a negative and positive couplet at about 208 nm and 190 nm. For β-sheet structures, the CD spectra have a negative band at about 216 nm and a positive band at about 198 nm (216). Random coil structures show a characteristic minimum at 200 nm.

On the other hand, near UV CD spectra (typically between 250 – 350 nm) provide information about the polypeptide tertiary structure and they reflect the contribution of aromatic side chains and disulphide bonds. Aromatic amino acids are intrinsically symmetric with planar chromophores. When the aromatic amino acids are mobile, the signal in the near UV range becomes zero. In an ordered structure, the environment of aromatic amino acids becomes asymmetric and a CD signal can be observed (217).

CD is measured as ellipticity (Θ) in degrees (often mdeg). The conversion of the measurement signal to mean residue ellipticity in deg.cm^2.dmol^{-1} is as follows:

$$\theta_{MRW} = \frac{\theta \cdot 100 \cdot M}{d \cdot c \cdot N_{aa}} \qquad [2.3]$$

**Equation 3: Determination of the mean residue ellipticity**
Θ$_{MRW}$ = mean residue ellipticity (deg cm^2 dmol^{-1}), Θ = ellipticity (deg), M = molecular mass (g/mol), d = path length (cm), c = concentration (M), N$_{aa}$= number of amino acid residues.

CD measurements were performed using a Jasco J-720 spectropolarimeter equipped with a Peltier element. Far-UV CD spectra were measured using 10 µM of protein in a 1-mm pathlength quartz cuvette between 260 and 200 nm. Near-UV CD was measured between 320 and 250 nm using 50 µM protein in a 1-mm quartz cuvette. Measurements were done at 20 °C and all spectra were accumulated 16 times and buffer corrected. Thermal transitions were recorded using 10 µM protein at 212 nm for V$_L$κ and 205 nm for V$_L$λ, VH and C$_L$κ variants with a heating and cooling rate of 20 °C /h.

Thermal transitions were fit to a Boltzmann function. Equilibrium un(re)folding transitions and folding kinetics for $C_L\kappa$ variants were monitored by measuring the ellipticity at 220 nm for 50 s.. In all cases, the reversibility of the GdmCl-induced unfolding transitions was confirmed. The data were fitted to a two-state model using a linear extrapolation method described previously (81,218). All measurements were done in PBS buffer for native samples.

2.6.3   Fluorescence spectroscopy

Luminescence is generally the emission of light from any substance, which occurs from electronically excited states. Depending on the nature of the excited state, luminescence can be divided into two categories: phosphorescence and fluorescence. A Jablonski diagram (Fig. 2.1) is usually used to describe both processes (219). Species undergoing these processes have a singlet ground electronic state (low energy state (S_0)) and excited electronic states (high energy state (S_1) and second excited state (S_2). At each of these electronic states are a number of vibrational energy levels at which a fluorophore can exist. When a photon (usually of higher energy – shorter wavelength) is absorbed by a fluorophore, it gets excited from its ground state to one of various vibrational levels in the excited electronic states (S_1 or S_2). Due to collision, molecules in the S_2 state or higher S_1 vibrational states rapidly relax ($\leq 10^{-12}$ s) to the lowest vibrational levels of S_1 in a process called internal conversion. From the S_1 state, molecules can undergo spin conversion to the first triplet state (T_1) in a process called intersystem crossing. The return of the molecule from the T_1 state to vibrations levels in the S_0 state results in emission (with lower energy – larger wavelength), termed phosphorescence (emission rate $10^{-3} - 10^0$ s^{-1}). On the other hand, the return of molecules from the S_1 state to one of various vibrational states in the S_0 level resulting in the emission of a photons of low energy (higher than for phosphorescence but lesser than the excitation energy) is termed fluorescence (emission rate 10^{-9} s^{-1}). The fluorescence intensity of a fluorophore depends on its absorption of the excitation wavelength (nm) and its quantum yield at the emission λ (nm).
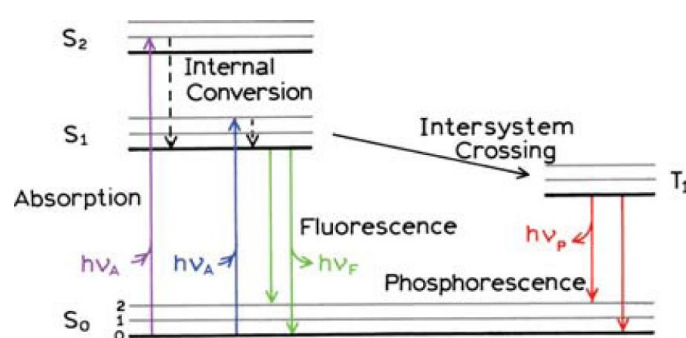
Figure 2.1. Simple representation of a Jablonski diagram. hν = E (energy of the photon), where, h is the Planck's constant and ν is the frequency (and ν = 1/λ, the wavelength). From. Lakowicz, (2006) (219).

Aromatic molecules are responsible for fluorescence. In proteins, intrinsic fluorescence originates from the aromatic amino acids, Trp, Tyr and Phe. The emission of Tyr in native proteins is often quenched due to the interaction with peptide chains or energy transfer (FRET) to Trp. Denaturation therefore results in Tyr emission and can be used in protein analysis due to its relative abundance. However,

the Pk$_a$ of Tyr decreases drastically upon excitation and ionization can occur. It is also less sensitive to the polarity of its environment. Phe emission only occurs when samples lack Tyr and Trp, a very rare case. Trp has the highest quantum yield (best emitter) and its emission maximum is highly sensitive to the polarity of its environment due to the presence of an indole ring (219,220). Thus it is often used as the intrinsic fluorescent probe in studying protein conformational changes (folding and dynamics). Emission spectra shift could be due to protein–protein association or protein unfolding. The fluorescence emission maxima of proteins reflect the average exposure of Trp residues to aqueous solvent. That is, depending on the solution's conditions, a protein might undergo conformational changes thereby exposing these hydrophobic residue (Trp) to the solvent. This then causes a variation in the relative fluorescent signal. However, it is usually not possible to predict how the fluorescence signal will change. While some proteins exhibit large increases in fluorescent intensities and lifetime, others show a decrease upon unfolding. Usually in a hydrophobic environment (buried within the core of the protein), Trp residues show a high quantum yield and therefore a high fluorescence intensity. If exposed to solvent, the quantum yield decreases resulting in a low fluorescence intensity. Trp residues show strong Stokes shifts which depend on the solvent, that is, the maximum wavelength of the emission band will differ depending on its environment. The peak maxima range from ~308 nm for relatively non polar environments to ~360 nm which represent the chromophore surrounded by polar solvent molecules (221). In multi Trp-containing proteins, the tryptophan emission spectrum is the result of the overlapping contribution of all these residues located in distinct regions of the protein experiencing different microenvironments. Trp fluorescence can be quenched by iodide, acrylamide and nearby disulfide bridges. All antibody domains possess at least one Trp residue conserved in the hydrophobic core. In the native state, the fluorescence of this Trp is quenched by a conserved disulfide bond in its close proximity (222,223). Upon unfolding, this Trp becomes exposed (farther away from the disulfide bond), thereby resulting in increased fluorescent signal. As a result, this Trp was used to study folding and dynamics of the different antibody domains.

Besides intrinsic fluorescent reporters there are extrinsic fluorophores such as 1-anilinonaphthalene-8-sulfonic acid (ANS) and Thioflavin T (ThT) (224). The latter, used to monitor the fibrillation process of proteins will be described elsewhere. Due to their sensitivity for hydrophobic surfaces, these two dyes can be used for structural studies of proteins. Upon binding to hydrophobic patches, the fluorescence of ANS increases with often a blue-shifted emission maximum. As a result, ANS is a versatile probe to study structural changes in proteins that result in a change in the hydrophobic environment such as folding and unfolding, whereby natively buried hydrophobic residues become surface exposed during unfolding. The dye is generally used to assess the hydrophobicity at protein surfaces (225) and for the detection of partially unfolded species or "molten globule" with usually large exposed hydrophobic patches (226) and therefore the characterization of intermediates during amyloid fibril formation (227). It should be noted that, the interaction with ANS is not always completely due to the hydrophobicity of the protein but sometimes also as a result of electrostatic interactions with the ANS' negatively charged sulfonyl group.

2.6.3.1   Tryptophan fluorescence, equilibrium unfolding and refolding transitions

Unless otherwise stated, all fluorescence measurements were carried out using a FluoroMax-4 Spectrofluorimeter. Trp fluorescence spectra measurements were performed with excitation and emission slit widths of 2 and 3 nm, respectively. The protein concentration in a 1-cm quartz cuvette was 1 µM. The sample was excited at 280 nm, and spectra were recorded between 300 and 450 nm. Equilibrium unfolding and refolding transitions were carried out by denaturing 1 µM protein overnight at 25 °C in different concentrations of GdmCl (between 0 and 2.5 or 3 M GdmCl) for unfolding transitions or by diluting a 3 M GdmCl-denatured protein sample to various concerntrations of GdmCl with PBS for refolding transitions (final protein concerntration 1 µM). Conformational changes were monitored by recording Trp fluorescence emission at 358 nm with excitation at 280 nm for 50 s, and the average taken. 358 nm was the wavelength with the maximum Trp fluorescence signal difference between the native and fully-denatured sample spectra. All measurements were performed at 25 °C. The data was evaluated using Origin (originLab) and the analysis was carried out using the linear extrapolation method assuming a two-state model described previously (Eq. 2.4) (81,218)

$$y(D) \; = \; y_N^0 + \; m_N \cdot [D] - \frac{(y_N^0 + m_N \cdot [D]) - (y_U^0 + m_U \cdot T)}{1 + exp\left(- \frac{\Delta G_{stab} + m_c \cdot [D]}{R\,T}\right)} \qquad [2.4]$$

Equation 2.4: Two-state model with the assumption of linear dependency of emission from native and unfolded protein. *y(D) = fluorescence signal; y^0 = y-intercepts; N = data of native protein; U = data of unfolded protein; m = slope of the best fit straight line; m_c = cooperativity; [D] = denaturant concentration; T = temperature in Kelvin (20 °C $\sim$ 293.1 K); ΔG_{stab} = free enthalpy of denaturant induced unfolding; R = universal gas constant (8.314 J mol^{-1} K^{-1}).*

2.6.3.2   Folding kinetic experiments by manual mixing

For refolding kinetics, 25 or 100 µM of protein was denatured in 2.5 M GdmCl overnight at 25 °C and then diluted into a refolding buffer (PBS) in a 1 cm path-length stirred cuvette. For unfolding experiments, native protein equilibrated overnight at 25 °C was diluted 25-fold or 100-fold into an unfolding buffer. The final protein concentration was 1 µM. The change in fluorescence intensity over time was monitored at the unfolded emission maximum of 358 nm, after excitation at 280 nm with excitation and emission slits set to 1 nm and 2 nm, respectively. At least three traces were averaged.

2.6.3.3   Folding kinetic with stopped-flow fluorescence spectroscopy

Stopped-flow is a spectroscopic technique used for studying fast reaction mechanisms in solution that occur within a few ms and up to a few 100 s. In general, two reagents are rapidly mixed together and then 'stopped' in an observation cell. How the reaction proceeds can be visualized by the change in the recorded signal as a function of time observed when the sample cell is irradiated with monochromatic light. Depending on whether the stopped-flow instrument is coupled to a fluorescence spectrometer or circular dichroism (CD) spectrometer, fluorescence and absorption or CD can be detected, respectively. These instrumentations can be used to observe rapid unfolding and/or refolding of proteins and parameters like reaction rates, the complexity of the reaction mechanism and information on short-lived intermediates etc can be determined. Effects of several factors

(temperature, pH, reagent concentration, additives etc) on the kinetics of a reaction can be measured by stopped flow experiments.

Rapid mixing experiments were carried out using an SX18-MV stopped-flow apparatus in single mixing mode with a mixing ratio of 1:10 at 20 °C. Protein (100 µM) was diluted in various concentrations of refolding or unfolding buffer. The final protein concentration was 1 µM. Fluorescence was excited at 280 nm, and a cut-off filter at 305 nm was used. Excitation and emission slits were set to 0.4 mm each. Five traces were averaged in each case.

2.6.3.4   Analysis of kinetic data

Exponential changes in fluorescence were fitted using Origin. Different exponential functions (single and double) were tested and the residual compared. The fit with the best residual was chosen for the single jumps. Global fitting of the chevron plot was performed using Berkeley Madonna (www.berkeleymadonna.com).

2.6.3.5   1-Anilino-8-naphthalene-sulfonate (ANS) binding assay

ANS binding for samples incubated with ANS for at least 2 h was measured by fluorescence emission recorded from 400 nm to 650 nm, with excitation and emission slit widths of 2 and 3 nm (for FluoroMax-4 Spectrofluorimeter) and 3 and 5 nm (for Jasco Spectrofluorimeter), respectively. Experiments were performed at a protein concentration of 10 µM and an ANS concentration of 100 µM in 1 cm quartz cuvette. All spectra were accumulated three times and averaged.

2.7   Amyloid experiments and aggregation assays

2.7.1   Thioflavin T fluorescence

The benzothiazole dye Thioflavin T (ThT) is a fluorescent dye widely used *in vitro* and *ex vivo* to monitor the process of amyloid fibril formation. In the unbound state ThT has its excitation and emission maxima at 342 nm and 430 nm, respectively, whereas upon interacting with amyloid fibrils, the excitation and emission bands undergo a red-shift to 444 nm and 482 nm maxima, respectively, accompanied with an enhanced fluorescence emission intensity. ThT emission intensity appears to be in a good approximation, proportional to the amount of amyloid fibril present (118,228,229). Interestingly, ThT quantum yield may vary considerably depending on the morphology of the amyloid, (230). The specificity and sensitivity of this dye is however greatly debated and the binding mechanism of ThT to amyloids remains unclear (119). This scenario is further complicated by the observation that ThT can binds structures other than amyloids, for example nucleic acids probably through charged interactions between thioflavin dye micelles and negatively charged nucleic acids as proposed by Kurana et al., (2005), who also suggested a role of micelle formation in the increase of fluorescence excitation and emission of ThT bound to the surface of amyloid fibrils (119). It was reported that ThT binds fibrils parallel to their long axis β-sheet (231,232). The steric constraints

imposed by this environment are unique to amyloid fibrils and might offer an explanation for the specificity of the dyes.

In this work, ThT was used as a probe to follow amyloid formation kinetic during amyloid induction assays by ultrosonication and to assess the presence of amyloid fibril in experimental (end-stage) samples after gentle agitation (without ultrasonication).

2.7.2   Amyloid fibril induction assay with ultrasonication (performed in the Goto-lab, Osaka, Japan)

The application of ultrasonic irradiation to samples accelerate the fibrillation process (233-238). Ultrasonication of multiple samples in a multi-well plate coupled with ThT fluorescence measurements enables the simultaneous investigation of several samples (233-238). For the standard reaction, the V_L variants were diluted at a concentration of 30 μM in PBS solution and 5 μM ThT. Then, 0.2 ml of the V_L variant solution was applied to each of the 96 wells of a microplate (Greiner bio-one 96 well microplate 675-074 made of polystyrene with a size of 128 mm x 85 mm). The micro plate was set on the HANABI (HANdai Amyloid Burst Inducer), which combines the use of a water-bath-type ultrasonicator (ELESTEIN, Elekon, Chiba, Japan) and a microplate reader (SH-9000, Corona Electric Co., Ibaraki, Japan). The water bath with 400 mm x 250 mm x 230 mm (height) contains about 12 L water and three ultrasonic transducers are set, two on the sides and one at the bottom. Pulses were applied to the microplate from three directions, focusing on the center position, for cycles of 1 min followed by a quiescent period of 9 min. The measurement temperature was maintained at 37 ºC. The formation of fibrils was monitored by ThT fluorescence with excitation and emission wavelengths of 450 nm and 490 nm, respectively. Microplates were incubated at 37°C and shaken before the measurements. TEM images were taken using a HITACHI H-7650 transmission microscope, (Hitachi, Tokyo, Japan), with an acceleration voltage of 80 kV. Aliquots (5 µl) of the samples were placed on a 400-mesh copper grid covered by a carbon-coated colloidal film for 60 s. Two percent (w/v) uranyl acetate solution (5 µl) was then placed on the grid to stain the samples for 60 sec. Excess sample solutions on the grid were removed with a filter paper. The magnification on the representative pictures was 30,000 fold*.*

2.7.3   Amyloid fibril induction assay with gentle agitation

Samples of V_L at a concentration of 30 µM in PBS buffer pH 7.4, (with or without 0.5 mM SDS) containing 0.05% sodium azide (or in acetate buffer pH 2) were incubated with shaking (gentle agitation) in a roller shaker at 37 °C for at least one week. At various intervals, 20 µl aliquot was withdrawn for ThT binding and TEM. For the ThT binding, 10 µl of protein aliquot was added to 480 µl of a PBS buffer pH 7.4 followed by the addition of 10 µl of 500 µM ThT prepared in PBS buffer/pH 7.4 (ThT final concentration 10 µM). This was left to incubate for 90 sec before acquiring the emitted signal in a 1-cm quartz cuvette from 430 – 570 nm with excitation at 440 nm. The spectra were recorded using a Jasco FP-6500 Spectrofluorimeter (Jasco, Grossumstadt, Germany) equipped with a Jasco ADP-303T Peltier as temperature controller. The excitation and emission slit widths were both 3

nm. A Scan speed of 100 nm/min, an integration time of 1 s and a 0.5 nm recording interval were used. Three scans were averaged for each sample. Rayleigh (elastic) light scattering at an angle of 90° was also measured at the excitation wavelength peak of 440 nm (93,239). That is, for the same sample, information on its light scattering ability (presence of macromolecular assemblies or aggregation) and ThT binding (fibril formation) was assessed. For comparison, the Rayleigh scattering intensity at 440 nm and the ThT fluorescence intensity at 480 nm were considered.

2.7.4   Aggregation assays

Rayleigh (elastic) scattering of light can be used to follow the growth of macromolecular assemblies in solutions or aggregation/polymerization kinetics in the low protein concentration rage (93,239). According to Rayleigh's theory, Rayleigh scattering applies only to non-absorbing particles with sizes smaller than the wavelength of the incident light, which easily holds true for visible light, monomeric proteins or small aggregates. Therefore, Rayleigh scattering was used to follow the generation of aggregates during thermal induced-unfolding and to monitor the growth of large molecular assemblies (aggregates) during amyloid induction assay with gentle agitation (section 2.6.3). For this, Rayleigh scattering at an angle of 90°, measured as the maximum of the elastic peak of excitation light during ThT fluorescence was used (see section 2.6.3).

2.7.4.1   Thermal-induced aggregation

V_L samples at a concentration of 10 µM in PBS buffer, pH 7.4 with 10 µM ThT were subjected to increasing temperatures from 20-80 °C in steps of 5 °C in a 1-cm path cuvette. After thermal equilibration for 3 min, emission spectra were recorded for Rayleigh scattering and ThT fluorescence (section 2.7.3) (93).

2.8   Quaternary structure analysis

2.8.1   Analytical ultracentrifugation

Analytical ultracentrifugation (AUC) is a versatile and powerful technique used extensively to characterize hydrodynamic and thermodynamic properties of proteins and other macromolecules in solution. AUC can be applied in the study of biomacromolecules in a wide range of solvents and over a wide range of solute concentrations. There are three optical systems (absorbance, interference and fluorescence) available in the analytical ultracentrifuge, that allow for a precise and a selective observation of sedimentation in real-time (240). AUC makes use of the application of centrifugal force together with the simultaneous real-time observation of the sedimentation of macromolecules in the centrifugal field. An advantage of AUC experiments is that proteins can be studied in solution without influencing their hydrodynamic and thermodynamic properties by any interaction with a matrix or a surface like in size exclusion chromatography (SEC).

Two basic types of ultracentrifugation experiments can be used for protein characterization: (1) sedimentation velocity (SV) and (2) sedimentation equilibrium (SE) methods. Typically SV

experiments are run at relatively high rotor speeds. That is, a sufficiently large centrifugal force field is applied that leads to a movement of molecules towards the bottom of the centrifuge cell. The sedimentation process is determined by three factors: the gravitational force, the buoyancy and the hydrodynamic friction. The gravitational force is proportional to the square of the rotor speed, thus adjustment of the rotor speed allows for the study of a wide range of particle sizes, ranging from kilodalton (kDa) to gigadalton (GDa) molecular masses. In other words, in SV experiments, each protein species in a mixture forms a unique boundary and sediments at a characteristic speed governed by its molecular mass, size and shape (220). The Svedberg equation accounts for the balance of all three forces:

$$ s = \frac{v}{\omega^2 r} = \frac{MD(1 - \overline{V}\rho)}{RT} \tag{2.5} $$

Equation 2.5: Svedberg equation.
S = sedimentation coefficient (s), v= observed radial velocity (m/s), ω = angular velocity of the rotor (m/s^2), $\omega^2 r$ = centrifugal field, M = molar mass (g/mol), D = diffusion coefficient (m^2/s), $\overline{V}$= partial specific volume (cm^3/g), ρ = density of the solvent (g/cm^3), R = gas constant (8.314472 J/K mol), T = absolute temperature (K).

The sedimentation coefficient describes the sedimentation velocity v in relation to the centrifugal field $\omega^2 r$. The s-values are commonly reported in Svedberg (S) units, which correspond to 10^{-13} s.

Sedimentation equilibrium experiments are conducted at relatively low rotor speeds, the centrifugal force is adjusted sufficiently small so that the process of diffusion significantly opposes the process of sedimentation and an equilibrium concentration distribution of macromolecules is obtained throughout the cell. For an ideal non-interacting single component system, the equilibrium distribution is an exponential function of the buoyant mass of the molecule, $M(1 - \overline{V}\rho)$, as described by:

$$ c(r) = c_0 \cdot e^{\frac{M(1 - \overline{V}\rho)\omega^2(r^2 - r_0^2)}{2RT}} \tag{2.6} $$

Equation 2.6: Equilibrium within an ideal non-interacting single component system.

$c(r)$ = sample concentration at radial position r, c_0 = sample concentration at reference radial distance r_0, $\overline{V}$ = partial specific volume, ρ = density of the solvent, ω = angular velocity of the rotor, R = gas constant, T= absolute temperature.

The concentration gradient of proteins at equilibrium is widely used to determine the molecular weight, stoichiometry, binding affinity and virial coefficient of hetero-interacting or self-associating systems (220).

2.8.1.1 Determination of oligomeric state of protein variants

Oligomeric state and the size of various native protein variants were determined by AUC sedimentation velocity (SV) runs at 25 °C. Experiments were carried out with a ProteomLab XL-I (Beckman) equipped with absorbance and interference optics. 400 µl of a 20 µM sample and 410 µl of PBS buffer were loaded into assembled cells with quartz windows and 12 mm pathlength charcoal-filled epon double-sector centerpieces and centrifuged at 42,000 rpm in an eight-hole Beckman-Coulter AN50-Ti rotor. Sedimentation was monitored at 280 nm and continuously scanned with a

radial resolution of 30 µm and three replicates were taken. Data analysis was carried out with SEDFIT using the continuous $c(S)$ distribution mode of SEDFIT (241,242).

2.8.2   Microscopy

2.8.2.1   Transmission Electron Microscopy

The transmission electron microscope (TEM) operates on the same basic principles as the light microscope but uses electrons instead of light. Light microscopes have a limited resolution due to the wavelength. As TEMs use electrons with much shorter wavelength as a "light source" it is possible to get a resolution a thousand times better than with a light microscope (resolution limit is approximately 10^{-10} m). The possibility for high magnifications has made the TEM a valuable tool in both medical, biological and materials research (243).

In this work, TEM was used to detect the presence of fibrils (244). For the measurements (performed by Bettina Richter, Dep. Chemistry, Technische Universität München), 10 µl protein aliquot were fixed on a 200-mesh activated copper grid and incubated for 60 seconds. After a washing step with 10 µl H_2O, the samples were stained with 10 µl of a 1.5 % uranyl-acetate solution. Samples were recorded with an acceleration voltage of 100 kV at a magnification of 33.000x on a Transmission Electron Microscope JEM100CX (JEOL, Tokyo, Japan).

2.8.2.2   Atomic Force Microscopy

The atomic force microscopy (AFM) is another high-resolution imaging technique that can even resolve features on a molecular and atomic level. It is used for determining the surface topography of biomolecules at sub-nanometer resolution (245-247). This technique is widely used to observe the structure and growth of amyloid fibrils (248). AFM works by bringing a cantilever tip in contact with the surface to be imaged. An ionic repulsive force (pN-nN range) from the surface applied to the tip bends the cantilever upwards and the amount of bending, which is measured by a laser spot reflected on to a split photodetector, can be used to calculate the force. By keeping the force constant while scanning the tip across the surface, the vertical movement of the tip follows the surface profile and is recorded as the surface topography map of the sample. The method can be used for imaging any conducting or non-conducting surface.

AFM imaging was performed as a complementary method to TEM (248). A benefit of AFM over TEM is that samples can be viewed in their native state (no special treatment of samples). For AFM, 20 µl of sample was transferred to a freshly cleaved mica disc. The disc was washed three times with water and air dried. AFM measurements were carried out with a Digital Instruments multimode scanning probe microscope (Veeco) in contact mode using DNP-S20 tips (Veeco), as described in (209).

2.8.3   Nuclear Magnetic Resonance (performed in the Reif-lab, Dep. Chemistry, TUM)

Structure determination by Nuclear Magnetic Resonance (NMR) is based on the magnetic properties of the nuclei in the atoms comprising a molecule. NMR occurs when a nucleus with nonzero spin quantum numbers are placed in a powerful magnetic field and subjected to radiofrequency irradiation. The most informative nuclei in biological materials are ^{1}H, ^{13}C, ^{15}N, ^{19}F and ^{31}P, with a magnetic spin of ½ (220). The resonance signals and their characteristics such as, chemical shifts, linewidth, *J*-coupling, cross peaks and Nuclear Overhauser Effect (NOE), are correlated to the nuclear environment and can be used to determine the 3D structure. NMR can be divided into solution-state and solid-state (ssNMR), they differ substantially in sample preparation, instrumentation parameters and the type of structural information obtained. On the one hand, solution-state NMR is a high resolution method for studying 3D structures of soluble samples. The most commonly used nuclei for solution-state NMR studies of proteins are ^{1}H, ^{13}C and ^{15}N. Resonances of these nuclei and interactions among them often resolve most or all the individual amino acids in small proteins of ≤ 60 kDa. On the other hand solid-state ssNMR is typically used in studies of insoluble and noncrystalline samples such as amyloid fibrils, to obtain high resolution structural information.

In this work, solution-state NMR experiments were performed to understand the influence of some residues on the structure the MAK33 V_L domain. All NMR spectra were recorded using uniformly ^{15}N, ^{13}C labeled proteins in 20 mM phosphate, 50 mM NaCl, pH 6.5 at 298 K. Backbone resonance assignments of MAK33 V_Lκ-wt and MAK33 V_Lκ-I2E and MAK33 V_Lκ-D70N were transferred from the assignments of MAK33 V_Lκ-S20N. The assignment was accomplished based on 3D HNCA, HNCACB, CBCACONH, HNCO and HNCACO experiments (249). Experiments were carried out at 1mM protein concentration. Those residues, for which the assignment could not be transferred from MAK33 V_Lκ-S20N, were assigned based on an HNCA-experiment using a 100µM MAK33 V_Lκ-I2E sample. All spectra were recorded on a Bruker AVANCE 900MHz spectrometer (Bruker BioSpin) equipped with a triple-resonance cryo-probe. Spectra were processed using TOPSPIN 3.2 (Bruker BioSpin) and analyzed with CcpNmr 2.2.2. (250). For determination of residues affected by the I2E or S20N substitution, ^{15}N-HSQC spectra of MAK33 V_Lκ-wt, MAK33 V_Lκ-I2E and MAK33 V_Lκ-S20N each at 50 µM were acquired on a Bruker Avance 600MHz spectrometer equipped with a cryo-probe. For all residues, chemical shift perturbations were determined using the following formula:

$$\Delta \delta^{res} = \sqrt{\left(\Delta \delta^{1H}\right)^2 + \frac{1}{25} \cdot \left(\Delta \delta^{15N}\right)^2} \qquad [2.7]$$

Equation 2.7: determination of chemical shift pertubations.
$\Delta\delta^{res}$ = weighted chemical shift difference, $\Delta\delta^{1H}$ = chemical shift of the amide proton in ppm and $\Delta\delta^{15N}$ = chemical shift of the amide nitrogen in ppm.

For comparison of signal intensities, all signals, which displayed significant overlap in one of the proteins were ignored. Intensities were normalized for both proteins to the same average intensity.

2.9   Molecular Dynamic simulations (performed by M. Zacharias, Dep. Physics, TUM)

Molecular dynamics (MD) is an important technique for computer simulation of complex systems modelled at atomic level. Biological macromolecules such as proteins and their environment can be studied *in silico* using MD simulations (251-254). Simulations can provide details of individual particle motion as a function of time. That is, the time evolution or dynamics of a system can be followed through equations of motions solved numerically and this allows for the derivation of kinetics and thermodynamic properties of the system under consideration (251).

MD simulations were used in this work to gain insights into the influence of some amino acid residues at specific positions on the structure of variable domains. For this, start structures of MAK33 $V_L\kappa$ domains, $V_L\lambda$ and V_H domains were obtained by extracting the corresponding coordinates from the crystal structures 1FH5 (MAK33 $V_L\kappa$), 1OPG, 1AQK and 1VGE, respectively. Residue 2 of MAK33 $V_L\kappa$, 1AQK $V_L\lambda$, 1VGE $V_L\kappa$ and 1VGE V_H was substituted *in silico* with Glu while residue 2 of 1OPG $V_L\kappa$ was replaced with Ile. In another variant, residue 20 and 70 of MAK33 were replaced with Asn (S20N and D70N). *In silico* substitutions were performed using the SPDBV package (255) while selecting the best fitting side chain rotamer. All Molecular Dynamics (MD) simulations and the analysis of root-mean square deviation (RMSD) and fluctuations (RMSF) were performed using the Amber12 package (256). Proteins were solvated in octahedral boxes including explicit ions and explicit (TIP3P) water molecules (257). The simulation systems were first energy minimized (5000 steps) followed by heating up to 300 K in steps of 100 K with position restraints on all heavy atoms of the proteins. Subsequently, positional restraints were gradually removed from an initial 25 kcal mol^{-1} Å^{-2} to 0.5 kcal mol^{-1} Å^{-2} within 0.5 ns followed by a 1 ns unrestrained equilibration at 300 K. All production simulations were performed at a temperature of 300 K and a pressure of 1 bar. Umbrella sampling (US) simulations were performed using the distance between the Cα atom of residue 2 and the Cα of residue 32 at the floor of the binding region for residue 2 in the $V_L\kappa$ domains as a reaction coordinate. A quadratic penalty potential ($k(d_{C\alpha\text{-}C\alpha} - d_{ref})^2$, force constant $k = 2.0$ kcal mol^{-1} Å^{-2}) for the Cα-Cα distance was used with reference distances varying from 11.5 -16 Å in 0.5 Å steps and from 16-20 Å in 1 Å steps. At ~ 12-13 Å residue 2 stays bound to the protein in the cavity as observed in the experimental X-ray structure whereas it adopts a fully exposed state at distances >16 Å. The associated potential of mean force was calculated using the weighted histogram analysis (WHAM) method (258).

CHAPTER 3 – A residue-specific shift in stability and amyloidogenicity of antibody variable domains[1]

3.1   Results and discussion

3.1.1   Biophysical characteristics, Stability and amyloidogenic propensity of two highly homologous $V_L\kappa$ domains

MAK33 is a well-studied IgG antibody with respect to folding and association. In this context, the folding pathway of the V_L domain has been analyzed in detail (193), its amino acid composition is typical for a murine κ/IgG1 light chain variable domain (PDB: 1FH5 (210)). Interestingly, another antibody $V_L\kappa$ domain exists (1OPG (259)) which has identical CDRs but five differences in the framework region (Fig. 3.1a). We were wondering whether these differences affect the structural properties of the domain. Both isolated domains were expressed, purified and characterized. The far-UV circular dichroism (CD) spectra of the $V_L\kappa$ domains were similar with a minimum at 218 nm (Fig. 3.1b) characteristic of the β-sheet conformation. However, the shapes of both native spectra suggest a large contribution of random coils, which might be due to long flexible CDR loops. Their near-UV CD spectra (Fig. 3.1c) with a minimum at 275 nm confirmed that the tertiary structure of both proteins is correctly formed (71,73). The presence of a buried tryptophan in close proximity to the disulfide in antibody domains make it a sensitive probe for studying conformational changes by fluorescence. Intrinsic tryptophan fluorescence spectra of both $V_L\kappa$ domains showed MAK33 $V_L\kappa$ to have a large (~2.5 fold) increase in fluorescence emission intensity (maximum at ~ 358 nm) in the presence of 3 M guanidinium chloride (GdmCl) (Fig. 3.1d), as opposed to a small increase observed in 1OPG $V_L\kappa$ (Fig. 3.1e). This suggests already a partial exposure of tryptophan in the native 1OPG $V_L\kappa$. To verify whether the $V_L\kappa$ domains exposed hydrophobic patches on their surfaces, ANS binding assay was performed. As depicted in Fig. 3.1f, 1OPG $V_L\kappa$ binds ANS with an increase and a blue-shifted fluorescence emission maximum (shift from 525 – 515 nm), suggesting the presence of exposed hydrophobic patches. MAK33 $V_L\kappa$ shows no ANS binding (Fig. 3.1f). The propensity of both MAK33 and 1OPG $V_L\kappa$ domains to aggregate with increasing temperatures was monitored by recording Rayleigh (elastic) scattering of ThT fluorescence excitation (440 nm) light (Fig. 3.1g) (93,239). Surprisingly, 1OPG $V_L\kappa$ already starts aggregating at ~ 35 °C while MAK33 $V_L\kappa$ only starts transforming into aggregates at ~ 50 °C (a difference of ~ 15 °C). Maximum aggregation for both proteins occurred at ~ 65 °C. No ThT binding was observed for any of the proteins at the experimental

---

temperature range of 20 – 80 °C (Fig. 3.2), which implies that the aggregates formed by thermal-induction did not contain amyloid fibrils.
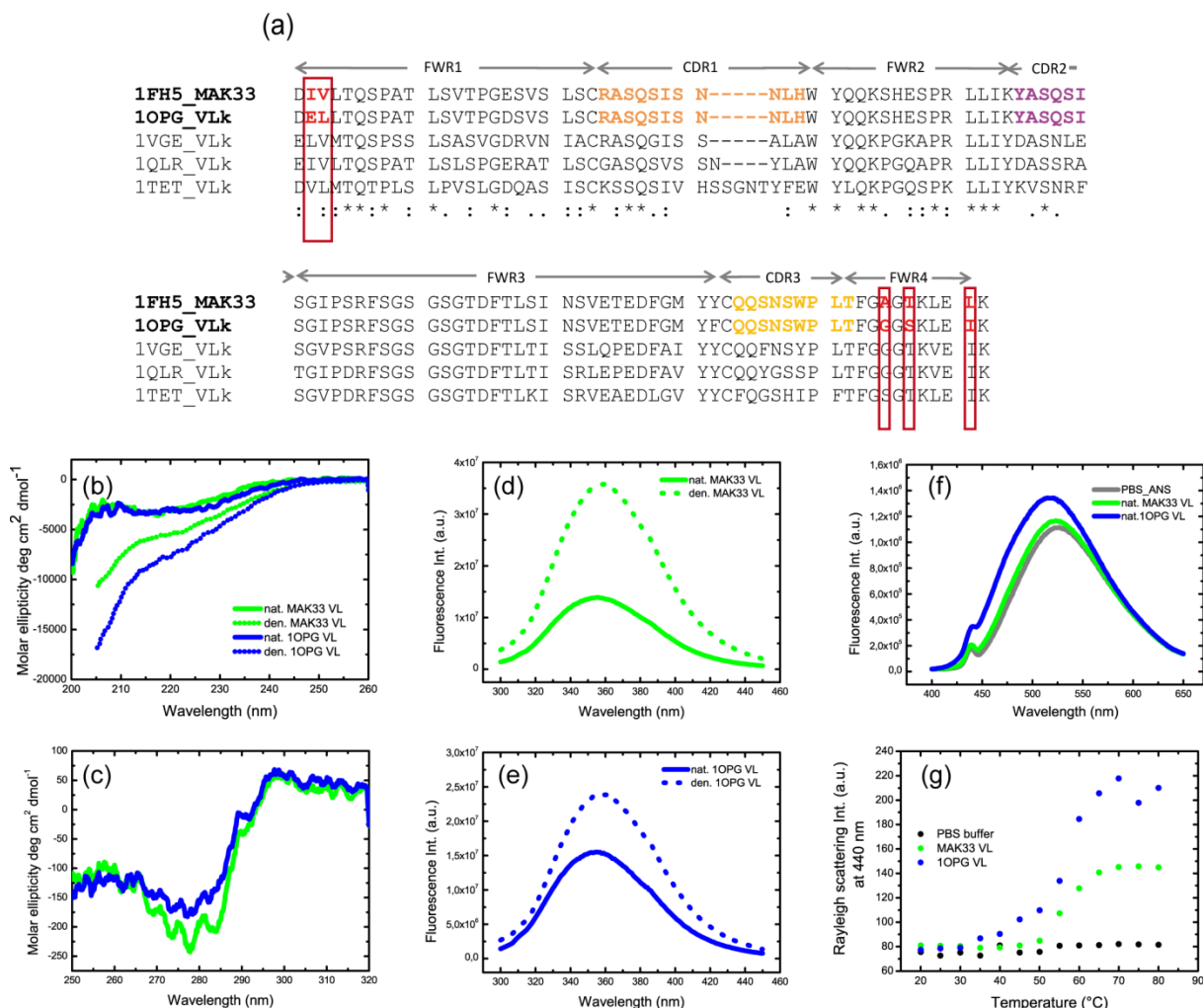

Figure 3.1. Sequence analysis of $V_L\kappa$ domains, spectroscopic characterization of MAK33 $V_L\kappa$ and 1OPG $V_L\kappa$. (a) Multiple sequence alignment of $V_L\kappa$ sequences. Five representative sequences are shown. MAK33 $V_L\kappa$ differs with 1VGE $V_L\kappa$ at 40 positions and with 1OPG $V_L\kappa$ at five positions marked by red rectangles. Identical CDRs between MAK33 $V_L\kappa$ and 1OPG $V_L\kappa$ domains are highlighted in different colors. (b) Far UV CD spectra of native (continuous line) and temperature denatured (dotted line; 60 °C) MAK33 $V_L\kappa$ and 1OPG $V_L\kappa$ in PBS buffer. (c) Near UV CD spectra of native MAK33 $V_L\kappa$ (green line) and 1OPG $V_L\kappa$ (blue line). Intrinsic tryptophan fluorescence spectra of native (continuous line) and 3 M GdmCl-denatured (dotted lines) MAK33 $V_L\kappa$ (d) and 1OPG $V_L\kappa$ (e) in PBS buffer. (f) ANS fluorescence spectra of native MAK33 $V_L\kappa$ (green line) and 1OPG $V_L\kappa$ (blue line), the grey line is PBS/ANS without protein. (g) Thermal-induced aggregation of MAK33 $V_L\kappa$ and 1OPG $V_L\kappa$ domains monitored by Rayleigh (elastic) light scattering at 440 nm.
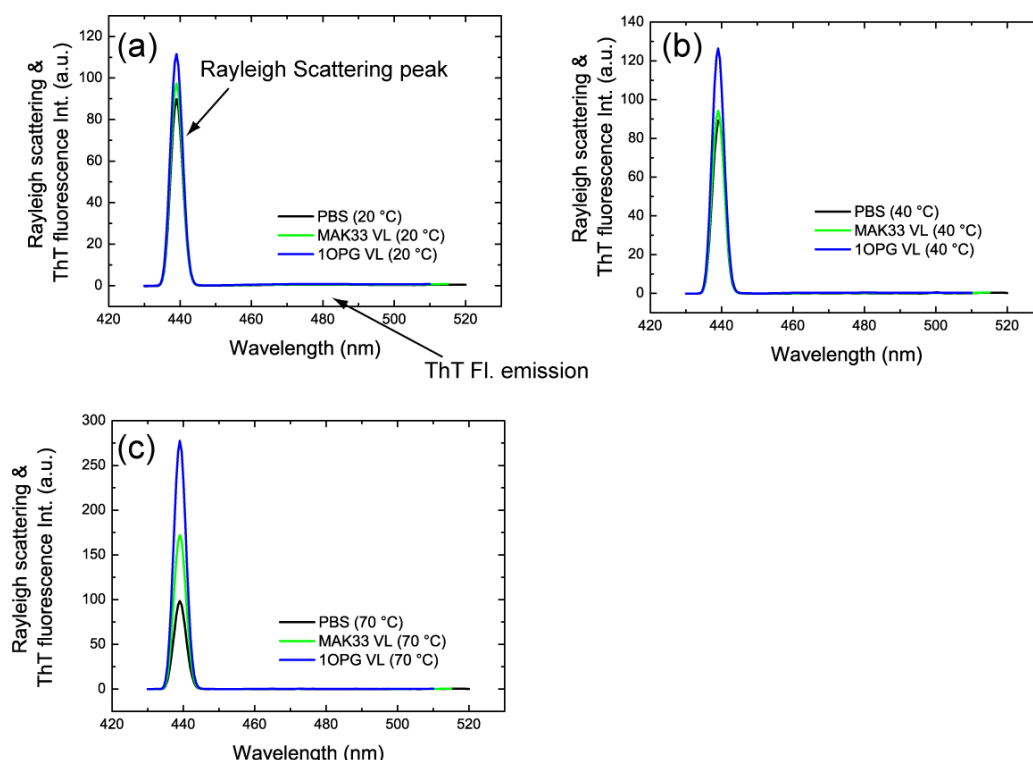
Figure 3.2 . Emission spectra of 1OPG V$_L$κ and MAK33 V$_L$κ recorded at different temperatures e.g. 20 °C (a), 40 °C (b) and 70 °C (c) to study thermal aggregation. Rayleigh (elastic) scattering intensity is recorded at the ThT fluorescence excitation of 440 nm. No ThT fluorescence emission is observed.

When we determined their stabilities against temperature monitored by far UV CD, we found interestingly that the 1OPG V$_L$κ is 16 °C less stable than MAK33 V$_L$κ as judged from their transition midpoints (Fig. 3.3b), which correlates very well with thermal-induced aggregation monitored by light scattering at 440 nm. Thermal-induced unfolding transitions were not reversible for any of the proteins. For equilibrium unfolding and refolding in the presence of GdmCl monitored by Trp fluorescence, the MAK33 V$_L$κ with a cooperative sigmoidal transition had identical midpoints in both the unfolding and refolding directions (Fig. 3.3c). This sigmoidal transition was not observed for 1OPG V$_L$κ for which a native baseline could not be obtained (Fig. 3.3c). The GdmCl-induced unfolding transitions were cooperative and reversible. Assuming two-state transitions, the data were fitted using linear extrapolation (81,218) which yielded an intrinsic stability of -17.8 kJ mol^{-1} for MAK33 V$_L$κ and -3.0 kJ mol^{-1} for 1OPG V$_L$κ and *m*-values of 18.0 ± 2.4 and 12.9 ± 5.3, respectively which reflects the different cooperativities observed. Though the stability of MAK33 V$_L$κ falls within the range of those reported for amyloidogenic V$_L$ domains (approximately 15–20 kJ mol^{-1}) (98,156), it is known to be non-amyloidogenic *in vitro* at physiological pH (193). Accordingly, after one week of incubation in PBS buffer at pH 7.4 and 37 °C with gentle agitation, MAK33 V$_L$κ did not form any amyloid fibrils (Fig. 3.3c). However, under these conditions, 1OPG V$_L$κ readily turned into well-defined fibrillar structures (Fig. 3.3c). When the proteins were incubated in an acidic buffer (pH 2) under similar conditions, fibrils were detected for both MAK33 V$_L$κ and 1OPG V$_L$κ samples (data not shown) confirming the previously published results of MAK33 V$_L$κ (193). This supports the notion that under appropriate destabilizing conditions, even proteins not known to form amyloid fibrils under normal physiological conditions can indeed fibrillize (1,105,260).
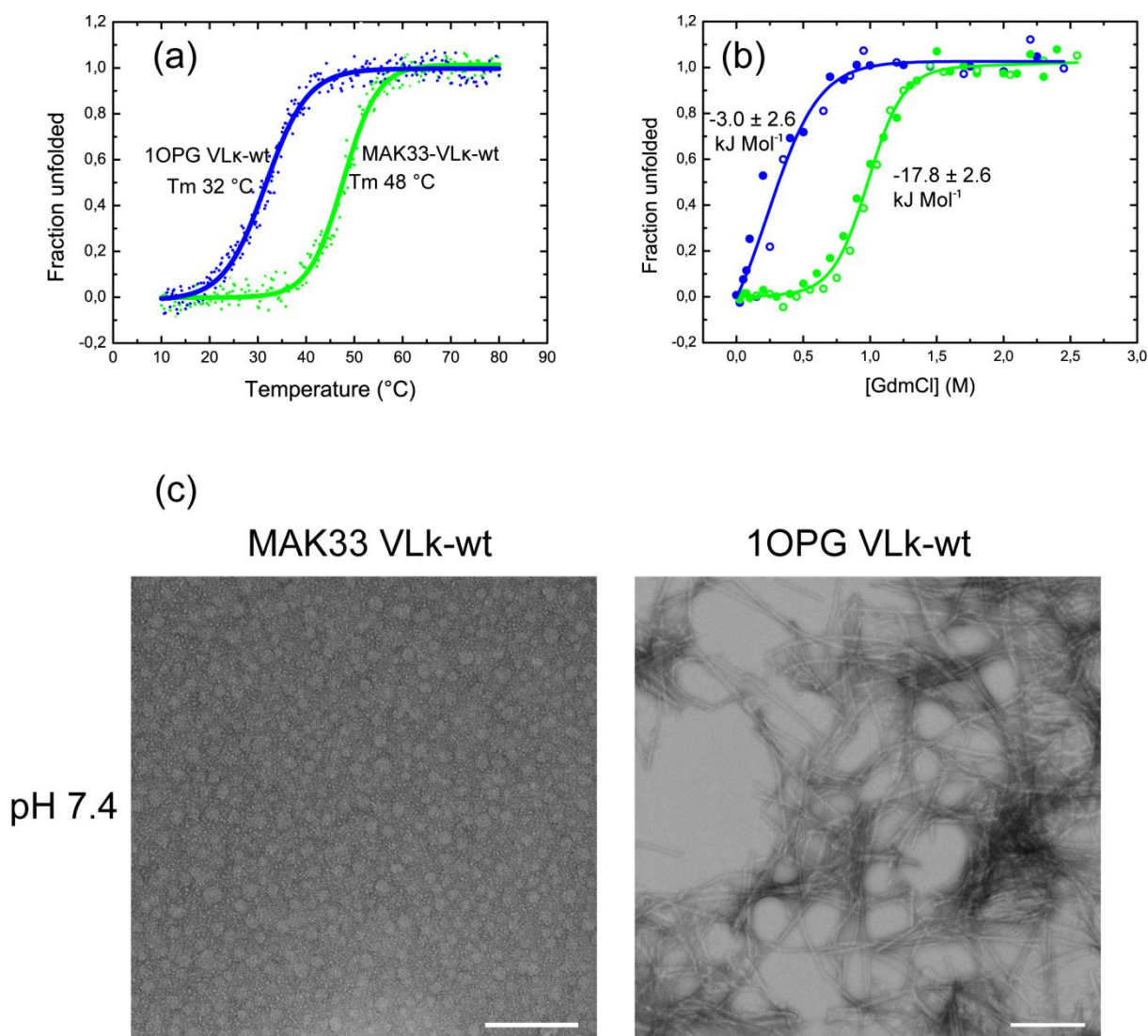
Figure 3.3. Stability and amyloidogenic propensity of MAK33 $V_L\kappa$ and 1OPG $V_L\kappa$ domains. (a) Thermal unfolding transitions of MAK33 $V_L\kappa$ (green symbols) and 1OPG $V_L\kappa$ (blue symbols). The solid lines indicate the theoretical curves derived by fitting the data to a Boltzmann function for MAK33 VLκ (green) and 1OPG $V_L\kappa$ (blue) to obtain transition midpoints (T_{melt}). (b) GdmCl-induced unfolding transitions of MAK33 VLκ (green) and 1OPG $V_L\kappa$ (blue), the reversibility of the GdmCl-induced unfolding process is shown by the overlay of unfolding symbols (open circles) and refolding (closed circles) experiments. The solid lines show the fit to a two-state mechanism for both MAK33 $V_L\kappa$ (green line) and 1OPG $V_L\kappa$ (blue line) to obtain thermodynamic stability values (ΔG_U) and the cooperativity parameters (*m-values*). (c) Transmission electron microscopy micrographs of MAK33 VLκ and 1OPG $V_L\kappa$ from amyloid induction experiments at neutral pH, 37 °C, one week incubation with gentle agitation. Scale bars represent 200 nm.

To assess how conserved the frameworks of both $V_L\kappa$ domains are in comparison to other $V_L\kappa$ sequences, a comprehensive sequence analysis was performed using the abYsis database (http://www.bioinf.org.uk/abysis/) which integrates sequences from the Kabat (261) and IMGT (262) databases. Figure 3.1a depicts five representative $V_L\kappa$ sequences from this alignment. Considering amino acid side chain chemistry, all framework residues are very similar, except for residue 2. In 1OPG $V_L\kappa$, this residue is a Glu while in most other $V_L\kappa$ frameworks an aliphatic residue is found at the same position (Fig. 3.1a). A residue frequency distribution analysis using the abYsis database with all available (~ 20,000 non-identical) LC sequences revealed that Ile occurs with the highest frequency (46 %) at position 2 (Fig. 3.4). Other aliphatic residues like Val, Leu and Ala represent 7-9 % at this position in V_L domains and charged residues (Glu, Asp, Lys and Arg) only < 1% (Fig. 3.4a). In addition, looking at the residue frequency distribution at position 102 of these V_L domains, we

observed a highly conserved Thr with a frequency of 98 % while all other residues had a < 1% frequency of occurrence (Fig. 3.4d). In the more conformationally stable MAK33 $V_L\kappa$, position 102 is occupied by Thr while in the less stable 1OPG $V_L\kappa$, the less frequent Ser is found. With position 2 and 102 showing huge frequency distribution discrepancies between the two $V_L\kappa$ domains, we wondered if these positions were responsible for the different behaviors of the two proteins. In consequence, we exchanged these residues between the two $V_L\kappa$ domains and monitored their effects on stability and amyloid fibril formation.
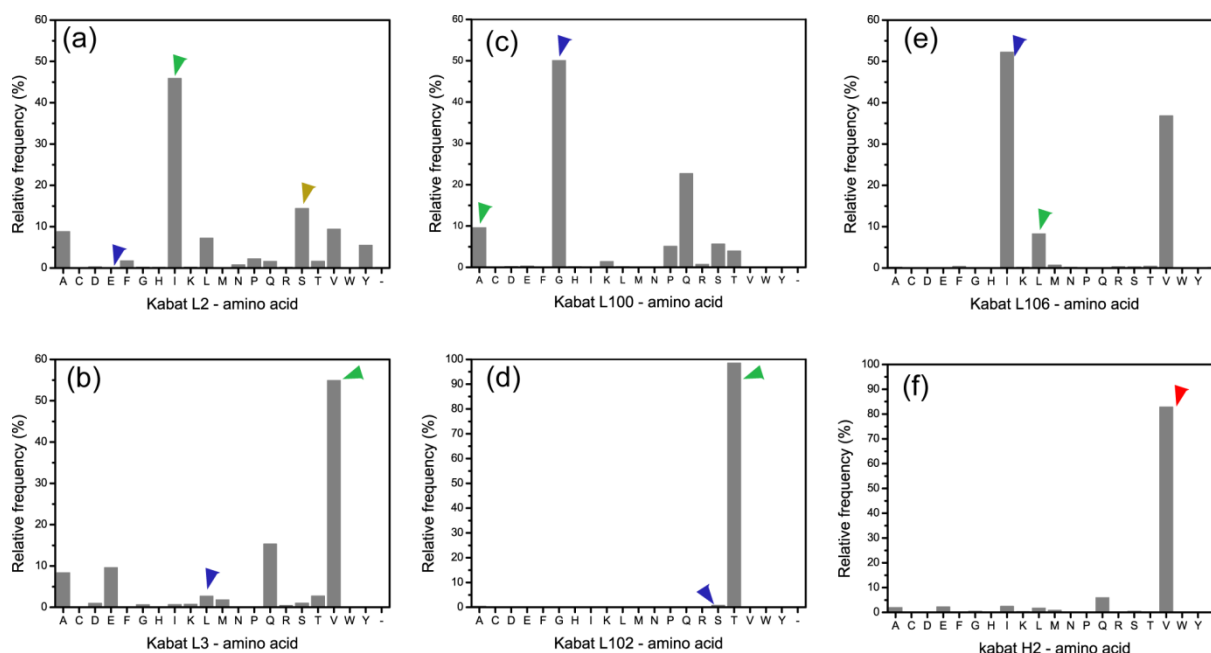

Figure 3.4. Residue frequency distribution analysis at position 2 of variable domains and of the different positions between MAK33 $V_L\kappa$ and 1OPG $V_L\kappa$ domains. Relative distribution of amino acid at various positions of light (a-e) and heavy chains (f). Green arrows indicate residue in MAK33 $V_L\kappa$, corresponding positions in 1OPG $V_L\kappa$ are shown in blue arrows. With the exception of positions 100 and 106, the residues found in MAK33 $V_L\kappa$ at positions (2, 3 and 102) are more frequent than those at corresponding positions in 1OPG $V_L\kappa$. The database could not separate lambda LCs from the kappa LCs during the distribution analysis. But an observation of most of the sequences show that Ser2 is almost entirely contributed by the lambda LCs (orange arrow in "a"). In "f" red arrows indicate the most frequent residue at position 2 of heavy chains. Sequences from all organisms were considered for the analysis. Analysis was based on the Kabat Numbering scheme accessed through the Antibody (abYsis data base) www.bioinf.org.uk: Dr. Andrew C.R. Martin's Group University College London.

3.1.2   Residue 2 is crucial for the stability of $V_L\kappa$ domains

To identify residue(s) that predispose 1OPG $V_L\kappa$ for lower stability and fibril formation, single point variants were generated by replacing residues in 1OPG $V_L\kappa$ with corresponding residues of MAK33 $V_L\kappa$ at any of the five different positions (OP-E2I, OP-L3V, OP-G100A, OP-S102T and OP-I106L). All five substitutions were correctly folded as revealed by UV CD spectra. The thermal unfolding of these variants showed a striking increase in thermal stability for OP-E2I with a T_{melt} value of 43.1 °C, compared to 31.6 °C for 1OPG $V_L\kappa$-wt. Also, OP-E2I was significantly more stable towards GdmCl than 1OPG $V_L\kappa$-wt (Fig. 3.5a, c; Table 3.1). OP-S102T was only slightly more stable and the other single point substitution variants were all similar to the less stable 1OPG $V_L\kappa$-wt (Table 3.1), implying that the exchange of these residues is not important for the conformational stability of the domain. With the high increase in stability seen in OP-E2I and the small increase observed in OP-S102T, we next tested whether a combination of both substitutions (OP-E2I/S102T) and a triple substitution OP-

E2I/G100A/S102T (OP-IAT) would further lead to increased stability. The double and triple substitution variants both unfolded with a T_{melt} of about 45 °C quite similar to that obtained for OP-E2I (Table 3.1), confirming only a small influence of the G100A and S102T substutions on the thermal stability of the protein. However, in terms of stability to GdmCl, both variants were more stable than the respective single substitution variants (Fig. 3.5c; Table 3.1). These stability values observed for the double and triple substitution variants of 1OPG $V_L\kappa$ are similar to that of MAK33 $V_L\kappa$ wild type (Table 3.1). This demonstrates an additive effect of exchanges at different positions.

Since Glu2 and Ser102 to a minor extent were determined to be the destabilizing residues in 1OPG $V_L\kappa$, we were also interested in how they affect MAK33 $V_L\kappa$. Interestingly, MK-I2E was destabilized while MK-T102S had stability values similar to that of MAK33 $V_L\kappa$-wt (Fig.3.5b, c; Table 3.1). Specifically, MK-I2E was ~ 10 °C less stable with a free energy of ~ 12 kJ mol^{-1} less than that of MAK33 $V_L\kappa$-wt. The combined substitution variant (MK-I2E/T102S) was less stable than both single substitution variants with a T_{melt} of 33.6 °C and a chemical stability of -8.0 kJ mol$^-$ (Fig. 3.5c; Table 3.1). This again emphasizes the additive effect of both positions. To demonstrate that our findings on the pivotal role of residue 2 are general to the $V_L\kappa$ family, we investigated a human κ/IgG1 antibody V_L domain (1VGE V_Lk) (263) which differs greatly in amino acid sequence from 1OPG and MAK33 $V_L\kappa$. The N-terminal motive is ELV as opposed to DIV in MAK33 and DEL in 1OPG $V_L\kappa$ (Fig. 3.1a). To assess the role of residue 2 in this human $V_L\kappa$, Leu was replaced with Glu (1VGE V_L-L2E). Notably, analysis of their thermal and thermodynamic stabilities also revealed a decrease in thermal stability of ~ 10 °C and a two-fold decrease in thermodynamic stability for 1VGE V_L-L2E compared to the wild type (Fig. 3.5d, e; Table 3.1). This correlates very well with results obtained for MAK33 and 1OPG $V_L\kappa$, implying that the effects mediated by residue 2 are conserved within the $V_L\kappa$ family.
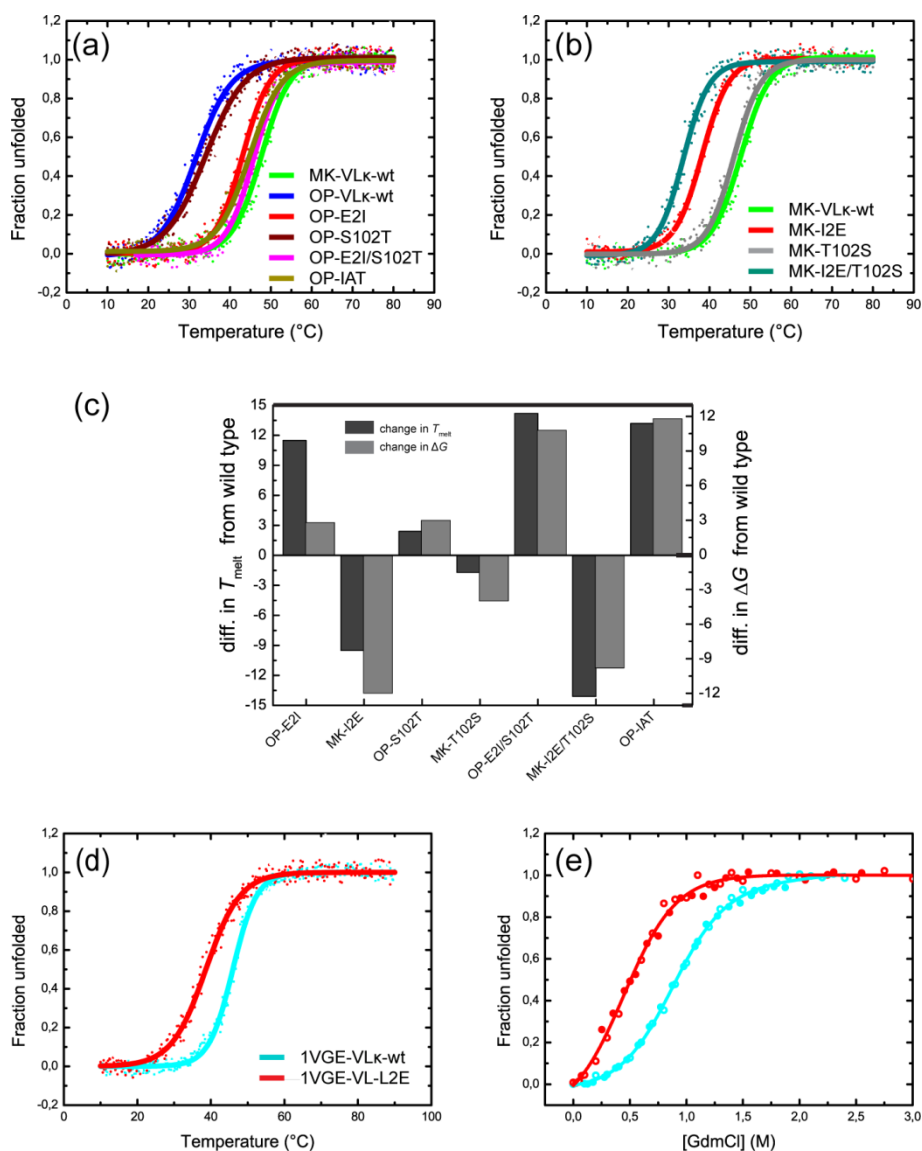
Figure 3.5. Stability of the different V_Lκ variants. Thermal unfolding transitions of (a) 1OPG V_Lκ variants (b) MAK33 V_Lκ variants and (d) 1VGE V_Lκ variants. The solid lines indicate the theoretical curves derived by fitting the data to a Boltzmann function to obtain transition midpoints (T_{melt}). (c) The increase or decrease in stability caused by the different substitutions in 1OPG V_Lκ and MAK33 V_Lκ, obtained by subtracting the T_{melt} and ΔG_U values of the wild type V_L from those of their corresponding mutants. (e) GdmCl-induced unfolding transitions of 1VGE V_Lκ variants, the reversibility of the unfolding process is shown by the overlay of unfolding symbols (open circles) and refolding (closed circles) experiments. The solid lines show the fit to a two-state mechanism for V_Lκ variants to obtain thermodynamic stability values (ΔG_U) and the cooperativity parameters (*m-values*).

**Table 3. 1. Thermal and chemical stabilities of 1OPG VLκ and MAK33 VLκ and 1VGE VLκ variants.**
Stabilities against the thermal and chemical (GdmCl) denaturation of different variants. Midpoints of thermal transitions are shown as T_{melt}. Since GdmCl-induced unfolding transitions were reversible, the data were fitted to a two-state equilibrium unfolding model to obtain the thermodynamic stability of unfolding (ΔGU) as well as the cooperativity parameter (m-value), for a qualitative comparison of the data.

Protein Variant	T_{melt} (°C)	ΔG_U (kJ mol^{-1})	m-value (kJ mol^{-1} M^{-1})
1OPG V_Lκ-wt	31.6 ± 0.2	- 3.0 ± 2.6	12.9 ± 5.3
OP-E2I	43.1 ± 0.1	- 5.8 ± 1.7	14.2 ± 1.7
OP-S102T	34.0 ± 0.2	- 6.0 ± 1.6	16.6± 3.3
OP-E2I/S102T	45.8 ± 0.2	-13.8 ± 3.9	18.1 ± 4.4
OP-E2I/G100A/S102T	44.8 ± 0.4	-14.8 ± 4.0	16.4 ± 4.0
MAK33 V_Lκ-wt	47.7 ± 0.2	-17.8 ± 2.6	18.0 ± 2.4
MK-I2E	38.2 ± 0.1	- 5.8 ± 1.6	15.0 ± 3.0
MK-T102S	46.0 ± 0.2	-13.8 ± 3.1	18.2 ± 3.2
MK –I2E/T102S	33.6 ± 0.2	- 8.0 ± 3.0	19.8 ± 2.3
MK-ΔD1	48.2 ± 0.2	- 14.2 ± 3.7	13.5 ± 3.3
MK-ΔD1/I2	41.7 ± 0.1	- 9.7 ± 1.7	15.6 ± 2.5
1VGE V_Lκ-wt	46.0 ± 0.1	- 9.3 ± 0.4	10.7 ± 0.5
1VGE V_L-L2E	38.6 ± 0.2	- 4.8 ± 0.6	10.7 ± 1.0

3.1.3   Influence of sequence variation on amyloid fibril formation

To examine whether the increase in stability of 1OPG V_Lκ variants also correlates with the resistance to amyloid formation, we set up ultrasonication amyloid-induction assays at neutral pH, 37 °C (233-238). Thioflavin T (ThT) fluorescence was used to monitor the fibrillation process (118,119). All variants that had a thermodynamic stability of < 10.0 kJ mol^{-1} (Table 3.1) bound ThT as indicated by an increase in fluorescence after a lag time of several hours whereas the more stable variants (OP-E2I/S102T, OP-IAT, MAK33 V_Lκ-wt and MK-T102S) showed no ThT fluorescence (Fig. 3.6a, b). The large fluctuations in ThT fluorescence amplitude is likely due to differences in higher order structure of amyloid fibrils as ThT fluorescence can change depending on the morphology of amyloid fibrils (234-236,264). These variations in ThT fluorescence amplitudes might also depend on the mutants and it is worth noting that ThT fluorescence is not always proportional to the amount of amyloid fibrils (264).

 To confirm the presence of fibrils, transmission electron microscopy was performed. As expected, amyloid fibrils were detected in all variants for which ThT fluorescence was observed (Fig. 3.6c). Similar results were obtained with gentle agitation without ultrasonication after 4 days or 10 days incubation of the V_Lκ variants (Fig. 3.7).
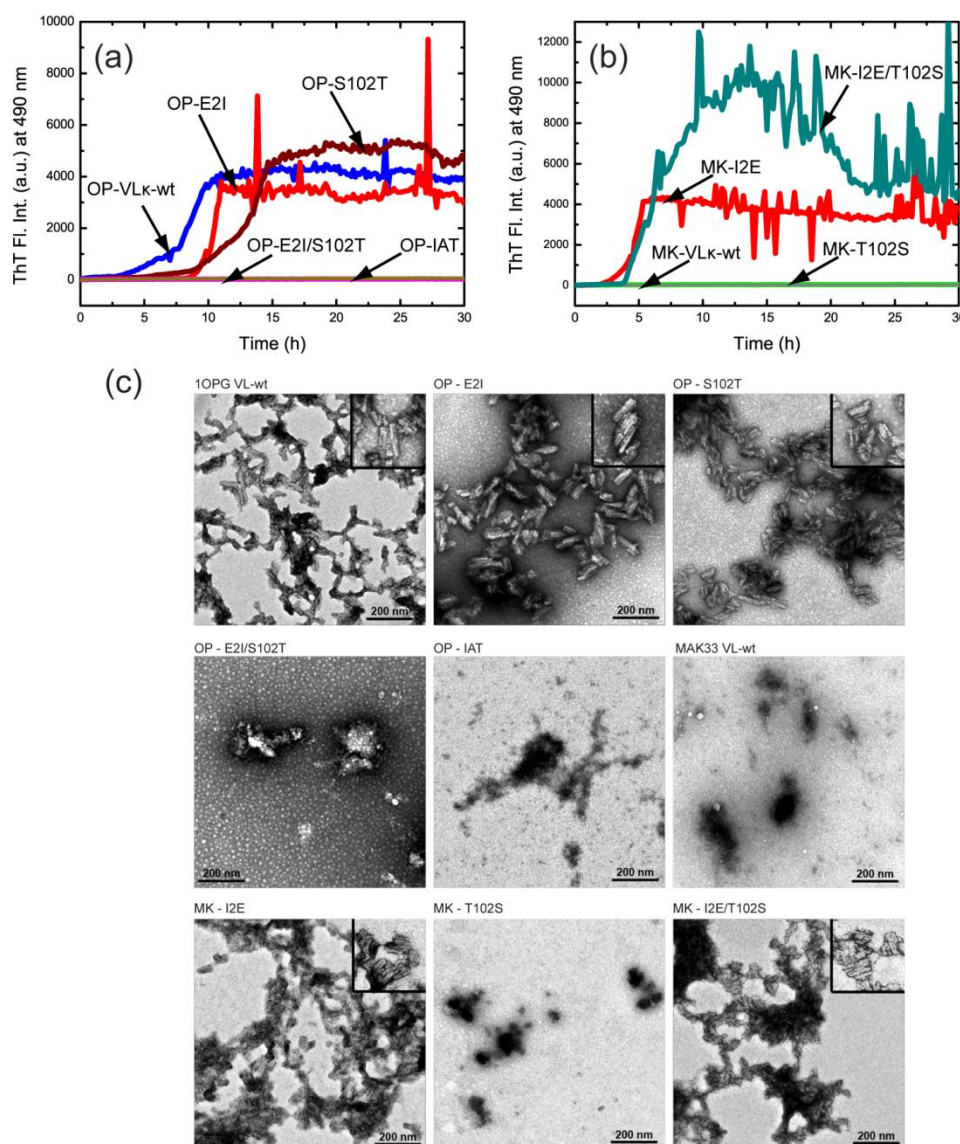
Figure 3.6. Amyloidogenicity of $V_L\kappa$ variants. 30 µM of each $V_L\kappa$ variant in a PBS buffer at pH 7.4 in the presence of ThT was subjected to ultrasonic pulses at 37°C (n=3). ThT fluorescence was monitored over time for 1OPG $V_L\kappa$ variants (a) and MAK33 $V_L\kappa$ variants (b) and TEM micrographs (c) were acquired at the end of the assay to detect the presence of fibrils. Variants with a higher thermodynamic stabilities (> 10 kJ mol^{-1}), OP-E2I/S102T and OP-IAT in (a) and MK-$V_L\kappa$-wt and MK-T102S in (b) did not bind ThT and as a result curves of both proteins are superimposed.

To address the issue of whether variants that do not form amyloid fibrils (do not bind ThT) at neutral pH and 37 °C might contain amorphous aggregates, we set up amyloid induction experiments for MAK33 $V_L\kappa$ and 1OPG $V_L\kappa$ variants without ultrasonication. Emission spectra for Rayleigh scattering intensity and ThT fluorescence were recorded before incubation (Fig. 3.7a), at 4 days of incubation (Fig. 3.7b, c) and 10 days of incubation (Fig. 3.7d, e). Before incubation, only the sample at pH 2 (MAK33 $V_L\kappa$, pH2) showed the presence of larger species (aggregates) but no ThT binding (Fig. 3.7a). However, after 4 days incubation all V_L variants revealed the presence of aggregates detected by an increase in the Rayleigh scattering intensity, though to varying extents. Interestingly, while 1OPG $V_L\kappa$-wt showed an increase in Rayleigh scattering and ThT binding, its stabilized mutants for example 1OPG VL-I2E/G100A/S102T (OP-VL-IAT), only showed an increase in Rayleigh scattering but no ThT fluorescence (Fig 3.7b below). On the other hand, MAK33 VLκ-wt at pH2 and its destabilized variant

MAK33 V_L-I2E both showed an increase in Rayleigh scattering and ThT fluorescence emission, whereas MAK33 V_Lκ-wt at pH 7.4 did not show any ThT emission fluorescence (Fig. 3.7c). Similar results were obtained after 10 days incubation (Fig. 3.7d, e), implying that already at day 4, saturation for aggregation and/or fibrillation was reached. These results confirm our finding that stable V_L variants are resistant to amyloid formation while destabilized variants readily form fibrils. However, although the stable variants were protected from amyloid fibril formation, they did transform into amorphous aggregates upon prolonged incubation at 37 °C, revealed by the increase in Rayleigh scattering intensity (Fig. 3.7b-e). Thus the changes in V_L domain stability induced by modulating residue 2 inversely correlated with its ability to form amyloid fibrils.
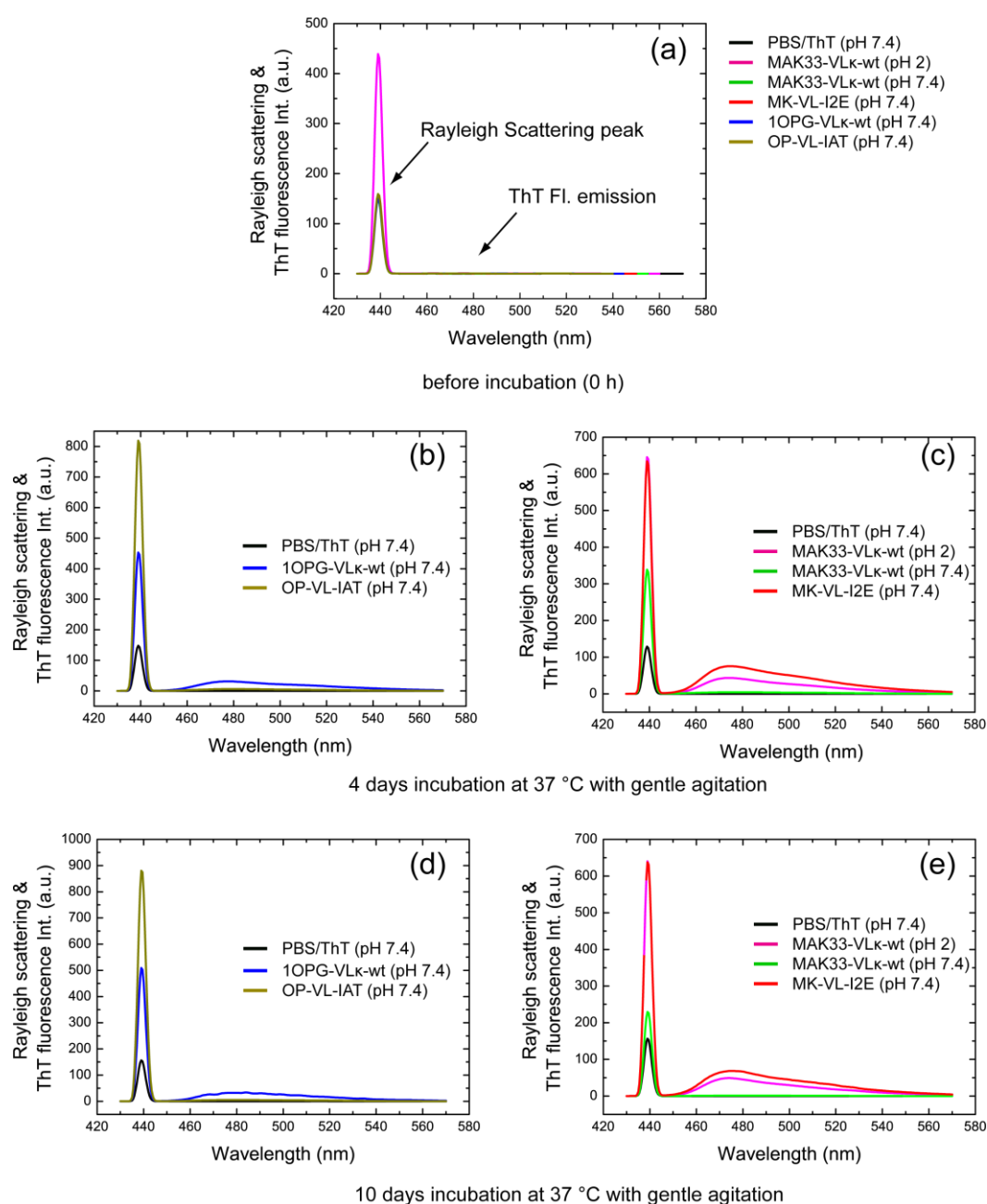

Figure 3.7. Emission spectra of 1OPG V_Lκ and MAK33 V_Lκ variants recorded at different time points, 0 h (a), 4 days (b and c) and 10 days (d and e) after incubation with gentle agitation at 37 °C to monitor aggregation and/or fibril formation. Rayleigh (elastic) scattering intensity was recorded at the ThT fluorescence excitation of 440 nm.

3.1.4   Residue 2 is a decisive factor for $V_L\kappa$ domain architecture

Alignment of $V_L\kappa$ sequences revealed a highly conserved N-terminus for this family, with residue 1 always being an acidic residue (Asp or Glu) and residue 2 mostly an aliphatic residue (Ile, Leu or Val) (Fig. 3.1a). We wondered whether the conserved residue 1 also affects the stability of $V_L\kappa$. Thus, N-terminal truncation variants of MAK33 $V_L\kappa$ were generated in which either the first residue (Asp1) was deleted (MK-$V_L\Delta$D1) or one in which both the first (Asp1) and the second (Ile2) residues were truncated (MK-$V_L\Delta$D1/I2). Deletion of these N-terminal residues did not influence the structure of the domain as determined by CD spectroscopy. Surprisingly, the deletion of Asp1 (MK-$V_L\Delta$D1) did not lead to changes in the thermodynamic stability when compared to the wild type domain. In contrast, the deletion of both Asp1 and Ile2 (MK-$V_L\Delta$D1/I2) resulted in a decrease in T_{melt} and thermodynamic stability of ~ 7 °C and ~ 10 kJ mol^{-1}, respectively (Fig. 3.8a, b; Table 3.1) which is similar to the stability obtained by substituting Ile2 alone with Glu (MK-I2E). These results show that residue 2 and not residue 1 is important for the stability of $V_L\kappa$ domains.

With Glu at position 2 identified to be destabilizing while Ile or Leu at the same position are stabilizing, we were interested to determine how residues with other side chains might affect the stability of these $V_L\kappa$ domains. Considering the residue frequency distribution at position 2 of V_L (Fig. 3.4a), Residue 2 of 1OPG $V_L\kappa$-wt and MAK33 $V_L\kappa$-wt were replaced with Ala, Leu, Asp, Lys or Gln. Interestingly, the thermal unfolding transitions of these variants revealed striking differences. All variants with an uncharged residue at position 2 had a high thermal stability (Table 3.2). While the respective 1OPG $V_L\kappa$ variants showed an increase in T_{melt} of 4-12 °C compared to their wild type, MAK33 $V_L\kappa$ variants with the same uncharged residues at position 2 had T_{melt} values similar to the more stable MAK33 $V_L\kappa$-wt (Fig. 3.8e; Table 3.2). On the other hand, all variants with a charged residue (whether positive or negative) at position 2 were strongly destabilized. OP-E2D and OP-E2K were as unstable as 1OPG $V_L\kappa$-wt while the same substitutions in MAK33 $V_L\kappa$ resulted in ~ 10 °C decrease in T_{melt} (Fig. 3.8e; Table 3.2).
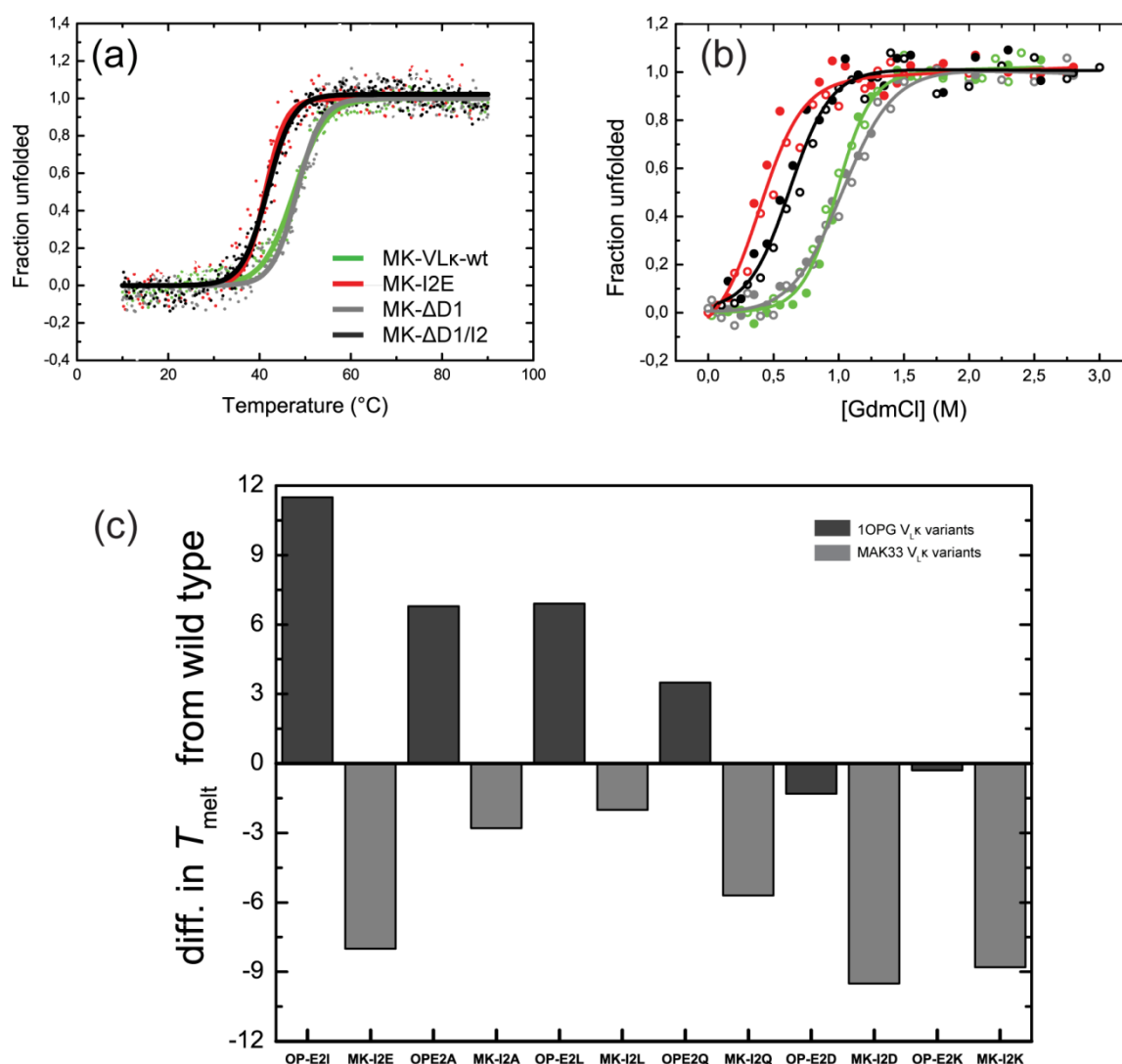
Figure 3.8. Stability of the different MAK33 $V_L\kappa$ N-terninal variants and the different position mutants. (a) Thermal unfolding transitions of MAK33 $V_L\kappa$ N-terminal variants, the solid lines indicate the theoretical curves derived by fitting the data to a Boltzmann function to obtain transition midpoints (T_{melt}). (b) GdmCl induced unfolding transitions of MAK33 $V_L\kappa$ N-terminal variants, the reversibility of the unfolding process is shown by the overlay of unfolding symbols (open circles) and refolding (closed circles) experiments. The solid lines show the fit to a two-state mechanism for $V_L\kappa$ variants to obtain thermodynamic stability values (ΔG_U) and the cooperativity parameters (*m-values*), for a qualitative comparison of the data. (c) The increase or decrease in stability caused by different amino acid substitutions at position 2 of 1OPG $V_L\kappa$ and MAK33 $V_L\kappa$, obtained by subtracting the T_{melt} value of the wild type V_L from those of their corresponding mutants.

Table 3.2. Thermal stabilities of different 1OPG $V_L\kappa$ and MAK33 $V_L\kappa$ residue two variants
Stabilities against thermal denaturation of the different 1OPG $V_L\kappa$ and MAK33 $V_L\kappa$ residue two variants. Midpoints of the thermal transitions are shown as T_{melt}.

1OPG $V_L\kappa$ Variant	T_{melt} (°C)	MAK 33 $V_L\kappa$ Variant	T_{melt} (°C)
1OPG $V_L\kappa$-wt	31.6 ± 0.2	MAK33 $V_L\kappa$-wt	52.0 ± 0.5
OP-E2I	43.1 ± 0.1	MK-I2E	44.0 ± 0.4
OP-E2A	38.4 ± 0.2	MK-I2A	49.2 ± 0.4
OP-E2L	38.5 ± 0.2	MK-I2L	50.0 ± 0.5
OP-E2Q	35.1 ± 0.2	MK-I2Q	46.3 ± 0.4
OP-E2D	30.3 ± 0.6	MK-I2D	42.5 ± 0.4
OP-E2K	31.3 ± 0.4	MK-I2K	43.4 ± 0.5

3.1.5 Residue 2 does not affect the stability of $V_L\lambda$ and V_H domains

Irrespective of the family, variable domains ($V_L\kappa$, $V_L\lambda$ and V_H) share a very similar topology of variable CDRs and a conserved framework. We were interested to determine whether residue 2 also plays a similar role in the lambda V_L and in the V_H domains. A multiple sequence alignment (Fig. 3.9a, b) and a residue frequency distribution analysis at position 2 of $V_L\lambda$ and V_H showed this position in $V_L\lambda$ to be mostly occupied by Ser or Ala (Fig. 3.4a) whereas at the same position in V_H domains, Val occurs with a frequency of 83 % (Fig. 3.4f). To ascertain the effects of residue 2 in these variable domains, a human λ/IgG $V_L\lambda$ (1AQK (265)) and a human IgG1 V_H (1VGE (263)) were examined. In the 1AQK $V_L\lambda$ domain, residue 2 (Asn2) was replaced with either a Glu (1AQK $V_L\lambda$-N2E) or Ser (1AQK $V_L\lambda$-N2S). Unexpectedly, stability studies did not reveal any differences between the wild type and the 1AQK $V_L\lambda$-N2E mutant (Fig. 3.9c and d; Table 3.3). This suggests that the stabilizing effects of residue 2 seen in the $V_L\kappa$ family cannot be ascribed to the $V_L\lambda$ family. Moreover, the mutant that contained Ser at position 2 (1AQK $V_L\lambda$-N2S) was slightly more stable than the wild type and 1AQK $V_L\lambda$-N2E mutant.

Figure 3.9. Multiple sequence alignment of $V_L\lambda$ (a) and V_H (b) domains, irrespective of organism. Five representative sequences are shown. Only the first 60 residues are presented. All sequences have a less diverse framework within each

family. The N-terminus of $V_L\lambda$ is mostly a QAV, ESV or PSV motive with Ser the most frequent at position 2. In V_H, the N-terminus is mostly an EVQ, AVQ or QVK motive with Val the most frequent residue at position 2. Stability of 1AQK $V_L\lambda$ and 1VGE V_H variants, (c) thermal unfolding transitions of the different 1AQK $V_L\lambda$ and (e) 1VGE V_H variants. The solid lines indicate the theoretical curves derived by fitting the data to a Boltzmann function to obtain transition midpoints (T_{melt}). (d) GdmCl-induced unfolding transitions of 1AQK $V_L\lambda$ variants monitored by tryptophan fluorescence, the reversibility of the unfolding process is shown by the overlay of unfolding symbols (open circles) and refolding (closed circles) experiments. The solid lines show the fit to a two-state mechanism for all 1AQK $V_L\lambda$ variants to obtain thermodynamic stabity values (ΔG_U) and the cooperativity parameters (m-values). GdmCl-induced unfolding transitions of 1VGE V_H variants could not be performed due to less protein amounts.

When Val2 of 1VGE V_H was substituted with a Glu (1VGE V_H-V2E) or Ala (1VGE V_H-V2A), again no pronounced thermal stability differences were observed (Fig. 3.9e; Table 3.3). Both 1VGE V_H-V2E and 1VGE V_H-V2A mutants were ~ 3 °C less stable than the wild type. Unlike in the $V_L\kappa$ family, the effects of residue 2 seen in V_H are minor and not specific for any amino acid side chain tested.
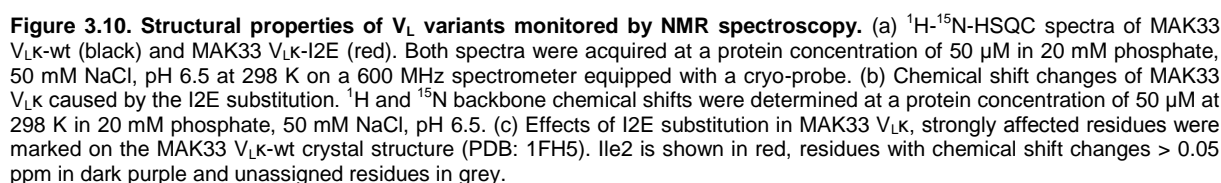
Table 3.3. Thermal and chemical stabilities of 1AQK $V_L\lambda$ and 1VGE V_H variants
Stabilities against thermal and chemical (GdmCl) denaturation of the different 1AQK $V_L\lambda$ and 1VGE V_H variants. Midpoints of thermal transitions are shown as T_{melt}. Since GdmCl-induced unfolding transitions acquired for 1AQK $V_L\lambda$ were reversible, the data were fitted to a two-state equilibrium unfolding model to obtain the thermodynamic stability of unfolding (ΔG_U) as well as the cooperativity parameter (m-value), for a qualitative comparison of the data. When due to too less amount of proteins chemical stability experiments could not be performed (NA).

Variable domain variant	T_{melt} (°C)	ΔG (kJ mol^{-1})	m-value (kJ mol^{-1} M^{-1})
1AQK $V_L\lambda$-wt	44.0 ± 0.2	-5.7 ± 1.3	13.1 ± 1.1
AQ-N2E	44.9 ± 0.5	-6.4 ± 1.7	15.1 ± 1.9
AQ-N2S	44.9 ± 0.4	-11.1 ± 3.0	17.8 ± 3.6
1VGE-V_H-wt	46.5 ± 0.6	NA	NA
VG-V2E	43.1 ± 0.6	NA	NA
VG-V2A	43.8 ± 0.5	NA	NA

3.1.6   Influence of residue 2 on domain structure and dynamics

To obtain a structural understanding of the influence of the residue at position 2 of $V_L\kappa$, both MAK33 $V_L\kappa$-wt and MAK33 $V_L\kappa$-I2E were analyzed by solution-state NMR spectroscopy. 71 % of the non-proline residues could be assigned (Fig. 3.10a). Backbone assignment was hampered by chemical exchange dynamics. The unassigned residues mainly comprise the regions from the N-terminus to Pro8, from His41 to Tyr50 and from Cys88 to Gly101. Figure 3.10b depicts chemical shift changes caused by the I2E substitution. We find particularly large chemical shift differences for residues Cys23 to Ser40 and Ser67 to Leu73. The first region contains the structurally important residues Cys23, which forms a conserved disulfide bond with Cys88. The adjacent Arg24 is involved in a conserved salt bridge with Asp70 and the conserved Trp35 lies in the core of Ig domains (266). Furthermore, the I2E substitution causes structural changes also at distant residues like Tyr86 and Phe87 (Fig. 3.10c). In consequence, a rearrangement of the hydrophobic core involving residues Ala25, Ile29, Trp35, Leu73 and Tyr86 which exhibit major chemical shifts changes is induced. The CDRs also show shift perturbations, especially the loop including residues Arg24 to Ile29 which is close to the N-terminus.

Figure 3.10. Structural properties of V$_L$ variants monitored by NMR spectroscopy. (a) ^{1}H-^{15}N-HSQC spectra of MAK33 V$_L$κ-wt (black) and MAK33 V$_L$κ-I2E (red). Both spectra were acquired at a protein concentration of 50 µM in 20 mM phosphate, 50 mM NaCl, pH 6.5 at 298 K on a 600 MHz spectrometer equipped with a cryo-probe. (b) Chemical shift changes of MAK33 V$_L$κ caused by the I2E substitution. ^{1}H and ^{15}N backbone chemical shifts were determined at a protein concentration of 50 µM at 298 K in 20 mM phosphate, 50 mM NaCl, pH 6.5. (c) Effects of I2E substitution in MAK33 V$_L$κ, strongly affected residues were marked on the MAK33 V$_L$κ-wt crystal structure (PDB: 1FH5). Ile2 is shown in red, residues with chemical shift changes > 0.05 ppm in dark purple and unassigned residues in grey.

To gain further insight in the dynamics of the V$_L$κ domains, molecular dynamics (MD) simulations in explicit solvent were performed on the MAK33 V$_L$κ-wt and 1OPG V$_L$κ-wt domains and on the MAK33 V$_L$κ-I2E and 1OPG V$_L$κ-E2I mutants. The substitutions were generated *in silico* based on the native crystal structures followed by extensive energy minimization before starting MD simulations (see Methods for details). For all four cases, the root-mean-square deviation (RMSD) of the backbone stayed close to the starting structure within < 2.0 Å (Fig. 3.11).
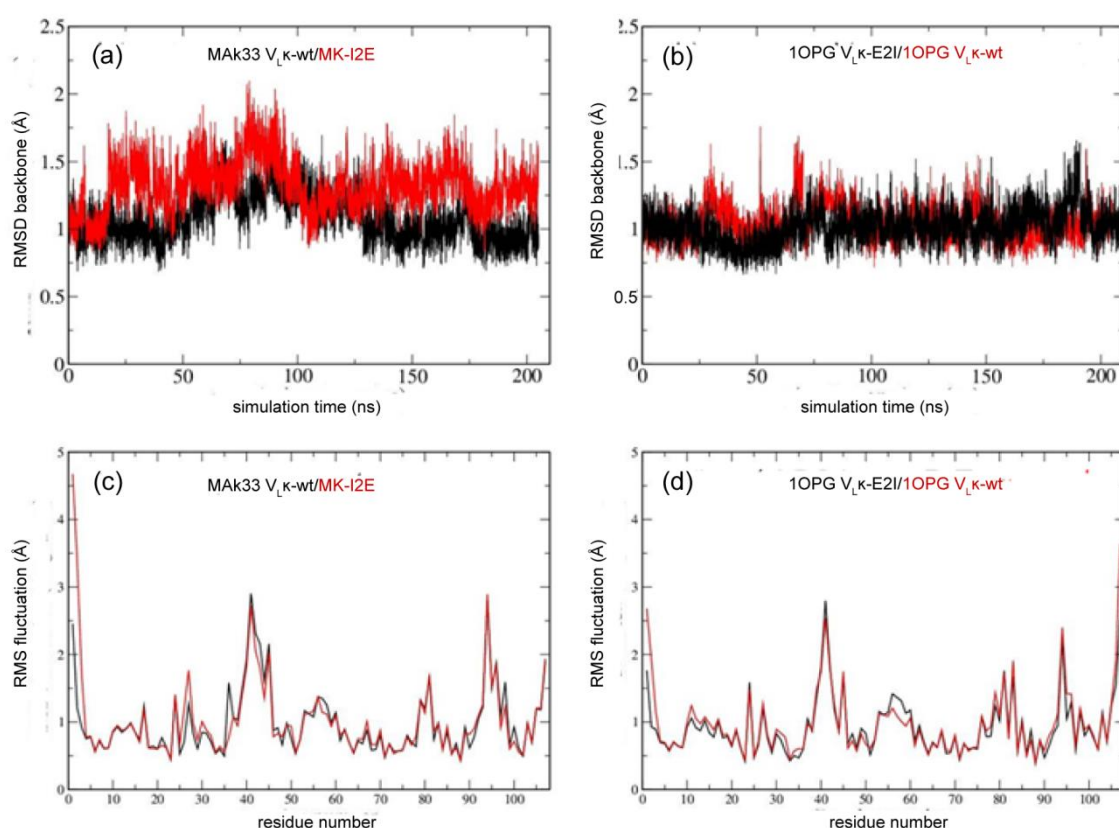
Figure 3.11. (a) Comparison of backbone root mean square deviations (RMSD) from the corresponding experimental start structure versus data gathering simulation time for MAK33 $V_L\kappa$-wt (black curve) and MAK33 $V_L\kappa$-I2E (red curve). (b) same for 1OPG $V_L\kappa$-wt (red curve) and 1OPG $V_L\kappa$-E2I (black curve). (c) Comparison of root mean square fluctuations (RMSF) of amino acid residues along the domain chain for 200 ns simulations of MAK33 VLκ-wt (black curve), and MAK33 VLκ-I2E (red curve). (d) same for 1OPG $V_L\kappa$-wt (red curve) and 1OPG $V_L\kappa$-E2I (black curve).

In the variants with an Ile at position 2 (MAK33 $V_L\kappa$-wt and 1OPG $V_L\kappa$-E2I), the Ile residue remained close to the starting conformation with the side chain located in a hydrophobic cavity (Fig. 3.12a, b, d). During the entire simulation, the RMSD of the heavy atoms of Ile2 stayed at a level similar or only slightly larger than that of the average backbone RMSD of the complete domains (black curves in Fig. 3.12a, b). Interestingly, one can distinguish two rotameric sub-states of the Ile side chain during the simulation with slightly different RMSDs with respect to the starting structure. Contrary to this, in the two variants with a Glu at position 2 (MAK33 $V_L\kappa$-I2E and 1OPG $V_L\kappa$-wt), significant deviations and fluctuations of the Glu2 side chains with respect to the starting structure were observed (red curves in Fig. 3.12a, b). Deviations of up to 6-8 Å from the initial placement were observed which correspond to a partial or full dissociation of the Glu2 side chain from the cavity, thereby adopting a fully solvent-exposed conformation (compare Fig. 3.12d, e). The frequent disruption of contacts between Glu2 and other side chains indicates that Glu does not contribute significantly to the stability.
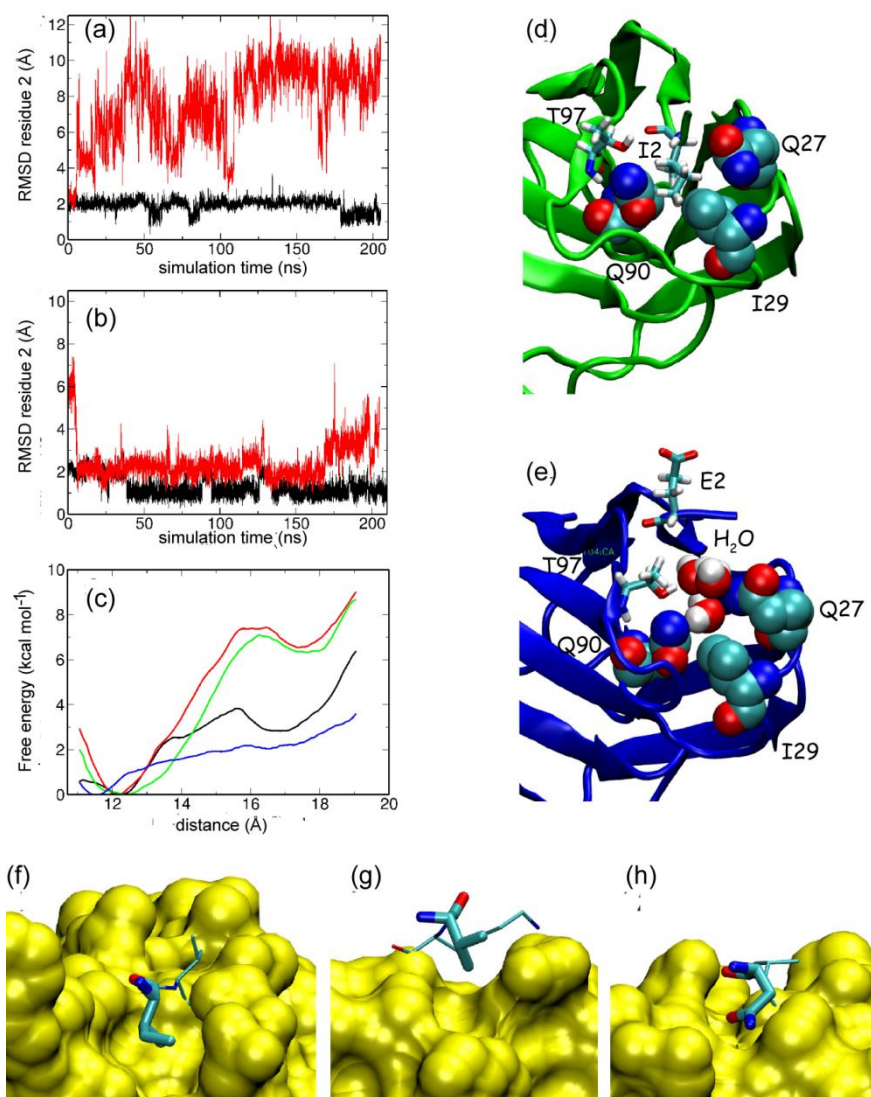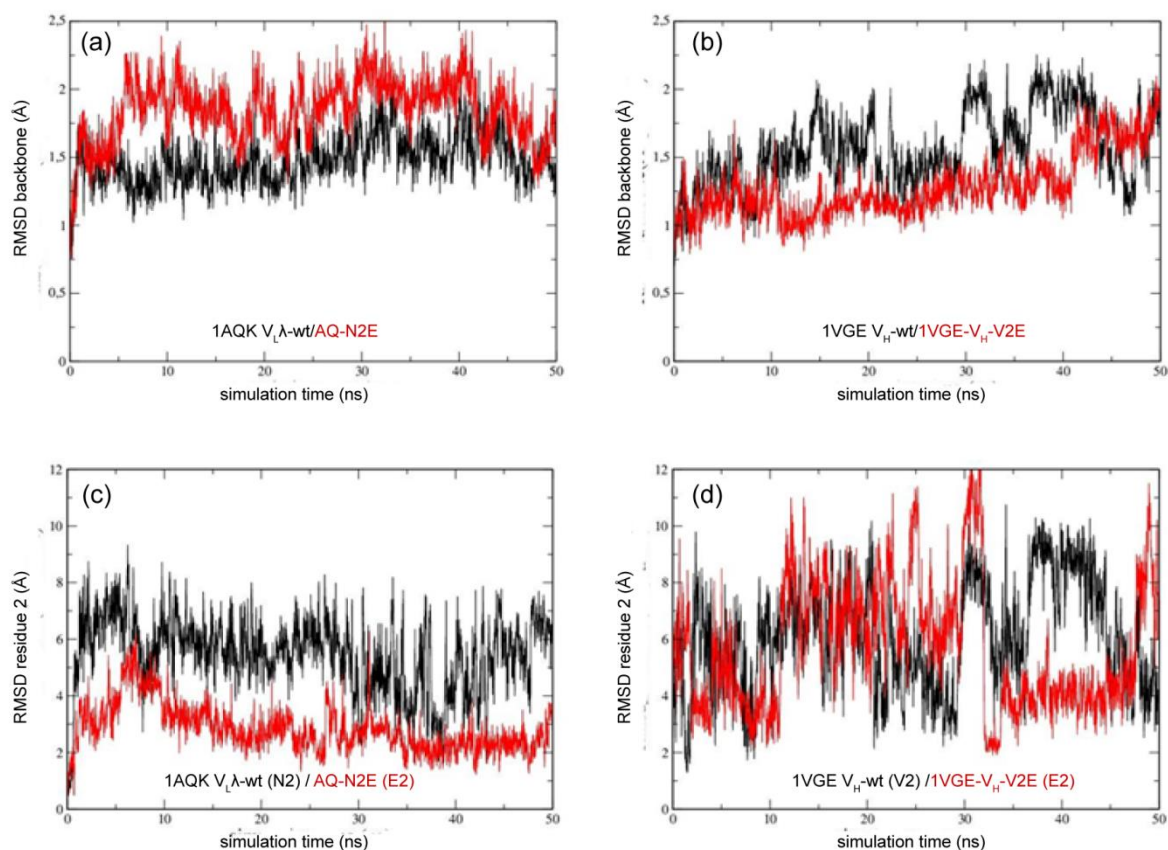
Figure 3.12. Heavy atom root mean square deviations (RMSD), of residue 2 versus simulation time after best superposition on the complete backbone of the start structure for (a) 1OPG V$_L$κ-wt (red: Glu2) and 1OPG V$_L$κ-E2I (black: Ile2) (b) MAK33 V$_L$κ-wt (black: Ile2) and MAK33 V$_L$κ-I2E (red: Glu2). (c) Calculated potential-of-mean force for the dissociation of the Glu2 in 1OPG V$_L$κ-wt (black curve), Ile2 in 1OPG V$_L$κ-E2I (green), Ile2 in MAK33 V$_L$κ-wt (red curve) and Glu2 in MAK33 V$_L$κ-I2E (blue curve) from the hydrophobic cavity region. (d) N-terminal cavity region of MAK33 V$_L$κ with the Ile2 side chain buried in the cavity (protein cartoon with Ile2 as sticks model and adjacent side chains as van der Waals spheres). (e) Example of a simulation snapshot with a fully solvent exposed Glu2 side chain and several water molecules at the rim of the N-terminal cavity region of 1OPG V$_L$κ. (f) Solvent accessible surface representation in yellow, of the binding cavity region for residue 2 in case of MAK33 V$_L$κ-wt (stick model of Ile2) (g) same for 1VGE V$_H$-wt with Val2 as stick representation and (h) same for 1AQK V$_L$λ-wt with Asn2.

The analysis of RMS fluctuations of each residue with respect to the mean structure (Fig. 3.11c, d) showed similar fluctuations for wild type and corresponding mutant chains. However, significantly larger fluctuations were observed for variants with a Glu at position 2 compared to those with Ile2. Slightly increased fluctuations were also seen at loop segments that flank residue 2 (e.g. the loop around Ala25), also identified in NMR experiments to have strong chemical shift perturbations. In order to quantify the relative stability of solvent-exposed and buried states of the side chain at residue 2, umbrella sampling (US) simulations were performed to induce a dissociation of the side chain from the N-terminal cavity region. The distance between the Cα atom of residue 2 and the Cα atom of a residue at the floor of the N-terminal contact region served as reaction coordinate (see Methods for

details). The US-simulations allowed for the calculation of a potential-of-mean force or associated free energy change for the dissociation of residue 2 (Fig. 3.12c). A distance of ~ 12 Å corresponded to a location of residue 2 in contact with the residues forming the N-terminal cavity region (smaller distances resulting in sterical repulsion). At distances of 16-18 Å the side chain of residue 2 had lost contacts with the N-terminal cavity region (larger distances resulting in disruption of additional contacts of subsequent residues along the chain). The US simulations indicate a significantly higher dissociation free energy barrier for the case of an Ile at position 2 compared to Glu2 in both MAK33 and 1OPG $V_L\kappa$ domains. Since in the folded form the Ile2 residue stays in an associated conformation, one can assign ~2.5 kcal/mol (~10 kJ mol^{-1}) as a free energy difference of dissociation compared to Glu2 as a folding stabilization free energy contribution of the Ile2 variants. This is in excellent agreement with the experimental values.

In addition to the V_L-kappa domains, the V_L-lambda and V_H domains (taken from the X-ray structures of 1AQK (265) and 1VGE (263), respectively, were also analyzed. MD simulations were performed starting from the X-ray conformations of the wild type structures and from variants with a Glu at position 2. In all simulations the domain conformations stayed close (within ~ 2 Å) to the start structure for 50 ns sampling time (Fig. 3.13a, b). In the case of 1AQK $V_L\lambda$, the Asn2 wild type residue is partially solvent-exposed forming contacts with polar and charged residues near the N-terminus (Fig. 3.12h). During the simulations, it transiently dissociates to more solvent-exposed states. Similarly, the Glu2 variant (1AQK $V_L\lambda$-N2E) also shows fluctuations to partially solvent-exposed states with limited contacts to the N-terminal region of the protein (Fig. 3.13c). In the case of 1VGE V_H, residue 2 is in a fully solvent-exposed state already in the X-ray starting structure (Fig. 3.12g). No cavity near the N-terminus is present that could serve as a stable anchor region for residue 2 as found for the kappa domains (Fig. 3.12f). Both the simulations of 1VGE V_H-wt and the V2E mutant showed large fluctuations of residue 2 and no stable binding mode could be identified (Fig. 3.13d). Thus, in these cases neither the wild type nor the Glu2 variants adopt a stable state with a buried side chain of residue 2 that could contribute to the stability of the protein, in contrast to what was observed for the Ile2 variants of MAK33 and 1OPG $V_L\kappa$. This is largely due to the absence of an appropriate cavity region near the N-terminus and explains why mutation of residue 2 has little or no influence on protein stability in the case of the $V_L\lambda$ and V_H domains.

Figure 3. 13. (a) Comparison of backbone root mean square deviations (RMSD) from the corresponding experimental start structure versus data gathering simulation time for 1AQK $V_L\lambda$-wt (black curve) and 1AQK $V_L\lambda$-N2E (red curve). (b) same for 1VGE V_H-wt (black curve) and 1VGE V_H-V2E (red curve). Heavy atom RMSD of residue 2 versus simulation time after best superposition on the complete backbone of the start structure for (c) 1AQK $V_L\lambda$-wt (black: Asn2) and 1AQK $V_L\lambda$-N2E (red: Glu2) (d) 1VGE V_H-wt (black: Val2) and 1VGE V_H-V2E (red: Glu2).

3.2   Discussion

We show here that residue 2 is crucial for the integrity and stability of V_L domains of the kappa family and this residue is a key factor which controls its amyloidogenic properties. Specifically, uncharged residues at position 2 of $V_L\kappa$ domains are important for the maintenance of the structural integrity and a high stability. In agreement with their extremely low frequency in naturally occurring $V_L\kappa$ domains, charged residues do not support this.

During NMR backbone assignment, we encountered several regions which are exchanged-broadened and therefore could not be assigned completely. Interestingly, these regions match well with dynamic residues in the dimer interface of the human $V_L\kappa$ sequence LEN, as identified by Mukherjee *et al*. (267). In this V_L domain (LEN), residues 1-9, 37-58 and 89-100 were shown to exhibit pH-dependent millisecond dynamics, which seem to be related to the process of amyloid fibril formation. The presence of residue 2 within these dynamic regions is another indicator for its relevance for amyloidogenicity. This is supported by our MD simulations which show an increase in N-terminal dynamics caused by a charged residue at position 2. Overall, the I2E substitution affects large parts of

the $V_L\kappa$ structure. Both the disruption of the conserved salt bridge and the changes in the hydrophobic core presumably contribute to the loss of thermodynamic stability. However, the short strand-connecting helix (residues 79 to 83) which is highly conserved in antibody variable domains and known to be important for early folding events of the domain (209) was not affected by the I2E substitution, implying that this replacement may not be involved in the amyloidogenicity of the V_L variants studied here.

Hydrophobic core packing is well known to be an important stabilizing factor in proteins in general (268-270) and antibody domains in particular (271). Therefore, the destabilizing effects of charged residues at position 2 could be explained by the fact that they are energetically unfavorable within the hydrophobic core. A charged residue at position 2 of a $V_L\kappa$ domain leads to a partial exposure of the hydrophobic interior and to the disruption of an optimal packing around residue 2. This in turn causes rearrangements of the whole hydrophobic core region. Consequently, this results in the destabilization of the fold and an increased aggregation propensity. For the $V_L\lambda$ and V_H domains, the residue at position 2 is largely solvent-exposed or forms only transient polar contacts with the surrounding protein surface. In MD simulations both the wild-type residues at position 2 or substitutions with Glu resulted in fluctuating solvent-exposed conformational states due to the lack of a stable binding region. Hence, in the case of the $V_L\lambda$ and V_H domains, residue 2 is not part of the hydrophobic core and substitution of this residue also does not disrupt the hydrophobic packing of the domains. Whereas the binding of the N-terminal segment to the hydrophobic cavity is crucial for stability in $V_L\kappa$ domains, it appears to be less important for $V_L\lambda$ and V_H domains. Unlike in $V_L\kappa$ domains where the hydrophobic cavity requires binding of the N-terminal residue 2 to prevent the unfavorable exposure of non-polar residues, in the $V_L\lambda$ and V_H domains, the respective hydrophobic cavities are less hydrophobic and do not require binding of residue 2 for protection.

It has been suggested that the formation of amyloid fibrils does not begin from the native state of a protein, but more likely from a partially folded state or an intermediate state (67,96,98). In this context, the occurrence of particularly non-conservative mutations that involve a change in side chain chemistry in a structurally important region and the correlation with an increased propensity to form amyloid have been reported (98,139). Mutations and changes in conditions that lead to destabilization of the native state might therefore serve to increase the population of an intermediate state or partially folded species, thereby enabling more molecules to be diverted into an amyloid-forming pathway (176,272). In this study, the less stable MAK33 and 10PG $V_L\kappa$ variants readily formed fibrils whereas their stable counterparts did not. Though stability is an important factor determining the amyloidogenic propensity of proteins, the amino acid sequence plays a major role (51) as this determines the amyloid fibril formation kinetics and the types of intermediates populated (96,98,209). Although some residues protect against amyloid formation, their chemistry and position within the protein appears to be more important. This protective role is largely attributed to edge strands of β-sheet proteins (273-275). Residue 2 in antibody variable domains appears in β-strand A which is one of four edge strands (A, D and C'',G) of the β-sandwich topology. Uncharged residues at this position might have evolved to

64

prevent undesirable β-sheet self-propagation (274). The frequent dissociation and solvent exposure of charged residues at position 2 of $V_L\kappa$ domains presumably leads to the disruption of native H-bonds and the subsequent exposure of H-bond donors and acceptors on the edge strand interface. This might result in V_L edge-to-edge aggregation (276). The rare occurrence of charged residues at this position within edge strand A and the high amyloidogenic propensity of these $V_L\kappa$ variants are in line with this explanation. Though there exists a huge similarity between the three different families of variable domains, it has also been shown that they differ significantly in their biophysical properties and propensity to aggregate to amyloid fibrils (271,277). Our study identifies an important property of the $V_L\kappa$ family, not pertaining to the $V_L\lambda$ and V_H domains, which controls its amyloidogenicity.

A compelling amount of evidence from studies on sequence and structural features that can predispose to amyloid formation has often been limited to given V_L subfamilies (e.g. κ1 or λ6) with hardly one feature pertaining to all or an entire variable domain family (187-189). Our results identify a key feature that can destabilize and as a result predispose an entire variable domain family to fibrillar aggregation. The importance of residue 2 for the V_L-κ family determined here adds to the understanding of antibody light chain (AL) amyloidosis linking the destabilization of native interactions with amyloid formation. In the context of AL, and also in the broader context of amyloid diseases, however, it remains to be seen whether point mutations promoting fibril formation generally affect stability. It seems reasonable to assume that other mechanisms may also apply.

CHAPTER 4 – Mechanisms of antibody V$_L$ amyloid formation

4.1   Results and discussion

4.2   Potential natural amyloidogenic mutations

To identify mutations that predispose to AL amyloidosis, several studies have been carried out comparing amyloidogenic LCs to non-amyloidogenic ones (139,150,153,154,159,173,177,184-186). In this study, after a thorough review of the literature, several potential naturally occurring AL predisposing mutations were introduced in the MAK33 V$_L$κ domain (Fig 4.1a). These mutations were selected based on how frequent they appeared in the literature and on their structural relevance to the V$_L$ domain. These single substitution variants were expressed in *E. coli* cells, purified and characterized with the view to determine the reason for their predisposition and mechanism of fibrillation. For example, the replacement of Ser at position 20 with Asn (S20N) (159,191) and Asp at position 70 with Asn (D70N) (139,159,191) introduces an *N*-linked glycosylation site at these positions. Several studies have reported *N*-linked glycosylation of the V$_L$ domain to be a major predisposing factor to amyloidogenesis (139,150,159,191). But despite this evidence, the precise role of glycosylation in AL proteins is yet to be determined. Therefore, in this study, the importance of glycans for the amyloidogenicity of these variants (MAK33 V$_L$-S20N and MAK33 V$_L$-D70N), was assessed by expressing the proteins in non-glycosylating *E. coli* cells. All variants were correctly folded as determined by far UV CD (Fig. 4.1a) and near UV CD (Fig. 4.1b) spectroscopy. The analysis of the oligomeric state of the native proteins by analytical ultracentrifugation sedimentation velocity runs revealed that all variants were monomeric with a sedimentation coefficient of ~ 13 S (Fig 4.1d).
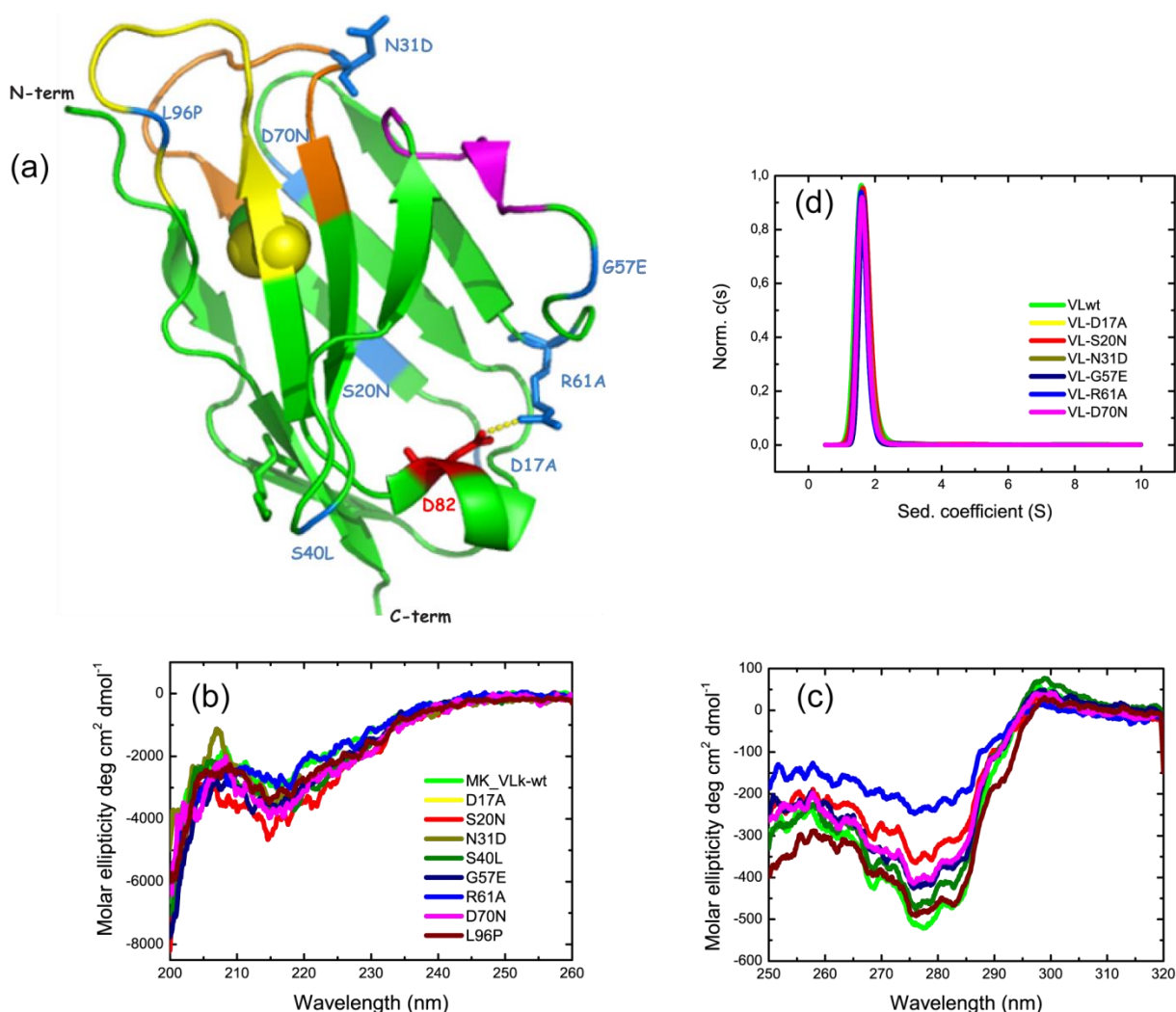
Figure 4.1. (a) The 3D structure of MAK33 V_L domain showing potential amyloidogenic mutantations. Blue patches on the structure depict regions of respective amyloidogenic mutation. D82 (red) with R61 (blue) depicted in "stick" are involved in a conserved salt bridge (yellow dashed line). The CDRs are highlighted in orange (CDR1), pink (CDR2) and yellow (CDR). Also shown is the disulfide bridge buried in the hydrophobic core. Structure based on PDB code 1FH5. Numbering is based on Kabat http://www.bioinf.org.uk/ or http://vbase.mrc-cpe.cam.ac.uk/. (b) Far UV and (c) Near UV CD spectra of native V_L variants in PBS buffer. (d) AUC sedimentation velocity runs, data was analyzed using the continuous c(S) distribution mode of SEDFIT (241,242).

4.2.1   Influence of mutations on the conformational state and stability of the V_L domain

Intrinsic tryptophan fluorescence was used to study the conformational changes of the different variants. Tryptophan fluorescence spectra showed all variants except for the R61A mutant to have a large (~2.5 fold) increase in fluorescence emission intensity (maximum at ~ 358 nm) in the presence of 3 M GdmCl (Fig. 4.2a, b, c, d, f). The R61A mutant had only a small increase and a pronounced red shifted emission maximum upon denaturation (Fig. 4.2e). This suggests already a partial exposure of Trp in the native R61A variant and a change in the polarity of its environment when denatured (219,220). For equilibrium unfolding and refolding in the presence of GdmCl monitored by Trp fluorescence, all variants except for N31D and R61A had a cooperative sigmoidal transition with identical midpoints in both the unfolding and refolding directions (Fig. 4.2g). This sigmoidal transition was not observed for the R61A variant for which a native baseline could not be obtained. In addition, the transitions for the R61A and N31D variants were irreversible, and therefore their ΔG_U could not be

determined (Fig. 4.2g; Table 4.1). Assuming two-state transitions, the data were fitted using linear extrapolation (81,218), which yielded similar intrinsic stability values of ~ -18 kJ mol^{-1} for the MAK33 V$_L$wt and for most of the variants (Fig. 4.2g; Table 4.1). Only G57E and R61A were significantly destabilized with the former having a stability value of ~ -13 kJ mol^{-1}, whereas that of the latter could not be determined due to irreversibility of the transition (Fig. 4.2g; Table 4.1). When their stabilities against temperature monitored by far UV CD was determined, the results correlated very well to those obtained by Trp fluorescence. All mutants with the exception of G57E and R61A had similar thermal stability values of ~ 50 °C like for the MAK33 V$_L$wt, as judged from their transition midpoints (Table 4.1). G57E and R61A variants had a T_{melt} value of 44 °C and 30 °C, respectively.
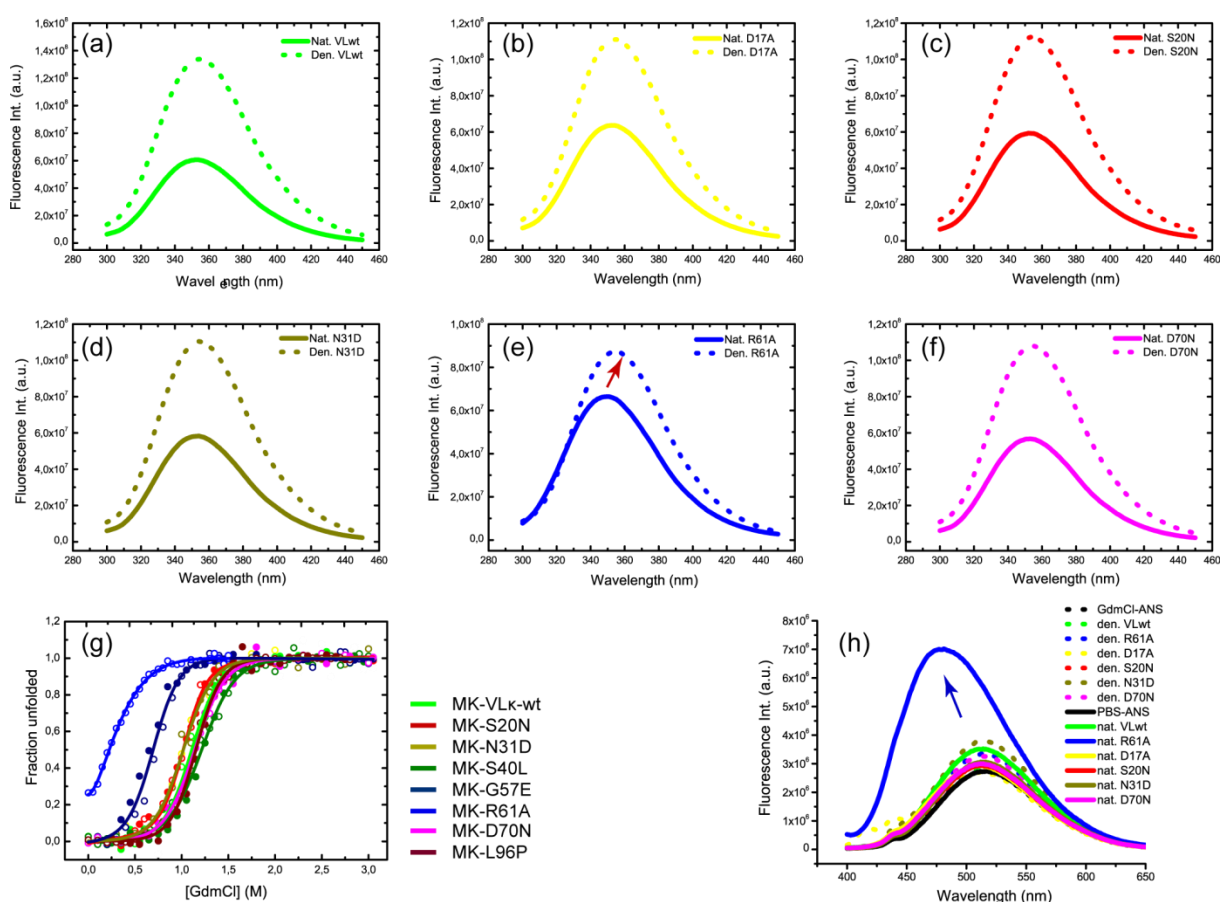
Figure 4.2. Spectroscopic characterization of MAK33 V$_L$ amyloidogenic variants. (a-f) Intrinsic tryptophan fluorescence spectra of native (continuous line) and 3 M GdmCl-denatured (dotted lines) MAK33 V$_L$ variants in PBS buffer. (g) GdmCl-induced unfolding transitions of the different variants, the reversibility of the unfolding process is shown by the overlay of unfolding symbols (open circles) and refolding (closed circles) experiments. The solid lines show the fit to a two-state mechanism for V$_L$ variants to obtain thermodynamic stability values (ΔG_U) and the cooperativity parameters (*m-values*). (h) ANS fluorescence spectra of native (continuous line) and 3 M GdmCl-denatured (dotted lines) protein variants.

To verify whether the different V$_L$ variants exposed hydrophobic patches on their surfaces, ANS binding assay were performed. As depicted in Fig. 4.2h, only the native R61A variant binds ANS with an increase and a blue-shifted fluorescence emission maximum (shift from 530 – 475 nm), suggesting the presence of exposed hydrophobic patches and likely the existence of a molten globule. None of the variants showed binding to ANS when completely denatured (Fig. 4.2h), which supports the idea that ANS does not bind to fully unfolded proteins (226).

Table 4. 1. Thermal and chemical stabilities of MAK33 VLκ variants.
Stabilities against the thermal and chemical (GdmCl) denaturation of different variants. Midpoints of thermal transitions are shown as T_{melt}. GdmCl-induced unfolding transitions were fitted to a two-state equilibrium unfolding model to obtain the thermodynamic stability of unfolding (ΔG_U) for variants with reversible transitions, as well as the cooperativity parameter (m-value), for a qualitative comparison of the data. When due to irreversibility of the transition, ΔG_U could not be performed (NA).

Protein Variant	T_{melt} (°C)	ΔG_U (kJ mol^{-1})	m-value (kJ mol^{-1} M^{-1})
MAK33 V_Lκ-wt	52.4 ± 0.5	- 18.0 ± 2.7	16.2 ± 2.3
MK-D17A	50.0 ± 0.6	- 18.4 ± 3.8	17.4 ± 3.4
MK-S20N	50.7 ± 0.3	- 20.5 ± 3.8	19.9 ± 3.5
MK-N31D	49.8 ± 0.4	NA	15.7 ± 3.2
MK-S40L	49.1 ± 0.2	-17.1 ± 2.0	14.0 ± 1.5
MK-G57E	44.4 ± 0.2	-12.7 ± 2.6	18.1 ± 2.8
MK-R61A	30.5 ± 0.4	NA	12.8 ± 2.9
MK-D70N	49.1 ± 0.4	- 17.4 ± 1.5	15.3 ± 1.3
MK –L96P	53.4 ± 0.2	- 21.1 ± 2.0	18.3 ± 1.6

4.2.2   Fibrillation of V_L variants does not only correlate with the conformational stability of the domain

Amyloid-induction assays with ultrasonication (233-238) or gentle agitation at neutral pH with or without 0.5 mM SDS, 37 °C, were used to determine the amyloidogenic propensity of the different naturally occurring amyloidogenic mutations. Thioflavin T (ThT) fluorescence was used to monitor the fibrillation process in both assays (118,119). As expected, the MAK33 V_L-wt did not bind ThT during the entire time of the experiment (Fig. 4.3a, b). Surprisingly, both variants (MK-D17A, MK-S20N and MK-D70N) with stability like the $V_L$wt and the destabilized variant (MK-R61A) (Table 4.1), bound ThT as indicated by an increase in fluorescence after a lag time of a few minutes to hours with ultrasonication in the absence of SDS (Fig. 4.3a) or after two weeks incubation with gentle agitation in the presence of SDS (Fig. 4.3b). Conversely, MK-N31D, MK-S40L and MK-L96P with similar stability as the $V_L$wt or the destabilized MK-G57E (Table 4.1), did not show any ThT binding with ultrasonication (Fig. 4.3a). However, after a prolonged incubation of several weeks with gentle agitation in the presence of SDS, MK-N31D and MK-G57E bound ThT, whereas MK-S40L and MK-L96P did not bind ThT (Fig. 4.3b).

To confirm the presence of fibrils, transmission electron microscopy was performed. As expected, amyloid fibrils were detected in all variants for which ThT fluorescence was observed in the ultrasonication assay (Fig. 4.3c). Only amorphous aggregates could be observed in samples that did not fibrillize. Similar results were obtained with gentle agitation without ultrasonication (data not shown). The fact that on the one hand both stable (MK-S20N and MK-D70N) and destabilized (MK-R61A) variants readily fibrillized, while on the other hand both stable (MK-V_L-wt, MK-S40L, MK-L96P) and the destabilized (MK-G57E) variant were resistant to fibril formation, implies that stability is not the

only factor controlling amyloid formation. It will therefore be interesting to understand particularly why stable variants readily fibrillize, since this is contrary to the generally recognized notion that increased stability protects from amyloid formation.
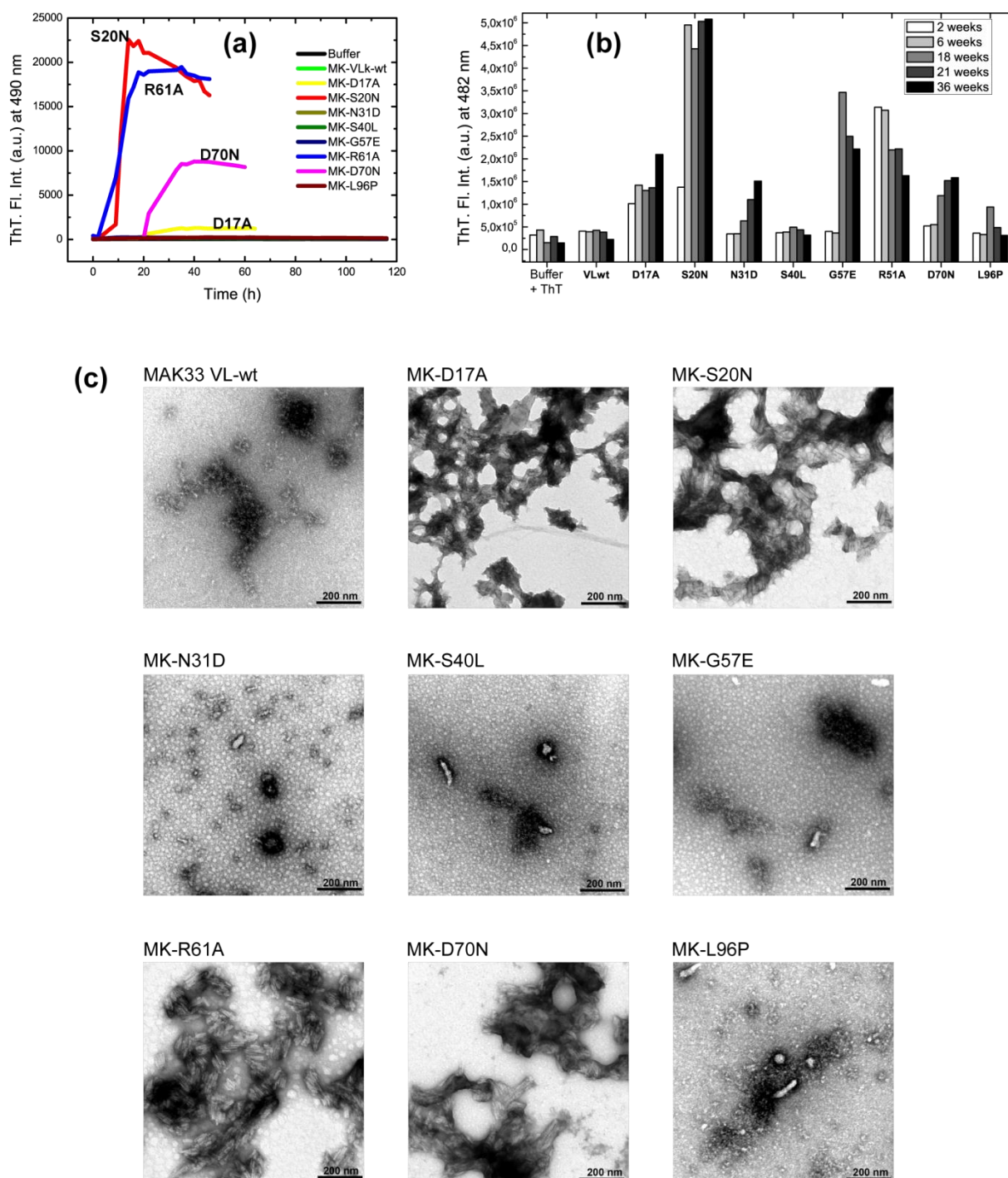
Figure 4.3. Amyloidogenicity of natural amyloidogenic MAK33 V$_L$ mutants. 30 μM of each V$_L$ variant in a PBS buffer at pH 7.4 in the presence of ThT was subjected to ultrasonic pulses at 37°C (n=4). ThT fluorescence was monitored over time for the different MAK33 V$_L$ variants (a). (b) ThT fluorescence of the V$_L$ variants in the presence of 0.5 mM SDS with incubation at 37 °C with gentle agitation (without ultrasonication). (c) TEM micrographs acquired at the end of the ultrasonication assay to detect the presence of fibrils. Fibrils are seen for variants (R61A, S20N, D70N and D17A) with a high ThT fluorescence in (a). Variants (V$_L$wt, N31D, S40L, G57E, and L96P) that did not form fibrils also did not bind ThT and their curves in (a) are superimposed.

4.3   Molecular mechanisms of V_L fibrillation

4.3.1   The folding kinetics and the exposed hydrophobic patches are not affected by amyloidogenic mutation

With the folding pathway of the non-amyloidogenic MAK33 V_L-wt elucidated (193), it was envisaged that the analysis of the folding pathway of the amyloidogenic MAK33 V_L-S20N mutant, with similar stability as the wild type, would reveal insights regarding folding kinetics and population of intermediates as to why this variant might readily fibrillize when compared to the wild type. To investigate this, refolding and unfolding kinetics of MAK33 V_L-wt and MAK33 V_L-S20N were determined at different GdmCl concentrations. The proteins were first unfolded in 2.5 M GdmCl, which was sufficient to fully denature them (data not shown). Once equilibrium had been reached, the proteins were diluted manually into a refolding buffer. For unfolding, the native proteins were diluted into different concentrations of GdmCl by manual mixing. The change in fluorescence intensity at 358 nm was monitored. Figure 4.4a shows the kinetic traces obtained for both MAK33 V_L-wt and the MAK33 V_L-S20N variant during refolding with a final GdmCl concentration of 25 mM and unfolding with a final GdmCl concentration of 2.56 M. Double exponential kinetics were observed for both variants for the refolding experiments and both variants had almost identical rate constants, with K_{obs1} of 0.0146 s^{-1} and K_{obs2} of 0.00404 s^{-1} for the MAK33 V_L-wt and K_{obs1} of 0.01399 s^{-1} and K_{obs2} of 0.00504 s^{-1} for the MAK33 V_L-S20N. This implies that both variants exhibit a fast (K_{obs1}) and a slow (K_{obs2}) refolding phase. For the unfolding of both variants, single exponential kinetics and also similar rates were observed, with K_{obs1} of 0.02101 s^{-1} and 0.02143 s^{-1} for the MAK33 V_L-wt and MAK33 V_L-S20N, respectively.

For a detailed characterization and comparison of the folding behavior of MAK33 V_L-wt and the amyloidogenic MAK33 V_L-S20N mutant, single jump refolding and unfolding kinetic experiments were performed to obtain chevron plots (Fig. 4.4b). The refolding and unfolding rates observed were obtained by manual mixing technique and these were compared to rates observed for MAK33 V_L-wt, previously published (black circles: Fig. 4.4b) (193). Interestingly, the chevron plots of both variants are very similar with almost identical rate constants for the refolding and unfolding kinetics (green and red circles: Fig. 4.4b). When compared to the chevron plot published for MAK33 V_L-wt, (black circles: Fig. 4.4b), the rate constants observed show that both MAK33 V_L-wt and the amyloidogenic MAK33 V_L-S20N have a similar folding behavior. These rates constants observed by manual mixing correspond to the λ_2 phase, which is a slow and one of three exponential phases observed during the folding of MAK33 V_L-wt (193). The rates of this phase depend on the GdmCl concentration, with higher GdmCl concentrations resulting in decreased refolding rates or increased unfolding rates due to conformational changes associated with burial or exposure of core residues, respectively. Due to the very similar refolding and unfolding behavior of MAK33 V_L-S20N compared to MAK33 V_L-wt, a conclusive statement on the direct implication of folding to its amyloidogenic propensity cannot be made.
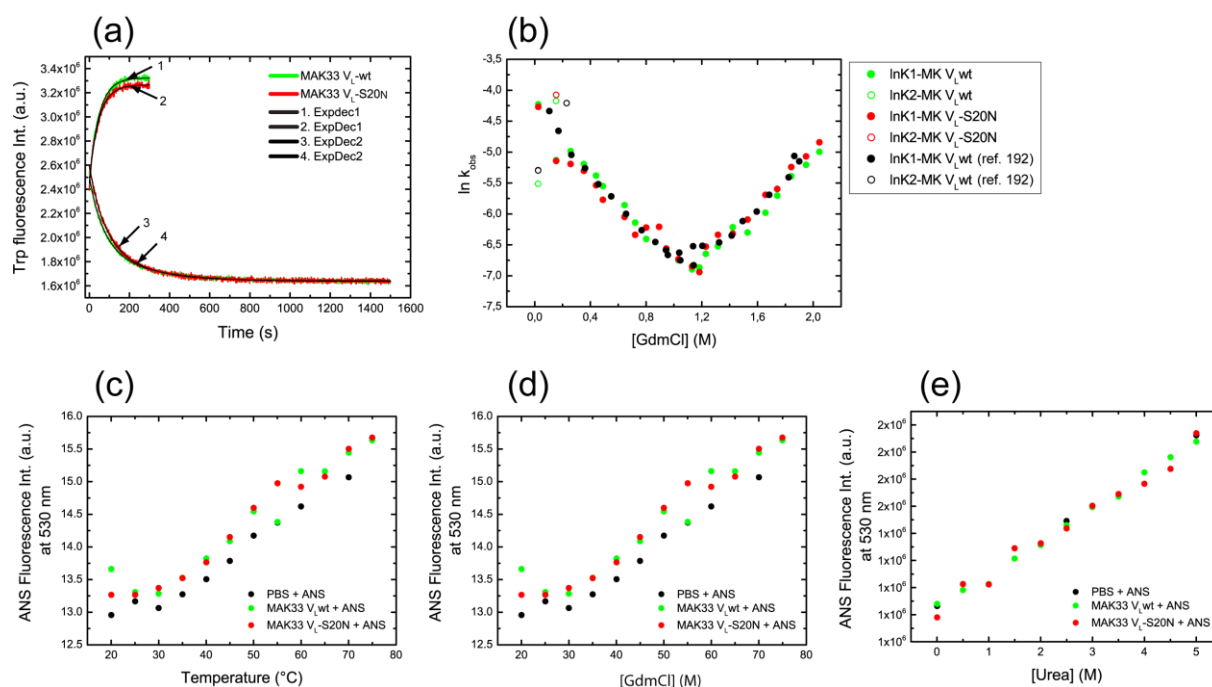
Figure 4.4. (a) Kinetic traces for unfolding and refolding of MAK33 V$_L$-wt and MAK33 V$_L$-S20N mutants, obtained by manual mixing. The final GdmCl concentration for the unfolding reaction was 2.56 M, while the refolding kinetics were observed at a GdmCl concentration of 25 mM, after diluting the native or denatured protein into a denaturing (GdmCl) or refolding buffer (PBS), respectively. The change in Trp fluorescence intensity at 358 nm was recorded. The arrows 1 and 2 (exponential decay 1) and 3 and 4 (exponential decay 2) point to the fit lines for the unfolding and refolding kinetics, respectively. (b) Chevron plot showing the variation in rates obtained from exponential fits of refolding and unfolding kinetics with GdmCl concentration. The V$_L$ domains in 0 or 2.5 M GdmCl were diluted 100-fold into an unfolding or refolding buffer with varying GdmCl concentrations, respectively, and the change in fluorescence intensity at 358 nm was monitored. Both the refolding and unfolding rates were obtained using manual mixing methods. The folding rates for MAK33 V$_L$wt (green circles) and MAK33 V$_L$-S20N (red circles) are compared with rates obtained for MAK33 V$_L$-wt (black circles), previously published by Simposon et al., (2010). ANS fluorescence recorded for MAK33 V$_L$-wt (green circles) and MAK33 V$_L$-S20N (red cicles) during thermal induced unfolding (c), GdmCl induced unfolding (d) and urea induced unfolding (e), compared with buffer (PBS: black circles).

ANS is a fluorescent hydrophobic probe for examining the non-polar character of proteins. The dye binds to clustered hydrophobic residues that are exposed to the solvent, giving rise to a blue shift and an increase in fluorescence emission intensity. This property makes ANS a versatile probe in monitoring protein folding and the detection of intermediates such as molten globules (226). To verify whether hydrophobic patches could be differentially exposed and hence the differential existence of intermediates in MAK33 V$_L$-wt and the amyloidogenic MAK33 V$_L$-S20N mutant during unfolding, both proteins were subjected to thermal- or chemical-induced unfolding in the presence of ANS. The ANS fluorescence spectra of the samples were recorded at 5 °C intervals, from 20 °C – 75 °C for the thermal induced unfolding, whereas for the chemical induced unfolding, ANS spectra of the protein variants in different concentrations of denaturant were acquired. Experiments were controlled with buffer (PBS)-only samples and the maximum ANS fluorescence emission intensity at 530 nm was compared (Fig. 4.4c-e). Surprisingly, for all modes of denaturation, the ANS binding pattern for MAK33 V$_L$-wt and MAK33 V$_L$-S20N were almost identical. Under thermal induced unfolding, only very little binding to the dye was observed when compared to the buffer (Fig. 4.4c). The fluorescence emissions of the samples including those of the buffer-only were linearly dependent on the temperature (Fig. 4.4c). Unfolding by GdmCl did not result in any significant binding of the dye to any of the V$_L$ variants when compared with the buffer (Fig. 4.4d). It was thought that this might be due electrostatic

interaction of the sulfonyl group of the ANS molecule with the GdmCl salt, which might interfere with the interaction of the dye with the protein variants. However, when urea (a non-salt denaturant) was used to denature the proteins, similar results were obtained (Fig. 4.4e). The fluorescence emission intensity was also proportional to the concentration of the chemical denaturants. The similar ANS fluorescence patterns of MAK33 V_L-wt and MAK33 V_L-S20N correlate very well with their folding kinetic behavior monitored by Trp fluorescence. The fact that ANS does not bind to any of the variants during unfolding might be an indication for the absence of clustered hydrophobic residues during the unfolding process. The lack of differences in terms of folding and exposed hydrophobic surfaces between the MAK33 V_L-wt and the amyloidogenic MAK33 V_L-S20N mutant lead to the speculation that localized events such as H-bonding within the MAK33 V_L-S20N might be responsible for its amyloidogenicity.

4.3.2   Influence of S20N and D70N substitution on V_L domain structure and dynamics

Like S20N, the D70N substitution in MAK33 V_L leads to fibril formation, but also with a conformational stability similar to the non fibrillar wild type protein. Since S20N showed no differences in terms of folding and exposed hydrophobic patches when compared to the wild type that could help explain why the mutation predisposes to amyloid formation, we assumed that the D70N substitution with the same stability would have a similar behavior, and that structural elements might be responsible for their amyloidogenicity. Therefore, to obtain a structural understanding of the influence of these mutations (S20N and D70N) on $V_L\kappa$ the domain, all three variants (MAK33 V_L-wt, MAK33 V_L-S20N and MAK33 V_L-D70N) were analyzed by solution-state NMR spectroscopy. 75 % of the non-proline residues could be assigned for the MAK33 V_L-wt and MAK33 V_L-D70N, whereas 78 % of the non-proline residues were assigned for MAK33 V_L-S20N (Fig. 4.5a). The ^{1}H-^{15}N-HSQC NMR spectra of all three variants are similar in general, which validates that the variants have comparable structures. Backbone assignment was hampered by chemical exchange dynamics. Figure 4.5b depicts chemical shift changes caused by the S20N substitution. These chemical shift perturbations are mainly observed for residues close to residue 20 in the 3D structure (Fig. 4.5c). Among the affected residues are Ser22, Ser63, Ser65, and Ser74, which are likely involved in pairwise hydrogen bonding. The structurally important residues Cys23, which forms a conserved disulfide bond with Cys88, and the adjacent Arg24, involved in a conserved salt bridge with Asp70 are also affected by the S20N substitution. Despite Ser20 or Asn20 being surface-exposed, the S20N substitution affects hydrophobic core residues (Gln6, Val13, Leu21, L47, G64, G66, L73, I75), especially Gln6, Val13, Leu21, Gly64, Leu73 and Ile75 that display considerable changes in their chemical shifts (Fig. 4.5b).

The D70N mutation also resulted in chemical shift perturbations for some residues (Fig. 4.5d). These effects are more localized and less pronounced, but otherwise similar to those induced by the S20N substitution. With the exception of Ile48 (a hydrophobic core residue) and Ile55 that are far away, all other assigned residues that are affected by the D70N substitution are very close to residue 70 in the crystal structure (Fig. 4.5e). Ser26, Ser67 and Thr69 that are likely involved in pairwise hydrogen

bonding are affected. The hydrophobic core residues Ala25 and Ile29 both located in the N-terminal part of the protein and in close proximity to residue 70 are significantly affected by the D70N mutation. Furthermore, like for the S20N mutation, the D70N substitution causes a chemical shift change for residue Cys23, which forms a conserved disulfide bond with Cys88. However, Arg24, which forms a salt bridge with Asp70 in the wild type and is also strongly affected by the S20N substitution, only displays minor changes in its backbone chemical shifts. Since its adjacent residues, Cys23 and Ala25 display very large chemical shift changes in the HSQC spectra, the D70N substitution probably leads to an altered side chain conformation of Arg24.

Although the thermodynamic stability and the overall folding mechanisms are not affected by the amyloid forming mutants (S20N and D70N), there are localized structural differences compared to the wild type protein, which may facilitate the conversion from a partially folded state to a state that is competent for fibril formation. Several of the residues strongly affected by the S20N and D70N mutations are Ser and Thr, which could be involved in local H-bonding and thus induce a conformational change to a fibrillar configuration. Moreover, N-terminal residues (Cys23, Ala25, Arg24 and Ile29) strongly affected by either mutations also displayed strong changes in the MAK33 V$_L$-I2E mutant, which significantly destabilizes the protein, leading to fibril formation (48).
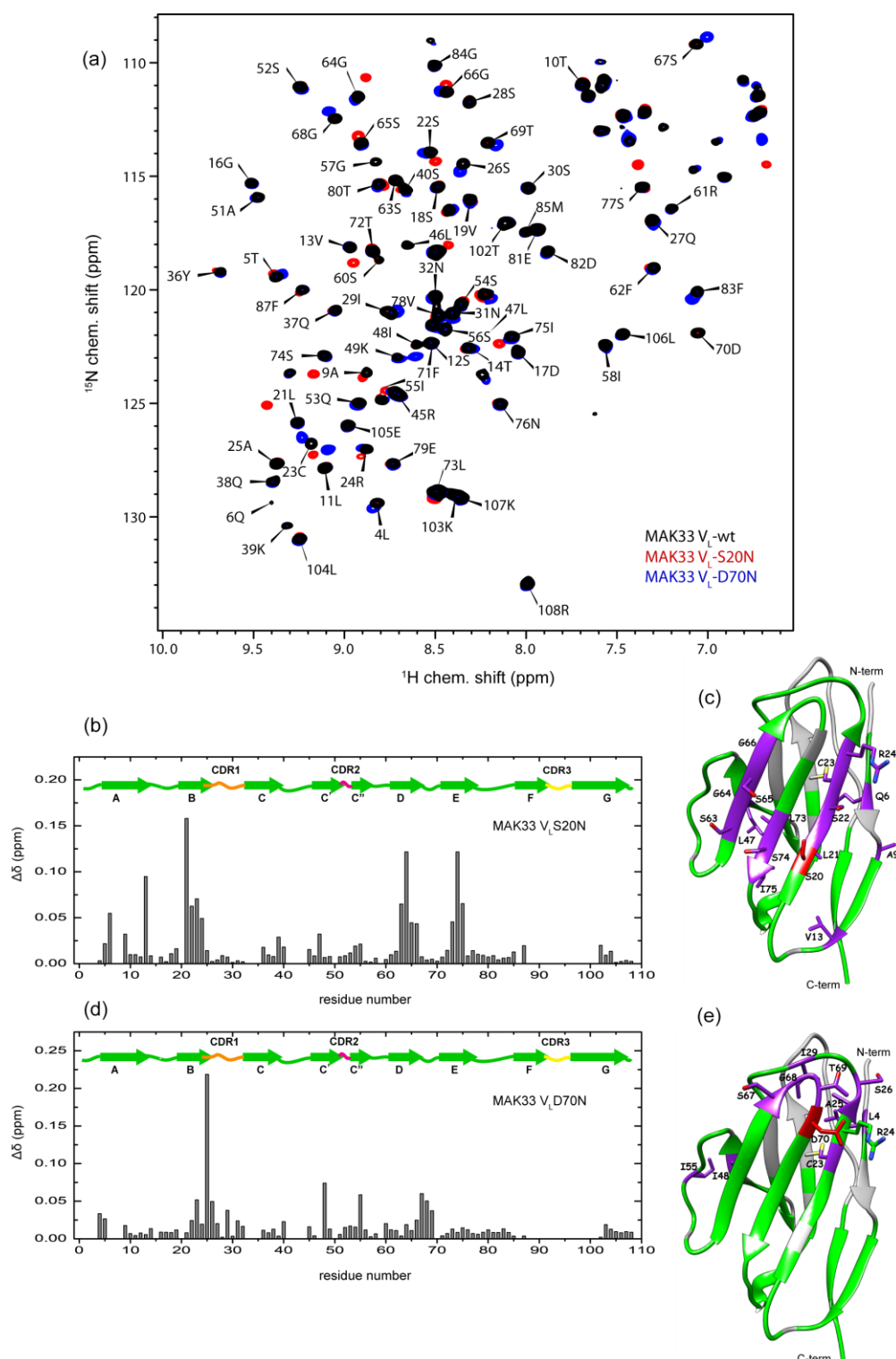
Figure 4.5. Structural properties of MAK33 V$_L$-wt, MAK33 V$_L$-S20N and MAK33 V$_L$-D70N mutant monitored by NMR spectroscopy. (a) ^{1}H-^{15}N-HSQC spectra of MAK33 V$_L$κ-wt, MAK33 V$_L$κ-S20N and MAK33 V$_L$κ-D70N. All spectra were acquired at a protein concentration of 50 μM in 20 mM phosphate, 50 mM NaCl, pH 6.5 at 298 K on a 500 MHz spectrometer equipped with a cryo-probe. Chemical shift changes of MAK33 V$_L$ caused by the S20N (b) and D70N (d) substitutions. [1]H and ^{15}N backbone chemical shifts were determined at a protein concentration of 50 μM at 298 K in 20 mM phosphate, 50 mM NaCl, pH 6.5. Effects of S20N (c) and D70N (e) substitutions in MAK33 V$_L$, strongly affected residues are marked on the MAK33 V$_L$-wt crystal structure (PDB:1FH5). S20N and D70N are shown in red, residues with chemical shift changes > 0.03 ppm in dark purple and unassigned residues in grey.

To gain further insight in the dynamics of the V$_L$ domain and to assess whether this could be influenced by the amyloidogenic mutations, molecular dynamics (MD) simulations in explicit solvent

were performed on the MAK33 V_Lκ-wt domains and on the MAK33 V_Lκ-S20N and MAK33 V_Lκ-D70N mutants. The substitutions were generated *in silico* based on the native crystal structures followed by extensive energy minimization before starting MD simulations (see Methods for details). For all three variants, stable trajectories were observed and the root-mean-square deviation (RMSD) of the backbone stayed close to the starting structure within < 2.0 Å (Fig. 4.6a). Small differences in residue fluctuations with respect to the wild type protein were seen for residues 15-20 and to a lesser degree for residues around 60 and 90 in the case of the S20N variant (Fig. 4.6b). For the D70N variant, small increases in fluctuations were observed for residues upstream of residue 70 and to a lesser extent for residues 25-30 (Fig. 4.6b). Both mutations altered the arrangement of possible side chain hydrogen bonds and other contacts at the protein surface near the mutations S20N and D70N (Fig. 4.6c). These results correlate very well with those obtained by NMR spectroscopy. The fact that Ser and Thr with large chemical shift perturbations are also seen involved in a H-bonding network around residue 20 and 70 as determined by MD simulations support the idea that localized events involving H-bonding might be largely responsible for the fibrillation of the mutants.
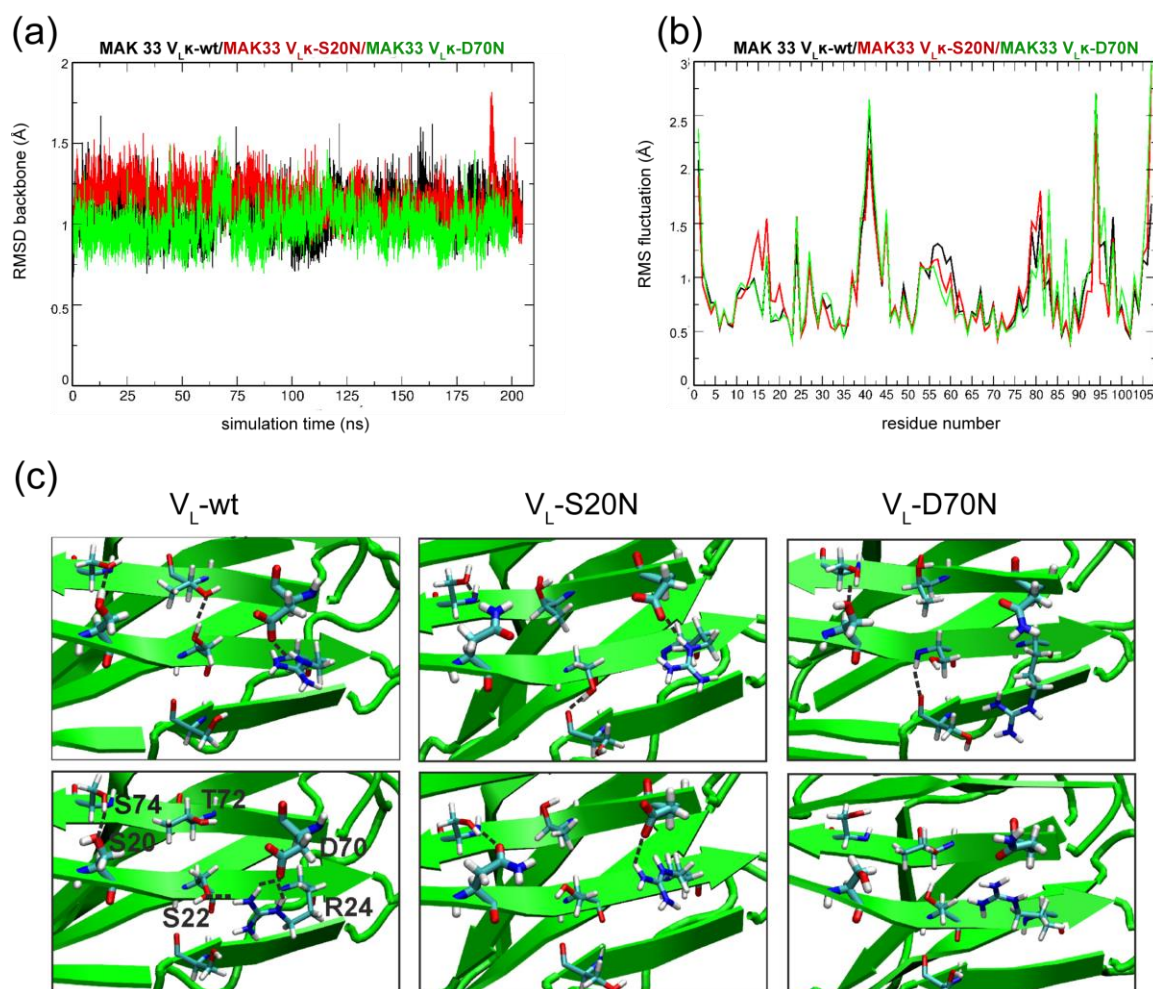
Figure 4.6. (a) The root mean square deviation (RMSD) of the protein backbone from the conformation in the experimental structure (PDB: 1F5N) were recorded during data gathering simulations (left panel) of the wild type (black line), the S20N mutation (red line) and the D70N mutation (green line). (b) The root mean square fluctuations (RMSF) indicate the atomic fluctuations of residues from the corresponding mean structures (same color coding as in the RMSD plot). (c) Typical snapshots of side chain arrangements (two for each case) around residues 20 and 70. The protein backbone is indicated as green cartoon, hydrogen bonds are shown as dashed black lines and surface side chains are labeled in the lower left snapshot.

4.4   Discussion

In addition to proteolytic fragmentation and high protein concentration, particular amino acid substitutions at specific positions within the V_L domain largely contribute to the tendency of the domain to aggregate in AL amyloidosis. There are several lines of evidence in support of this idea, however, the underlying mechanisms for why and how specific mutations induce amyloid formation remain poorly understood. MAK33 $V_L\kappa$ domain is resistant to fibril formation (48,193), but in this study we show that naturally occurring amyloidogenic mutations differentially influence the amyloidogenic propensity of the domain. None of these mutations investigated in this study had an effect on the native structure of the protein, as revealed by our CD data. This implies that although they might be amyloidogenic, they do not prevent the correct folding of the V_L to its native conformation.

While most of the natural amyloidic mutations studied here had no effect on the stability and amyloidogenicity of the MAK33 V_L domain, the R61A substitution significantly destabilized the protein and resulted in rapid fibril formation. The importance of Arg61 for the stability of the domain is due to its involvement in a key, conserved salt bridge with Asp82 located on an adjacent loop (Fig. 4.1a) (192). As reported by others (98,156), and presented by us elsewhere (48), our current findings support the view that destabilization of the V_L domain by a specific residue, including the disruption of a salt bridge with an R61A substitution, is consistent with the hypothesis that domain destabilization plays an important role in the etiology of AL. The role of the replacement of Arg61 with an uncharged residue or the disruption of the Arg61-Asp82 salt bridge resulting in destabilization and fibril formation was initially identified in a Bence-Jones protein REI (98,192). Arg61 and Asp82 are invariably encoded by all V_L germ line genes and largely conserved throughout the Ig superfamily (278). The observation of the same effects in the MAK33 V_L domain as in the REI protein as a result of the R61A mutation, suggest that this position is important for the maintenance of the structural integrity of variable domains. Therefore the replacement of the residue *per se* might not be the cause for fibrillation but rather the loss of structural integrity that follows. In addition, the R61A replacement also led to a significant exposure of hydrophobic clusters, shown by a large blue shift and an increased ANS fluorescence emission intensity. This is a likely indication for the existence of a molten globule, an intermediate rich in native-like secondary structure with fewer tertiary contacts and marginal stability, which can be prone to fibril formation (67-70). As described for other proteins (104,156,279,280), the destabilization of the native structure and the population of nonnative states is likely the primary mechanism through which the R61A mutation mediates its fibrillogenic effects (Fig. 4.7a)

Contrary to the R61A mutation, the G57E substitution that also destabilized the MAK33 V_L domain did not form fibrils, suggesting that although destabilization of the protein structure might be important, some specific interactions are also necessary for fibrillation. Surprisingly, the S20N and the D70N mutations that resulted in similar thermodynamic stability as the wild type protein readily transformed into amyloid fibrils. This is an indication that stabilization of the native fold in itself is not the sole factor that control amyloid formation, but that other elements such as specific interactions of the same kind that govern the folding and stability of globular proteins are possibly involved (281). It is likely that fibril

formation also depends to some extent on other V_L sequence elements - sequences that are directly involved in the aggregation interaction after they have been made accessible by the partial or complete unfolding of the globular protein (281). Asn residues at specific positions within the V_L domain might be particularly important in this respect. Several studies have reported the implicative role of Asn residues in the formation and stabilization of β-sheet-rich fibrillar structures derived from soluble globular proteins or Asn containing peptides derived from these proteins (104,282-286). In hereditary systemic amyloidosis due to a D76N substitution in $β_2$-microglobulin, the Asn76 amide establishes new H-bonds with surrounding residues including a nearby Thr, which has implication for the disease (287). Our structural data obtained by NMR spectroscopy and MD simulations, revealed that both the S20N and the D70N replacements in the MAK33 V_L domain result in localized structural changes that give rise to new H-bond formation. The ability to form stable or transient hydrogen bonds by the side-chains of polar residues (Asn, Gln, Ser, Thr, etc.) can have a major impact on the conformation of a protein. Although the presence of Asn at position 20 did not have any effect on the stability and folding of the V_L domain, on the one hand, it is possible that the residue induces a mild conformational change that results in a very transient intermediate state, which shifts the equilibrium from the native state to a fibril structure that is also stabilized by Asn20. On the other hand it is likely that a major conformational change due to the S20N or D70N mutation may not have taken place until after unfolding and an initial aggregation step of the V_L domain (Fig. 4.7b), such as that reported for insulin fibrillation at low pH (104,288), and the protein S6 from *Thermus thermophilus* or pathogenic variants of ataxin 3, where the rate of fibril formation were also not seen to correlate with the unfolding rate or conformational stability of the proteins (104,289,290). This leads to the proposal that the pathway for fibril formation of these V_L variants is distinct from that of unfolding (Fig. 4.7b) (104,290), and strongly supports the off-pathway mechanism of aggregation (281).

In our NMR experiments, particularly Ser and Thr were largely affected by the S20N or D70N substitution as revealed by their large chemical shift perturbations. In the MD simulations, while the same Ser and Thr residues were involved in a H-bond network in close proximity to position 20 and 70, Asn at these position (Asn20 and Asn70) also had a higher tendency for multiple H-bonds with neighboring side chains as opposed to the wild type residues (Ser20 and Asp70). As shown by several studies, electrostatic interactions of peptide backbone dipoles and a H-bond network between the Asn side-chain and adjacent residues are crucial for the stability of the cross-β amyloid structure (286,291,292). Therefore the H-bond network established by the mutations and the side-chain interactions of the Asn20 or Asn70 are possible mechanisms through which the mutations induce and stabilize the fibrillar state of the V_L protein. While it is well recognized that Gln and Asn-rich regions feature in several proteins involved in neurodegenerative diseases and amyloid-like aggregation (293) for example, Huntingtin in Huntington disease. It is also widely established that Gln and poly Asn stretches act as "dry, polar zippers", which stabilize β-sheet amyloid-like aggregates by forming interstrand H-bonds both along the backbone and between polar side-chains of adjacent strands (292,294,295). This is in good agreement with our MD simulation data, which show both Asn20 and Asn70 residues to be frequently involved in more H-bonds than their respective wild type residues,

Ser20 and Asp70. The fact that the S20N or the D70N mutation both have similar effects on the structure and fibrillogenesis of the V_L domain, implies that a similar mechanism largely dependent on the Asn residue underlies the process of fibril formation. Whether the replacements at positions 20 and 70 with Gln or residues with other side chains will result in similar structural changes and fibril formation remains to be determined. Furthermore, the investigation of the fibrillogenic propensity of short peptide fragments carrying Asn20 or Asn70 residues and structural studies on these peptides would help shed more light on the underlying mechanisms of Asn-induced V_L fibrillation.

Reports show that *N*-linked glycans that may be incorporated as a result of the acquisition of *N*-linked glycosylation sites (Asn-Xaa-Ser/Thr) through somatic hypermutations (150), are more frequent in AL proteins (157,158), and that these glycans might be responsible for their increased propensity to fibrillize (150). The S20N or the D70N mutations both introduced an *N*-linked glycosylation sites in the MAK33 V_L protein, but despite the lack of glycans at either position due to protein expression in non-glycosylating *E. coli* cells, both V_L variants readily formed fibrils. This leads to the proposal that the Asn in closed proximity to Ser/Thr, might be involved in side chain interactions that facilitate V_L fibril formation and not the glycans. This holds true based on our findings and the underlying mechanisms of Asn20- or Asn70-induced V_L fibril formation that we have proposed.

Taken together, we propose two possible underlying mechanisms for V_L fibril formation; (1) the destabilization of the native fold that results in the accumulation of nonnative structures that give rise to fibril formation (Fig. 4.7a), a widely recognized mechanism (104,106), and (2) the induction of specific native-like interactions similar to those that stabilize the native conformation, to induce and stabilize the fibril structure (Fig. 4.7b). The latter mechanism is reminiscent of the off-pathway mechanism of aggregation, which to the best of our knowledge has not been previously described for the fibrillation of V_L proteins.
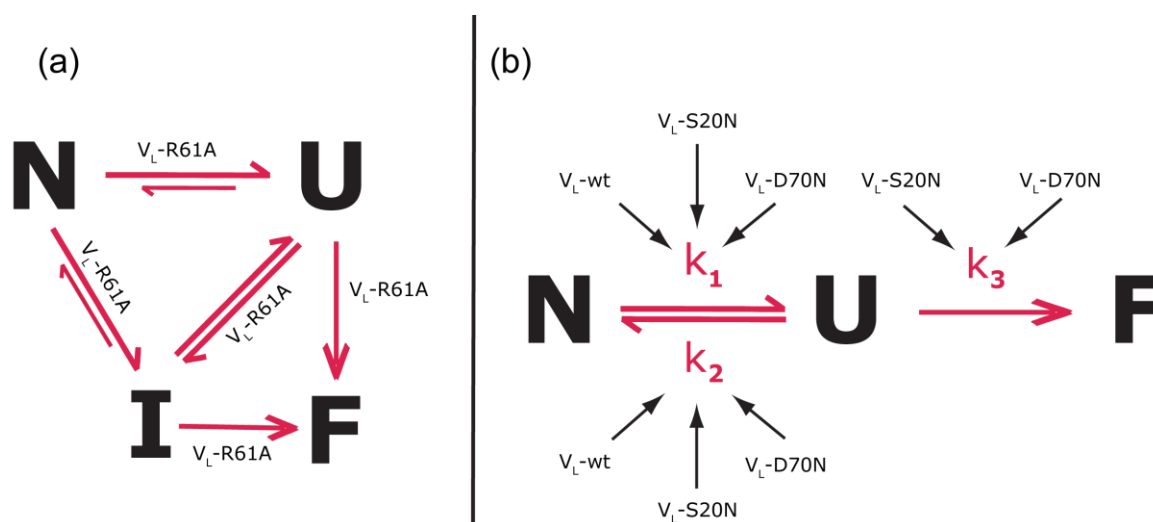
Figure 4.7 Proposed models for the underlying mechanisms of fibrillation, induced by (a) the R61A substitution in the MAK33 $V_L\kappa$ domain, where N; native state, I; intermediate state, U; unfolded state and F; fibrillar state. The R61A mutation destabilizes the protein leading to the population of intermediate and unfolded species which give rise to fibrils. (b) The S20N and D70N substitution in the MAK33 $V_L\kappa$ domain, where N; native state, U; unfolded state and F; fibrillar state. The mutations have the same unfolding (K_1) and refolding (K_2) rates like the wild type, and when the protein is unfolded, both mutation induce and stabilize (K_3) the fibrillar state, a state not induced by the wild type.

CHAPTER 5 – Role of V$_L$κ/C$_L$κ connecting-liker residues in antibody kappa light chains

5.1   Results and discussion

5.2   V$_L$/C$_L$ connecting-linker residues in antibody light chains

Immunoglobulin HCs assemble either with a κLC or a λLC, two classes of proteins encoded by different genes loci (296). Despite only a few similarities in their amino acid sequences, the Ig-fold and the domain architecture is identical for both chains (Fig. 5.1a, c). In AL, λ is represented in the ratio of 3:1 compared to healthy individuals or multiple myeloma patients (λ/κ = 1:2) (146). The reason for this occurrence is completely unknown. The identification of mostly the V$_L$ fragments of either the κLC or a λLC within aggregated deposits in AL patients (148) is an indication that common features exist between the two LCs that predispose the N-terminal V$_L$ domain to amyloid formation. The proteolytic cleavage of the C$_L$ domain (152), which contributes to the stability of the LC, and might protect it from fibrillation is possibly one key factor. In addition, since the C$_L$ domain is not involved in the deposits, it seems that structural elements exist that protect it from misfolding (209), whereas some factors within the V$_L$ domain might predispose it to the disease state.

As V$_L$s with different C-terminal length have been reported to show varying aggregation kinetics, it seems that some residues at the linker connecting the V$_L$ and the C$_L$ domain of the LC (V$_L$/C$_L$ boundary residues) might be important in this respect. In the study of Goto and Hamagushi (213), the truncation of boundary residues (N-terminus) of C$_L$λ (Fig.1a), results in a decrease in stability of ~ 2.5 °C, a decrease in the GdmCl transition midpoint of 0.2 M and a higher unfolding rate for the domain (213). However, these effects could not be assigned to a specific residue. In the context of the V$_L$, the effects of these residues at the C-terminus of the domain have not been described. Furthermore, the effects of boundary residues to the V$_L$ and C$_L$ have not been reported for the κLC (Fig. 5.1c). Identifying features of the connecting-linker that are common to the κLC and λLC families will help understand why mostly the V$_L$ appears in amyloid deposits. In addition, studying the effects of these connecting-linker residues in the context of the κLC might lead to the identification of common features that help maintain the structural integrity of antibody domains. Therefore, through the generation of various truncation variants, the effects of specific V$_L$κ/C$_L$κ-connecting-linker residues on the MAK33 V$_L$κ and the C$_L$κ domain are analyzed in terms of stability, folding and fibril formation in the following sections.
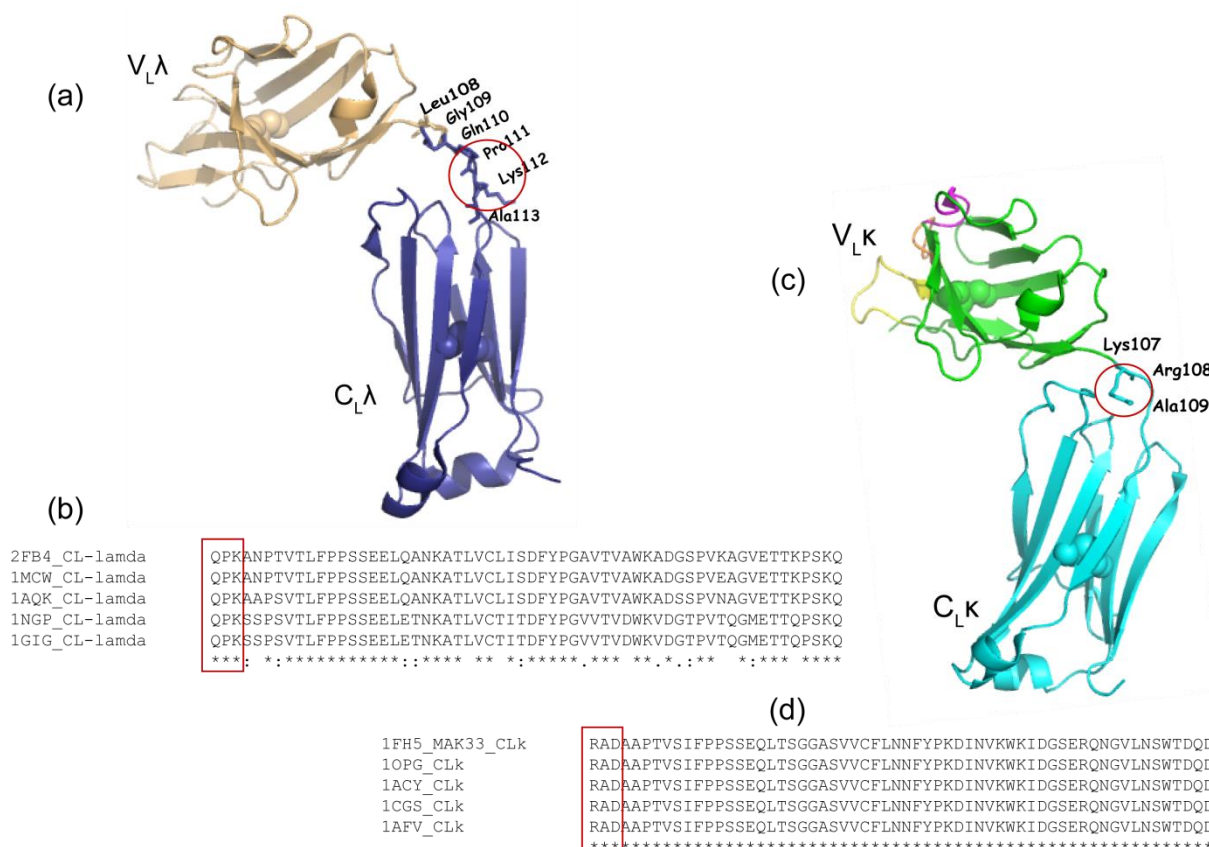
Figure 5.1. (a) 3D structure of a λLC (PDB: 2FB4), for which the role of residues 105-113 (N-term of C$_L$λ) for the folding and stability of the C$_L$λ domain has been published by Goto and Hamagushi 1987 (213). Pro111 and Lys112 marked by red circle were the two residues suggested to be important for folding and stability (213). (b) Multiple sequence alignment of C$_L$λ, irrespective of organism. Five representative sequences are shown. Only the first 60 residues are presented. All sequences show a less diverse framework. The N-terminus of C$_L$λ is a QPK motive (red rectangle). (c) 3D structure of a κLC (PDB: 1FH5). The residues that link the V$_L$κ domain and the C$_L$κ domain (boundary residues: Lys107, Arg108 and Ala109) are highlighted between the two domains. (d) Multiple sequence alignment of C$_L$κ domain, irrespective of organism. Five representative sequences are shown. Only the first 60 residues are presented. All sequences show a highly conserved framework. The N-terminus of C$_L$κ is an RAD motive (red rectangle).

## 5.3   The influence of the C-terminal Arg108 on the V$_L$κ domain

To analyze the effects of the conserved Arg108 on the V$_L$κ domain, Arg108 deletion variants of the MAK33 V$_L$κ-wt (MAK33 V$_L$wt-ΔR108) and of mutated variants of MAK33 V$_L$ (MAK33 V$_L$-I2EΔR108, MAK33 V$_L$-I2E/T102SΔR108), were generated (Fig. 5.2a). The mutated MAK33 V$_L$κ variants were already investigated within this thesis as reported in Chapter 3. All V$_L$κ variants with and without the C-term Arg108 were expressed in *E. coli* cells, purified and spectroscopically characterized. All variants were correctly folded as determined by far UV (Fig. 5.2b) and near UV (Fig. 5.2c) CD spectroscopy. When the proteins were investigated for ANS binding, there were no significant differences in ANS fluorescence emission between the V$_L$κ variants that carried the Arg108 and those that lacked the residue (Fig. 5.2d). This indicates that the presence or absence Arg108 does not induce a conformational change in the V$_L$κ domain that could result in the exposure of hydrophobic clusters.
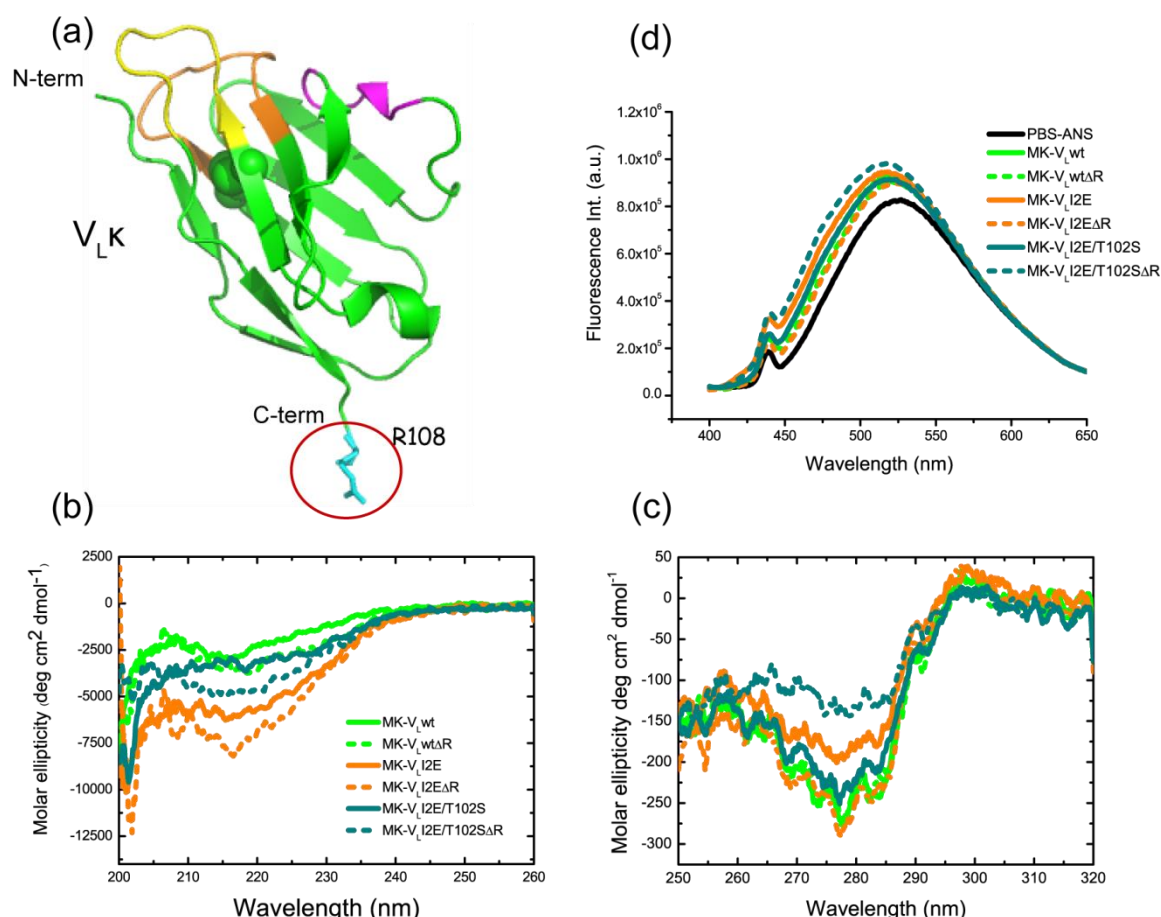
Figure 5.2. (a) The 3D structure of MAK33 $V_L\kappa$ domain with the C-terminal Arg108 highlighted. The CDRs are highlighted in orange (CDR1), pink (CDR2) and yellow (CDR). Also shown is the disulfide bridge buried in the hydrophobic core. Structure based on PDB code 1FH5. Numbering is based on Kabat http://www.bioinf.org.uk/ or http://vbase.mrc-cpe.cam.ac.uk/. (b) Far UV and (c) Near UV CD spectra of native MAK33 $V_L\kappa$ and MAK33 $V_L\kappa$-ΔR108 variants in PBS buffer. (d) ANS fluorescence spectra of native protein variants in PBS buffer.

5.3.1   C-terminal Arg108 affects the stability of $V_L\kappa$ domain

The effects of Arg108 on the stability of $V_L\kappa$ variants were assessed spectroscopically. Interestingly, all $V_L\kappa$ variants that lacked Arg108 were destabilized (Table 5.1). Specifically, the Arg108-deleted $V_L\kappa$ variants were ~ 3 °c less stable with a free energy of ~ 2 kJ mol^{-1} less than their counterparts carrying Arg108, as determined by thermal- and GdmCl-induced unfolding transitions, respectively (Fig. 5.3; Table 5. 1)
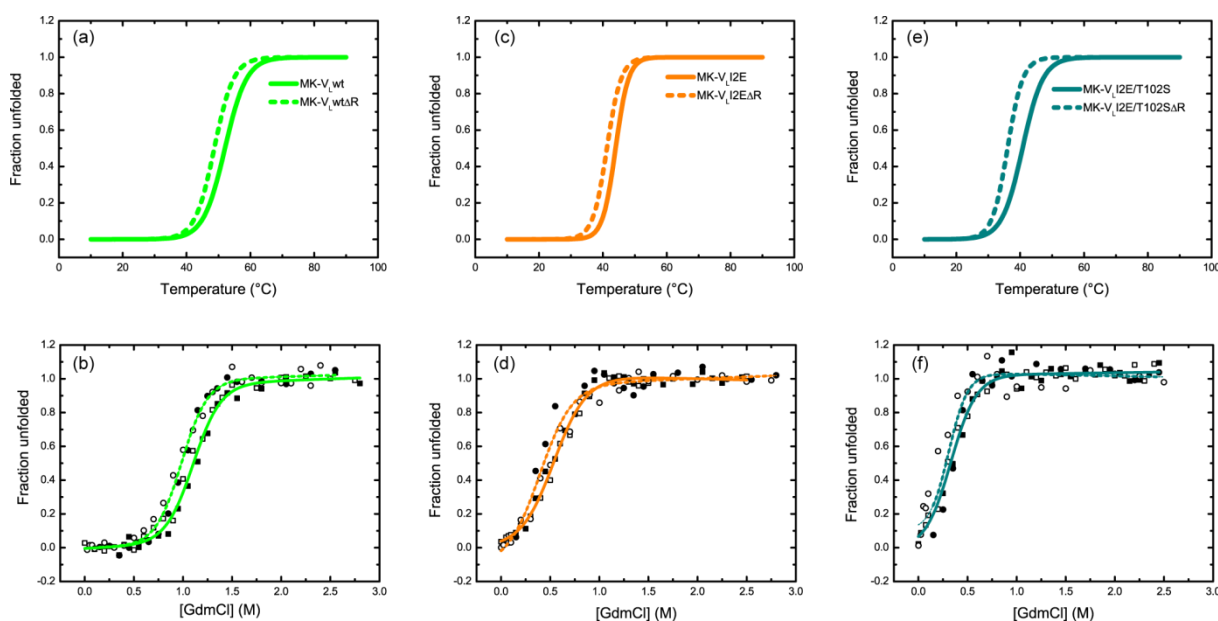
Figure 5.3. Stability of the different MAK33 V$_L$κ and C-term Arg108 deletion variants. (a, c and e) Thermal unfolding transitions of MAK33 V$_L$κ and V$_L$κ-ΔR108 variants, shown are the solid lines that indicate the theoretical curves derived by fitting the data to a Boltzmann function to obtain transition midpoints (*T*$_{melt}$). (b, d and f) GdmCl-induced unfolding transitions of MAK33 V$_L$κ and V$_L$κ-ΔR108 variants monitored by Trp fluorescence at 358 nm, the reversibility of the unfolding process is shown by the overlay of unfolding symbols (open circles) and refolding (closed circles) experiments. The solid lines show the fit to a two-state mechanism for V$_L$κ variants to obtain thermodynamic stability values (Δ*G*$_U$) and the cooperativity parameters (*m-values*), for a qualitative comparison of the data.

This stability data obtained here implies that Arg108 is important for stabilizing V$_L$κ domains and its deletion might influence the folding and amyloidogenicity of the domain.

Table 5. 1. Thermal and chemical stabilities of the different MAK33 V$_L$κ and C-term Arg108 deletion variants.
Stabilities against the thermal and chemical (GdmCl) denaturation of different variants. Midpoints of thermal transitions are shown as *T*$_{melt}$. Since GdmCl-induced unfolding transitions were reversible, the data were fitted to a two-state equilibrium unfolding model to obtain the thermodynamic stability of unfolding (ΔG$_U$) as well as the cooperativity parameter (m-value), for a qualitative comparison of the data.

Protein Variant	*T*$_{melt}$ (°C)	Δ*G*$_U$ (kJ mol^{-1})	*m*-value (kJ mol^{-1} M^{-1})
MAK33 V$_L$κ-wt	52.0 ± 0.5	-19.4 ± 3.0	17.5 ± 2.6
MAK33 V$_L$κ-wtΔR	47.7 ± 0.2	-17.8 ± 2.6	18.0 ± 2.4
MK-I2E	41.3 ± 0.4	-8.0 ± 0.7	14.8 ± 1.2
MK-I2EΔR	38.2 ± 0.1	- 5.8 ± 1.6	15.0 ± 3.0
MK -I2E/T102S	36.6 ± 0.4	-6.5 ± 1.0	19.8 ± 2,3
MK -I2E/T102SΔR	33.6 ± 0.2	- 8.0 ± 3.0	19.8 ± 2.3

5.3.2   The lack of C-terminal Arg108 predisposes V$_L$κ domains to amyloid formation

Stability is known to be an important factor that controls amyloid formation of V$_L$ proteins (48,98). To examine whether the decrease in stability caused by the truncation of C-terminal Arg108 induces amyloid formation, we set up ultrasonication amyloid-induction assays at neutral pH, 37 °C with or without 0.5 mM SDS, for MAK33 V$_L$κ-wt and the MAK33 V$_L$-ΔR108 variants (233-238). Thioflavin T

(ThT) fluorescence was used to monitor the fibrillation process (118,119). Interestingly, none of the variants bound ThT at neutral pH in the absence of SDS (Fig. 5.4a, c). But in the presence of SDS, an amyloid-inducing detergent (182,183,297), the MAK33 V$_L$-ΔR108 variant bound ThT as indicated by an increase in fluorescence after a lag time of several hours whereas the more stable MAK33 V$_L$κ-wt showed no ThT fluorescence during the experimental time (Fig. 5.4c). The large fluctuation in ThT fluorescence amplitude is likely due to differences in higher order structure of amyloid fibrils as ThT fluorescence can change depending on the morphology of amyloid fibrils (234-236,264). These variation in ThT fluorescence amplitude might also depend on the protein and it is worth noting that ThT fluorescence is not always proportional to the amount of amyloid fibrils (264).

To confirm the presence of fibrils, transmission electron microscopy was performed. As expected, amyloid fibrils were only detected in the MAK33 V$_L$-ΔR108 variant for which ThT fluorescence was observed (Fig. 5b, d)**.**
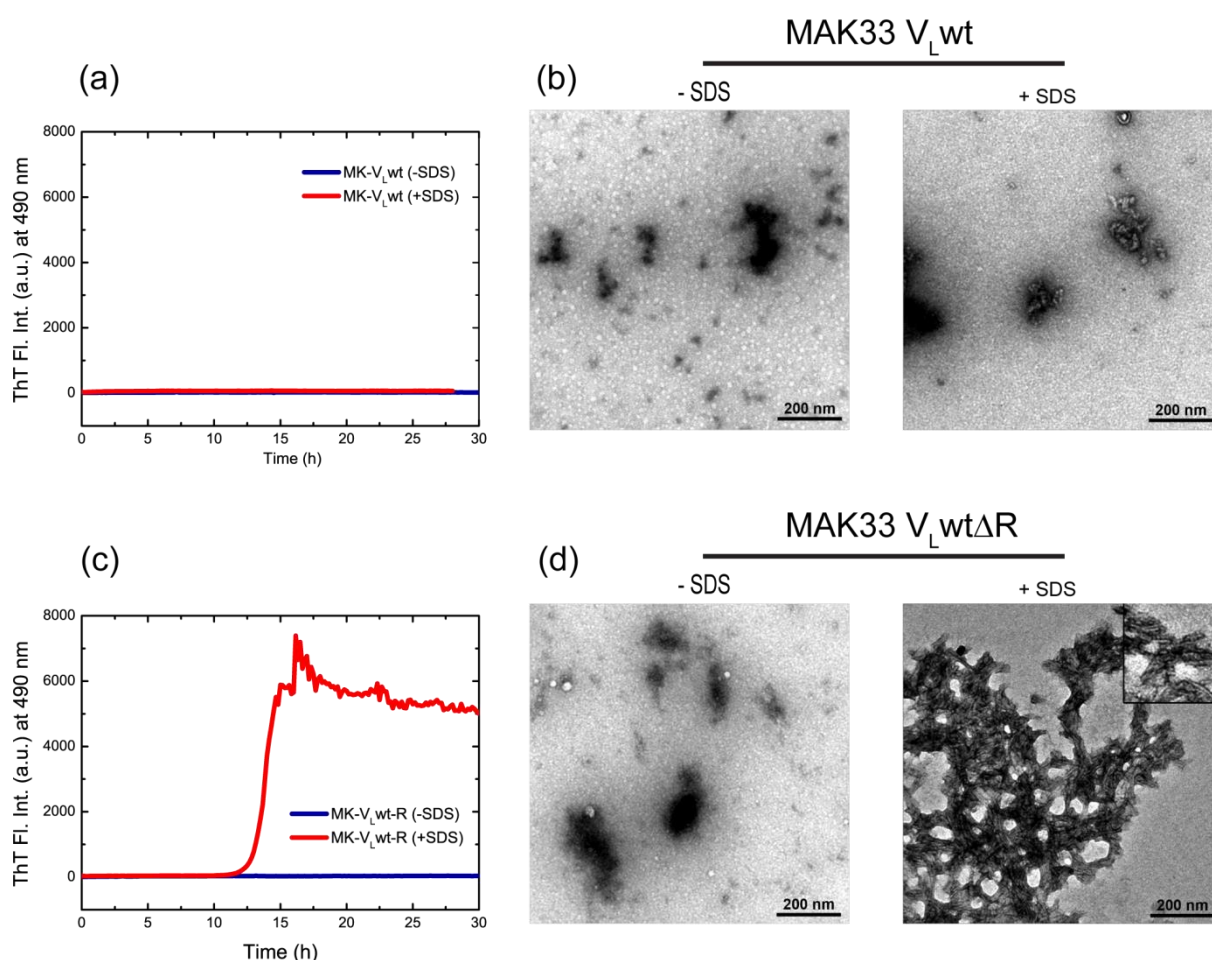
Figure 5.4. Amyloidogenicity of MAK33 V$_L$κ and MAK33 V$_L$κ-ΔR108. 30 µM of each V$_L$ variant in a PBS buffer at pH 7.4 in the presence of ThT, with or without 0.5 mM SDS was subjected to ultrasonic pulses at 37°C. (a and c) ThT fluorescence was monitored over time for the variants. (b and d) TEM micrographs acquired at the end of the ultrasonication assay to detect the presence of fibrils. Fibrils are seen for MAK33 VLκ-ΔR108 variant in the presence of SDS, with a high ThT fluorescence in (c).

5.3.3   C-terminal Arg108 influences the folding of $V_L\kappa$ domains

To assess the role of C-terminal Arg108 on the folding of the $V_L\kappa$ domain, refolding and unfolding kinetics of MAK33 $V_L\kappa$-wt and MAK33 $V_L\kappa$-ΔR108 variants were determined. For the refolding experiment, the proteins were first unfolded in 3.0 M GdmCl, which was sufficient to fully denature them (data not shown). Once equilibrium had been reached, the proteins were diluted manually into a refolding buffer. For unfolding, the native proteins were diluted into a denaturing buffer containing GdmCl by manual mixing. The change in fluorescence intensity at 358 nm was monitored. Figure 5.5-left shows the kinetic traces obtained for both MAK33 $V_L\kappa$-wt and MAK33 $V_L\kappa$-ΔR108 variant during refolding with a final GdmCl concentration of 0.6 M and unfolding with a final GdmCl concentration of 2.04 M. Double exponential kinetics were observed for both variants for the refolding experiments and both variants had very similar rate constants, with K_{obs1} of 0.00238 s^{-1} and K_{obs2} of 0.01255 s^{-1} for the MAK33 $V_L\kappa$-wt and K_{obs1} of 0.00226 s^{-1} and K_{obs2} of 0.0102 s^{-1} for the MAK33 $V_L\kappa$-ΔR108. This implies that both variants exhibit a fast (K_{obs1}) and a slow (K_{obs2}) refolding phase and the refolding rate is not affected by the Arg108. For the unfolding of both variants, single exponential kinetic with different rates were observed, with K_{obs1} of 0.00828 s^{-1} and 0.01076 s^{-1} for the MAK33 V_L-wt and MAK33 $V_L\kappa$-ΔR108, respectively. The plots (1, 2, 3 and 4) to the right are residuals for the single and double exponential fits to demonstrate the fit quality. These unfolding rates observed indicate that the deletion of the C-terminal Arg108 leads to a faster unfolding of the $V_L\kappa$ domain, which fits well to its stabilizing character and might explain why the $V_L\kappa$ variant lacking this residue readily fibrillize.
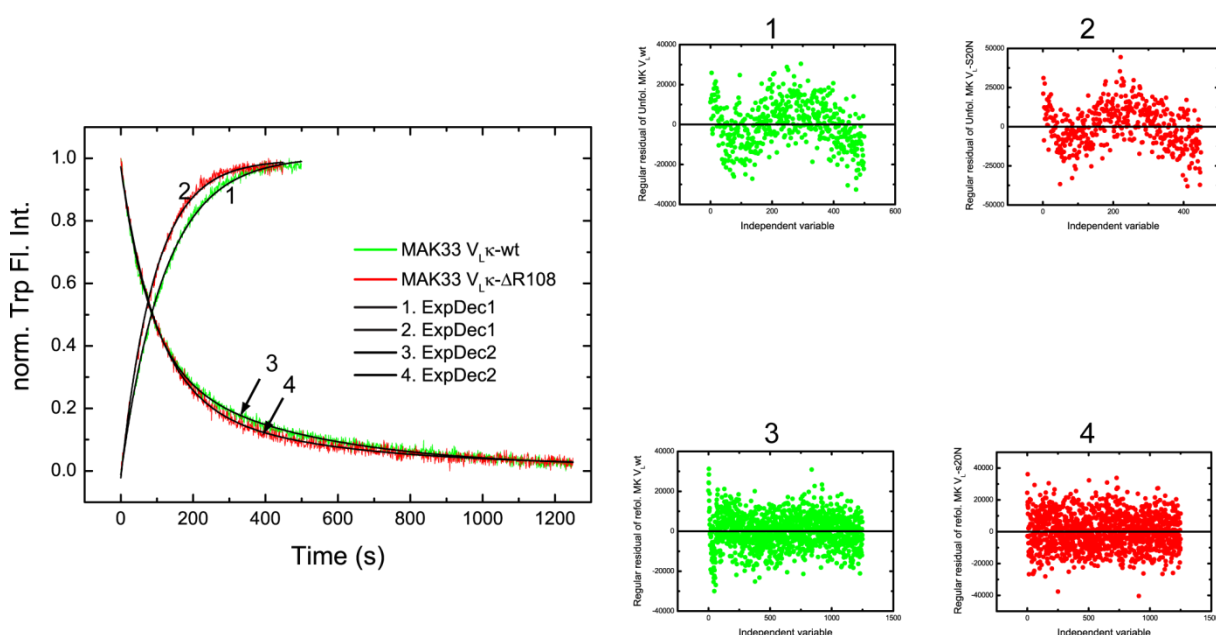
Figure 5.5. (a) Kinetic traces for unfolding and refolding of MAK33 V_L-wt and MAK33 V_L-ΔR108 mutants, obtained by manual mixing. The final GdmCl concentration for the unfolding reaction was 2.04 M, while the refolding kinetics were observed at a GdmCl concentration of 0.64 M. The change in Trp fluorescence intensity at 358 nm was recorded. The numbers 1 and 2 (exponential decay 1) and 3 and 4 (exponential decay 2) are the fit lines for the unfolding and refolding kinetics, respectively. Residual for the single (1 and 2) and double (3 and 4) exponential fits are depicted in plots, green dots (MAK33 V_L-wt) and red dots (MAK33 V_L-S20N), to demonstrate the fit quality.

5.4   The influence of N-terminal Arg108 on the C$_L$κ domain

Based on the annotation of LC sequences in the protein data bank (PDB), the conserved Arg108 is the first (N-terminal) residue of C$_L$κ domains (R108-E214), according to the Kabat numbering scheme (Fig. 5.1b, c). The cloned C$_L$ domain of the MAK33 κLC (PDB: 1FH5) previously published, only contains residues A112-E214 and in addition includes four N-terminal residues (GlySerHisMet) derived from the pET28a(+) vector after Thrombin cleavage of the His-tag during purification (Fig. 5.6b) (56). To ascertain the role of N-terminal Arg108 on the C$_L$κ domain, the C$_L$ domain of MAK33 κLC was recloned and its natural sequence (R108-E214) that starts with Arg108 generated (Fig. 5.6b). In addition, two N-terminal deletion variants C$_L$κ-ΔR108 (A109-E214) and C$_L$κ-ΔR108/A109 (D110-E214) were also generated (Fig. 5.6b). The three variants were expressed in *E. coli* cells, purified and biophysically characterized. All variants were correctly folded as determined by far UV (Fig. 5.6c) and near UV (Fig. 5.6d) CD spectroscopy. The far-UV CD spectra of all C$_L$κ variants were similar with a minimum at 218 nm (Fig. 5.6c) characteristic of the β-sheet conformation. Their near-UV CD spectra (Fig. 5.1d) with a minimum at 280 nm confirmed that the tertiary structure of all protein variants is correctly formed (71,73). The analysis of the oligomeric state of the C$_L$κ-wt and its N-terminal deletion variants by analytical ultracentrifugation sedimentation velocity runs revealed that all variants were monomeric with a sedimentation coefficient of ~ 14 S (Fig. 5.6e).
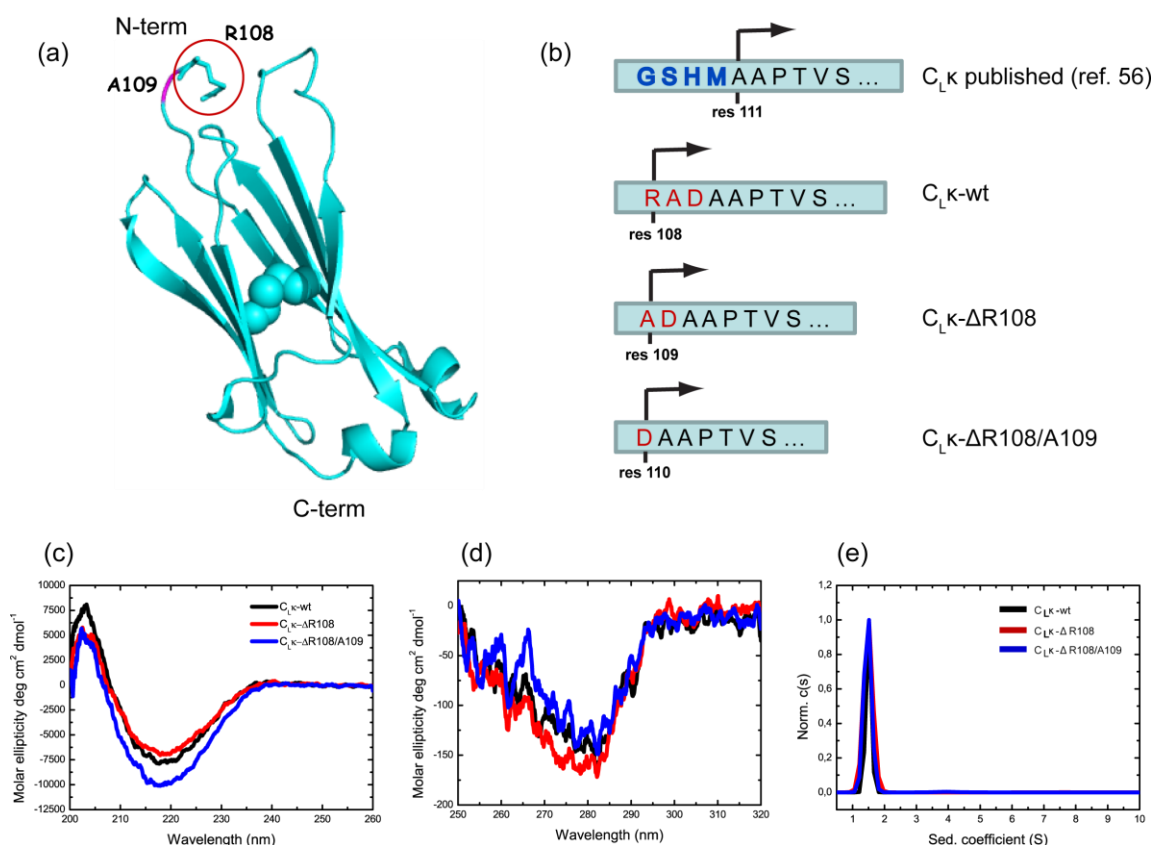
Figure 5.6. (a) The 3D structure of MAK33 C$_L$κ domain with the N-terminal Arg108/Ala109 highlighted. The disulfide bridge buried in the hydrophobic core is shown in van der Waals spheres. Structure based on PDB code 1FH5. Numbering is based on Kabat http://www.bioinf.org.uk/ or http://vbase.mrc-cpe.cam.ac.uk/. (b) The different C$_L$κ constructs generated. (c) Far UV and (d) Near UV CD spectra of native MAK33 C$_L$κ-wt**,** C$_L$κ-ΔR108 and C$_L$κ-ΔR108/A109 variants in PBS buffer. (e) AUC sedimentation velocity runs, data was analyzed using the continuous c(S) distribution mode of SEDFIT (241,242).

5.4.1   N-terminal Arg108 is important for the conformational stability of C$_L$κ domain

The presence of a buried tryptophan in close proximity to the disulfide in antibody domains make it a sensitive probe for studying conformational changes by fluorescence. Intrinsic tryptophan fluorescence spectra of the C$_L$κ variants showed all proteins to have only a very small increase in fluorescence emission intensity with a red shifted emission maxing of ~ 10 nm (355 nm – 365 nm), in the presence of 2.5 M GdmCl (Fig. 5.7d –c). This suggests a partial exposure of Trp in the native states and a change in the environment of the Trp during unfolding of the C$_L$κ-wt and its N-terminal deletion variants. Though differences were not observed in the Trp fluorescence emission for the different C$_L$κ variants, we were wondering whether Arg108 could have an effect on the conformation, which might lead to the differential exposure of hydrophobic patches on the surface. Therefore, ANS binding assays were performed. Interestingly, both C$_L$κ variants that lacked Arg108 (C$_L$κ-ΔR108 and C$_L$κ-ΔR108/A109) bound ANS with the same amplitude and a small blue-shifted fluorescence emission maximum, whereas the C$_L$κ-wt only showed a small increase in ANS fluorescence (Fig. 5.7d). This suggests that the lack of Arg108 and not Ala109 induces a conformational change within the C$_L$κ domain that leads to clustering and the surface exposure of hydrophobic residues.
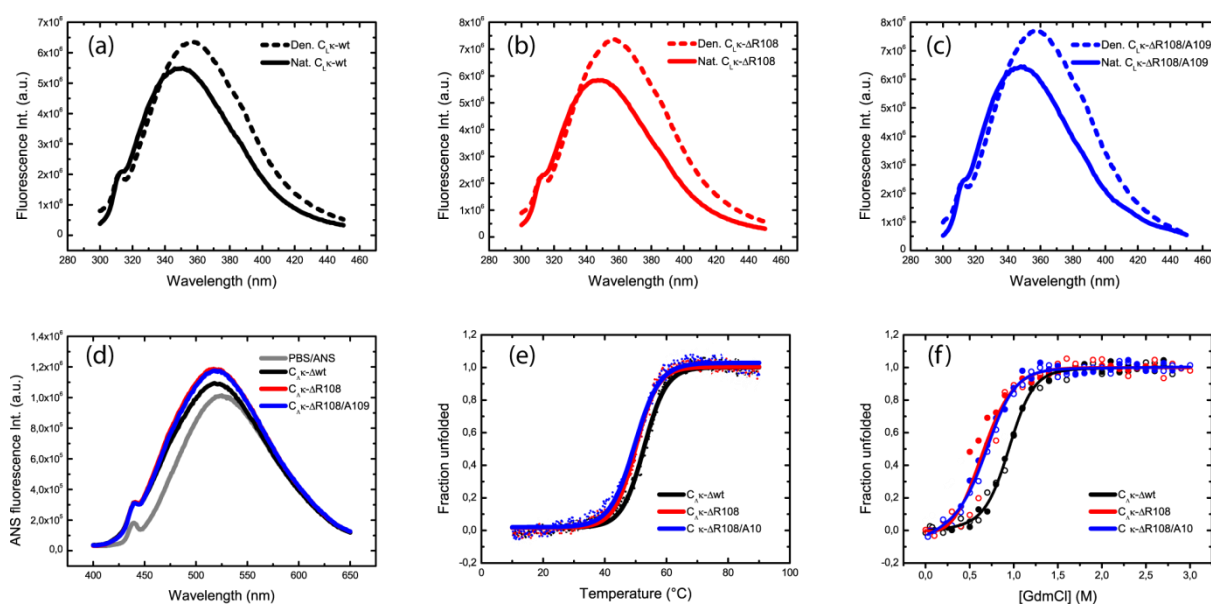
Figure 5.7. Spectroscopic characterization of MAK33 C$_L$κ and N-term Arg108/Ala109 deletion variants. (a-c) Intrinsic tryptophan fluorescence spectra of native (continuous line) and 2.5 M GdmCl-denatured (dotted lines) MAK33 C$_L$κ variants in PBS buffer. (d) ANS fluorescence spectra of native protein variants in PBS buffer. (e) Thermal unfolding transitions of MAK33 C$_L$κ variants, the solid lines indicate the theoretical curves derived by fitting the data to a Boltzmann function to obtain transition midpoints (T_{melt}). (f) GdmCl-induced unfolding transitions of the different C$_L$κ variants monitored by far UV CD by recording the change in ellipticity at 220 nm, the reversibility of the unfolding process is shown by the overlay of unfolding symbols (open circles) and refolding (closed circles) experiments. The solid lines show the fit to a two-state mechanism for V$_L$ variants to obtain thermodynamic stability values (ΔG_U) and the cooperativity parameters (*m-values*), for a qualitative comparison of the data.

To determine whether Arg108 affects the conformational stability of the C$_L$κ, thermal-induced unfolding transitions of the different variants were monitored by a change in ellipticity at 205 nm. Both C$_L$κ-ΔR108 and C$_L$κ-ΔR108/A109 were destabilized by ~ 3 °C when compared to the C$_L$κ-wt protein (Fig. 5.7e; Table 5.2). Similar results were obtained from GdmCl-induced unfolding transitions monitored by a change in ellipticity at 220 nm, with the deletion of Arg108 resulting in a decrease in free energy of ~ 5 kJ mol^{-1} (Fig. 5.7e; Table 5.2). The similarity in stability values of C$_L$κ-ΔR108 and C$_L$κ-ΔR108/A109

indicates that the N-terminal Arg108 and not Ala109 is important for the conformational stability of C$_L$κ domains.

**Table 5. 2. Thermal and chemical stabilities of the different MAK33 C$_L$κ and N-term Arg108/Ala109 deletion variants.**
Stabilities against the thermal and chemical (GdmCl) denaturation of different variants. Midpoints of thermal transitions are shown as T_{melt}. Since GdmCl-induced unfolding transitions were reversible, the data were fitted to a two-state equilibrium unfolding model to obtain the thermodynamic stability of unfolding (ΔG_U) as well as the cooperativity parameter (m-value), for a qualitative comparison of the data.

Protein Variant	T_{melt} (°C)	ΔG_U (kJ mol^{-1})	m-value (kJ mol^{-1} M^{-1})
C$_L$κ-wt	52.7 ± 0.1	- 16.8 ± 1.7	17.6 ± 1.7
C$_L$κ-ΔR108	50.0 ± 0.1	- 7.5 ± 3.2	12.9 ± 2.8
C$_L$κ-ΔR108/A109	49.7 ± 0.1	- 11.2 ± 2.0	15.3 ± 2.0

5.4.2   N-terminal Arg108 influences the folding of C$_L$κ domains

With the N-terminal Arg108 identified to be important for the stability of C$_L$κ domains, we were interested in whether it also affects its folding rates. Therefore, refolding kinetics of MAK33 C$_L$κ-wt and of ΔR108 and ΔR108/A109 variants were determined. For the refolding experiments, the proteins were first unfolded in 2.0 M GdmCl, which was sufficient to fully denature them (data not shown). Once equilibrium had been reached, the proteins were diluted manually into a refolding buffer by manual mixing. The change in fluorescence intensity at 370 nm was monitored. Figure 5.8-left shows the kinetic traces obtained for both C$_L$κ-wt, C$_L$κ-ΔR108 and C$_L$κ-ΔR108/A109 variants during refolding with a final GdmCl concentration of 20 mM. While a single exponential kinetic was observed for the C$_L$κ-wt domain, both variants lacking Arg108 refolded with double exponential kinetics (Fig. 5.8; Table 5.3). The C$_L$κ-wt protein refolded much faster with a $K_{obs1\ of}$ 0.04457 s^{-1} than C$_L$κ-ΔR108, K_{obs1} of 0.01983 s^{-1} and C$_L$κ-ΔR108/A109 with a K_{obs1} of 0.03208 s^{-1}. In addition, both variants lacking the Arg108 also had a second slow refolding phase with the same rate constants, K_{obs2} of 0.1 s^{-1} (Table 5.3). Since C$_L$κ-ΔR108 and C$_L$κ-ΔR108/A109 had similar slow refolding rates that differed from that of the C$_L$κ wild type which refolded much faster, it implies that N-terminal Arg108 is important in the productive folding of C$_L$κ domains, which is a good explanation for why the residue also stabilizes the domain.
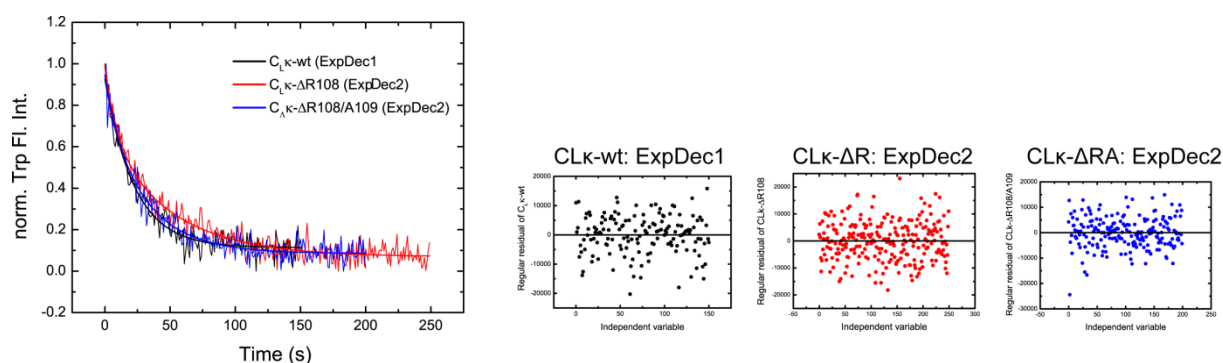
Figure 5.8. Kinetic traces for the refolding of MAK33 C$_L$κ-wt and N-term deletion mutants, obtained by manual mixing. The final GdmCl concentration for the refolding kinetics was 20 mM. The change in Trp fluorescence intensity at 370 nm was recorded. The solid lines represent fits for the refolding kinetics, single exponential decay for the C$_L$κ-wt and double exponential decays for both C$_L$κ-ΔR108 and C$_L$κ-ΔR108/A109. Residual for the single and double exponential fits are depicted in plots to the right, black dots (C$_L$κ-wt), red dots (C$_L$κ-ΔR108) and blue dots (C$_L$κ-ΔR108/A109), to demonstrate the fit quality.

Table 5. 3. Refolding rate constants for the refolding of C$_L$κ variants with a final GdmCl of 20 mM

Protein Variant	k_1	lnk_1	k_2	lnk_2
MK C$_L$κ-wt	0.04457	-3.11069		
MK C$_L$κ-ΔR108	0.01983	-3.92077	0.10554	-2.24868
MK C$_L$κ-ΔR108/A109	0.03208	-3.4394	0.10984	-2.20876

5.5   The interaction of Arg108 within the κLC

Haven determined that Arg108 is important for the stability and folding of both the V$_L$κ and the C$_L$κ domain, we wondered if this residue could be interacting with residues in both domains. A detailed analysis of the interacting partners of Arg108 within the crystal structure of several κLC was performed with the use of the Pymol software. Figure 5.9 depicts the interaction partners of Arg108 as seen in MAK33 κLC. Although it also stabilized and influenced the folding of the V$_L$κ protein, Arg108 only interacts with residues within the C$_L$κ domain, were it is the authentic N-terminal residue. Whether Arg108 affects the dynamic and thus the stability and folding of the V$_L$κ domain is still to be determined. Within the C$_L$κ domain, Arg108 interacts with Ala109, Tyr140 and Asp170 via side chain H-bonds. The interaction with Asp170 is conserved and might be a good explanation for the stabilizing character of Arg108. The dynamics of the domain might also be affected by this residue. Similar interacting residues were seen for other C$_L$κ domains (data not shown). The effects of this N-terminal Arg108 to the C$_L$κ domain are similar to those identified for the N-terminal residue of the C$_L$λ domain (213), which implies that the effects of these residues are conserved to both C$_L$ families. These results are in line with our previously published findings on the role of N-terminal residues for the integrity and stability of V$_L$κ domains (48), which suggest that the N-termini of antibody domains are crucial for their integrity.
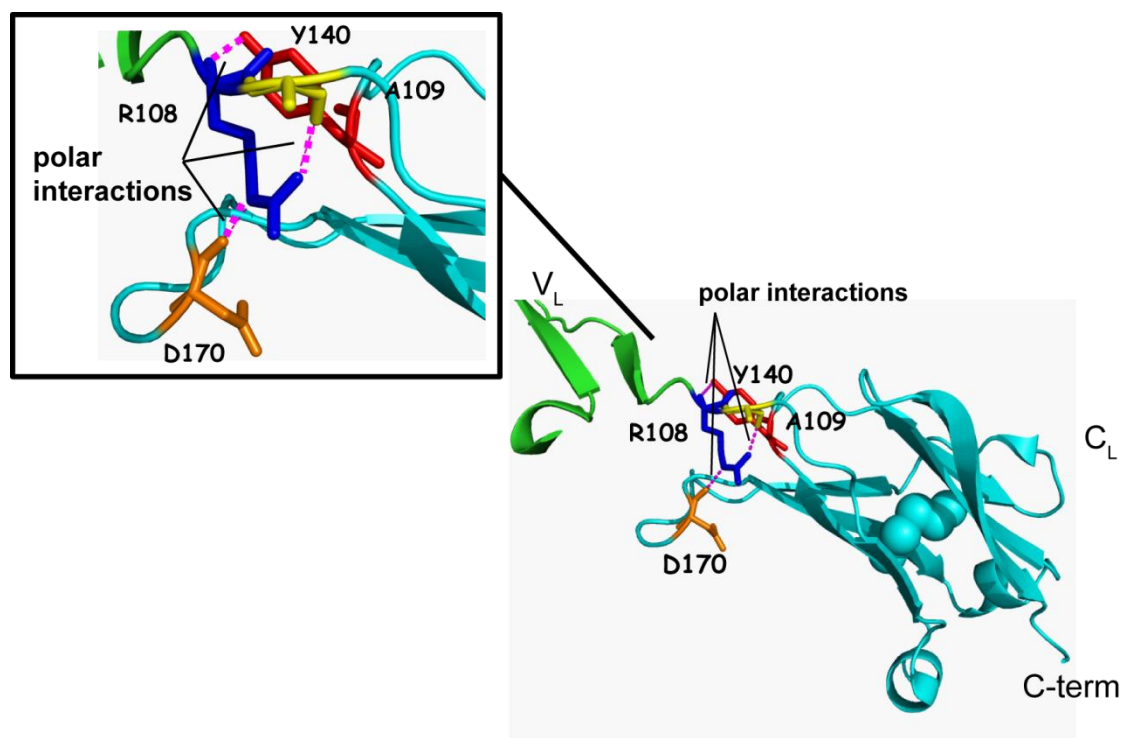
Figure 5.9. Interaction partners of Arg108, based on the crystal structure of MAK33 κLC (PDB: 1FH5). Arg108 does not interact with any residue within the V$_L$κ domain. It interacts with Ala109 and Tyr140 through H-bonding (pink dashed lines) and via a conserved salt bridge (pink dash line) with Asp170 within the C$_L$κ.

5.6   Discussion

Here we have described the role of Arg108, a linker residue between the V$_L$κ and C$_L$κ domain of the κLC, in terms of folding, stability and amyloidogenicity. The conformation of V$_L$κ and C$_L$κ wild type and their respective variants lacking Arg108 were very similar on the basis of CD and tryptophan fluorescence spectra, indicating that Arg108 is not essential for maintaining the global conformation of the V$_L$κ or C$_L$κ domain. The measurements of unfolding equilibria with heat and GdmCl showed that V$_L$κ or C$_L$κ variants lacking the Arg108 residue were slightly but distinctively less stable than their corresponding wild type, which is quite similar to the stabilizing effects of N-terminal residues reported for the C$_L$λ (213). When these stability values were compared to those previously published for the C$_L$κ, which lacked the authentic Arg108 but included four additional N-term residues (Fig. 5.6a)(56), the values obtained here for the C$_L$κ –wt which has the authentic N-term (Arg108) were slightly higher than those reported for the published C$_L$κ protein. While we determined a stability to GdmCl of - 16.8 kJ mol^{-1} and a T_{melt} of 52.7 °C for the authentic wild type C$_L$κ, - 14.6 mol^{-1} and 51 °C, respectively, were reported for the published C$_L$κ protein. This implies that the length of the N-terminal segment *per se* is not the crucial factor that controls the stability of the domain, but rather a specific residue at a given position, which in the case of the C$_L$κ domain is Arg108.

A look at how Arg108, which is at the C-terminus of V$_L$κ might affect the amyloidogenicity of the domain showed that the lack of the residue predisposes the V$_L$κ domain to fibril formation (Fig. 5.4a, d). This implies that the additional V$_L$-C-term residues help stabilize the domain and hence decrease

the propensity to fibril formation, which is also consistent with the idea that the entire LC protein is a more stable entity largely resistant to amyloid formation. This is one possible explanation for the fact that larger V$_L$ fragments with polypeptide sequences extended by residues of the C$_L$ N-terminus, show slower fibril formation kinetics than shorter V$_L$ fragments (152).

The analysis of unfolding and refolding kinetics with GdmCl for the V$_L$κ-wt and it C-terminal ΔR108 variant showed that the V$_L$κ-ΔR108 unfolds faster than the V$_L$κ-wt and that both variants refold with the same kinetics. These findings indicate that C-terminal Arg108 is implicated in the rate limiting step of the unfolding kinetics but not with the rate limiting step of the refolding kinetics of the V$_L$k domain. This agrees well with the notion that unfolding events precede fibrillation (104,156,279,280), which would explain why the faster unfolding V$_L$κ-ΔR108 variant is more prone to fibril formation. For the C$_L$κ domain and its N-terminal truncation variants (C$_L$κ-ΔR108 and C$_L$κ-ΔR108/A109), the refolding kinetics were slower for the variants that lacked the N-terminal Arg108. The unfolding kinetics of the C$_L$κ variants could not be assessed by manual mixing experiments, due to the rapid unfolding of the proteins during the dead time of the experiment. Despite the observation of slower refolding kinetics for both variants that lacked Arg108 than for the C$_L$κ-wt (Fig. 5.8; Table 5.2), C$_L$κ-ΔR108 variant which had an exposed Ala109 residue refolded even slower than C$_L$κ-ΔR108/A109, which had and exposed Asp110 residue (Fig. 5.6b) (refolding kinetics: C$_L$κ-ΔR108 < CLκ-ΔR108/A109 < C$_L$κ-wt). This can be explained by the fact that a surface-exposed charged residue as in the C$_L$κ-wt (exposed N-terminal Arg108) or CLκ-ΔR108/A109 (exposed N-terminal Asp110) might influence the overall solubility of the protein and thereby guide its folding to a certain extent (298-301). These findings indicate that the charged N-term Arg108 residue is concerned with the rate limiting step of folding of C$_L$κ domains.

A closer look at the interaction partners of Arg108 within the crystal structure revealed that Arg108 interacts with residues within the C$_L$κ, but not the V$_L$κ domain. From the structural point of view, on the one hand, the interaction of Arg108 within the C$_L$κ would be expected since it is the conserved N-term residue 1 of the domains, whereas on the other hand, it would be unexpected that the same residue is important for the stability and folding of both the V$_L$κ and the C$_L$κ domains. While the interaction of Arg108 with residues within the C$_L$κ domain appears to be a more rational explanation for its stabilizing effects on the domain, this does not hold true for the V$_L$κ domain. This increased stability in the V$_L$κ may originate from the introduction of a charge from the Arg108 residue, which is exposed at the C-terminus of the protein, where it likely provides a new coulombic interaction on the protein surface (298). This could also be a result of the suppression of unfavorable interactions of nonpolar residues with water or from addition of new hydrogen bonds with the solvent (298,301). Furthermore, the presence of a surface charge might have led to an optimal solvation of the protein molecule and thereby an increased stability (301). It is also possible that the Arg108 residue influences the dynamics of the V$_L$κ and/or the C$_L$κ domain. Thus structural studies by NMR and molecular dynamic simulations might help reveal such details.

Similar effects of the N-terminal residues on the integrity of antibody domains where described by Goto and Hamaguchi (213), for the C$_L$λ domain and by Nokwe et al.,(48), for antibody variable

domains. From these reports and our current findings, it is therefore tempting to speculate that the N-termini of immunoglobulin domains have been selected by evolution to stabilize the native state and to direct the folding process. This would be useful in the rational design and engineering of antibody fragments.

CONCLUSIONS AND PERSPECTIVES

In this work, several key questions pertaining to the relationship between sequences of antibody domains and the structures they adopt were addressed. Based on our findings, we are now able to rationalize some of the issues regarding the molecular mechanism of LC amyloid formation. Specific residues that are crucial for the integrity of antibody domains were identified. AL amyloidosis is a complex disease because each patient possesses an LC protein with a unique protein sequence, due to a recombination of germ line sequence and somatic mutations. Our study identifies the N-terminal residue 2 of V_L domains as a key feature that can destabilize and as a result predispose an entire variable domain family to fibrillar aggregation. The importance of residue 2 for the $V_L\kappa$ family determined here adds to the understanding of AL amyloidosis linking the destabilization of native interactions with amyloid formation. Furthermore, we show that a C-terminal Arg108 residue also stabilizes the $V_L\kappa$ domain and thus protects it from fibrillation by limiting its unfolding rate. In addition, our results revealed that the N-terminal Arg108 is also crucial for the stability and folding of the $C_L\kappa$ domains. These findings shed new light on the architecture of antibody domains and the prerequisites for formation of amyloid fibrils. This might also contribute to the rational design and engineering of functional antibody fragments.

A growing body of evidence suggests that protein folding and fibril formation are parallel and competitive processes (1,104), guided by thermodynamic stability. Also within this thesis, we show that thermodynamic stability is not the sole factor *per se* which controls the fibril formation of V_L domains, but that specific interactions are also required such as those established by Asn residues at specific positions within the domain. The two mechanisms underlying the V_L fibrillation process proposed in this work show that fibril formation is preceded by an unfolding event. On the one hand destabilization of the native fold is necessary for the population of intermediates and unfolded species that transform to amyloid fibrils. On the other hand, the fibrillation process is independent of the stability and unfolding of the native state, but depends on the interactions established once the protein has unfolded that give rise to the fibrillar structures. Identification of the regions of the sequence that form and stabilize the fibril core and/or play a primary role in fibril formation is therefore of prime importance for the diagnosis of AL amyloidosis and for the rational design of therapeutics.

Although considerable progress has been made in the elucidation of amyloid fibril properties at a molecular level, a key question in amyloid research remains the molecular mechanisms by which amyloid fibrils or the process of amyloidogenesis cause cellular toxicity and organ damage (94,104,145,199-207). In this context, the use of cellular model systems such as myocardial cells (202,302) and mesangial cells (303), and organismal models like transgenic mice (207,304) and *Caenorhabditis elegans* (305), would not only help in understanding of the process of AL diseases in general but will also help in the identification of the toxic species and the screening of potential therapeutic agents in particular. Also, very little is yet known about the structure of the amyloid protofibrils and unstructured intermediate species that precede their formation and are likely to play a

key role in the pathogenesis of the amyloid disease. The elucidation of the amyloid forming pathway of AL proteins with a clear identification of species (intermediates and oligomers) *en route* to the fibrils will also lead to deeper insights into the nature of the toxic species involved in AL amylodosis and hence the design of therapeutic options that directly target such species.

.

REFERENCES

1. Dobson, C. M. (2003) Protein folding and misfolding. *Nature* **426**, 884-890
2. Clore, G. M., and Gronenborn, A. M. (1989) Determination of three-dimensional structures of proteins and nucleic acids in solution by nuclear magnetic resonance spectroscopy. *Critical reviews in biochemistry and molecular biology* **24**, 479-564
3. Eisenberg, D., and Hill, C. P. (1989) Protein crystallography: more surprises ahead. *Trends Biochem Sci* **14**, 260-264
4. Tzakos, A. G., Grace, C. R., Lukavsky, P. J., and Riek, R. (2006) NMR techniques for very large proteins and rnas in solution. *Annual review of biophysics and biomolecular structure* **35**, 319-342
5. Smyth, M. S., and Martin, J. H. (2000) x ray crystallography. *Molecular pathology : MP* **53**, 8-14
6. Chothia, C. (1992) Proteins. One thousand families for the molecular biologist. *Nature* **357**, 543-544
7. Addou, S., Rentzsch, R., Lee, D., and Orengo, C. A. (2009) Domain-based and family-specific sequence identity thresholds increase the levels of reliable protein function transfer. *J Mol Biol* **387**, 416-430
8. Sjolander, K., Datta, R. S., Shen, Y., and Shoffner, G. M. (2011) Ortholog identification in the presence of domain architecture rearrangement. *Briefings in bioinformatics* **12**, 413-422
9. Orengo, C. A., Todd, A. E., and Thornton, J. M. (1999) From protein structure to function. *Curr Opin Struct Biol* **9**, 374-382
10. Thornton, J. M., Orengo, C. A., Todd, A. E., and Pearl, F. M. (1999) Protein folds, functions and evolution. *J Mol Biol* **293**, 333-342
11. Radford, S. E. (2000) Protein folding: progress made and promises ahead. *Trends Biochem Sci* **25**, 611-618
12. Baldwin, R. L. (1996) Why is protein folding so fast? *Proc Natl Acad Sci U S A.* **93**, 2627-2628
13. Tramontano, A. (2003) Of men and machines. *Nature structural biology* **10**, 87-90
14. Daggett, V., and Fersht, A. (2003) The present view of the mechanism of protein folding. *Nat Rev Mol Cell Biol* **4**, 497-502
15. Ellis, R. J., and Hartl, F. U. (1999) Principles of protein folding in the cellular environment. *Curr Opin Struct Biol* **9**, 102-110
16. Ellis, R. J. (2001) Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Curr Opin Struct Biol* **11**, 114-119
17. Bukau, B., and Horwich, A. L. (1998) The Hsp70 and Hsp60 chaperone machines. *Cell* **92**, 351-366
18. Hartl, F. U., and Hayer-Hartl, M. (2002) Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295**, 1852-1858
19. Hartl, F. U. (1996) Molecular chaperones in cellular protein folding. *Nature* **381**, 571-579
20. Walter, S., and Buchner, J. (2002) Molecular chaperones--cellular machines for protein folding. *Angewandte Chemie* **41**, 1098-1113
21. Gething, M. J., and Sambrook, J. (1992) Protein folding in the cell. *Nature* **355**, 33-45
22. Dobson, C. M. (2001) The structural basis of protein folding and its links with human disease. *Philos Trans R Soc Lond B Biol Sci* **356**, 133-145
23. Wang, S., and Kaufman, R. J. (2012) The impact of the unfolded protein response on human disease. *J Cell Biol* **197**, 857-867
24. Radford, S. E., and Dobson, C. M. (1999) From computer simulations to human disease: emerging themes in protein folding. *Cell* **97**, 291-298
25. Anfinsen, C. B. (1973) Principles that govern the folding of protein chains. *Science* **181**, 223-230
26. Seckler, R., and Jaenicke, R. (1992) Protein folding and protein refolding. *FASEB J* **6**, 2545-2552
27. Fersht, A. R., and Daggett, V. (2002) Protein folding and unfolding at atomic resolution. *Cell* **108**, 573-582
28. Levinthal, C. (1968) Are there pathways for protein folding? *The Journal of Chemical Physics* **65**, 44-45
29. Levinthal, C. (1969) How to Fold Graciously. in *Mossbauer Spectroscopy in Biological Systems* (Ed.. DeBrunner, T. P. a. M., E. ed., University of Illinois Press (1969), Allerton House, Monticello, Illinois

30.   Baldwin, R. L. (1996) Why is protein folding so fast? *Proc Natl Acad Sci U S A* **93**, 2627-2628
31.   Chan, H. S., and Dill, K. A. (1998) Protein folding in the landscape perspective: chevron plots and non-Arrhenius kinetics. *Proteins* **30**, 2-33
32.   Dobson, C. M. (1999) Protein misfolding, evolution and disease. *Trends Biochem Sci* **24**, 329-332
33.   Kubelka, J., Hofrichter, J., and Eaton, W. A. (2004) The protein folding 'speed limit'. *Curr Opin Struct Biol* **14**, 76-88
34.   Dobson, C. M. (2002) Getting out of shape. *Nature* **418**, 729-730
35.   Honig, B. (1999) Protein folding: from the levinthal paradox to structure prediction. *J Mol Biol* **293**, 283-293
36.   Feige, M. J., Hendershot, L. M., and Buchner, J. (2010) How antibodies fold. *Trends Biochem Sci* **35**, 189-198
37.   Rochet, J. C., and Lansbury, P. T., Jr. (2000) Amyloid fibrillogenesis: themes and variations. *Curr Opin Struct Biol* **10**, 60-68
38.   Sunde, M., Serpell, L. C., Bartlam, M., Fraser, P. E., Pepys, M. B., and Blake, C. C. (1997) Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J Mol Biol* **273**, 729-739
39.   Woolfson, D. N. (2001) Core-directed protein design. *Curr Opin Struct Biol* **11**, 464-471
40.   Koga, N., Tatsumi-Koga, R., Liu, G., Xiao, R., Acton, T. B., Montelione, G. T., and Baker, D. (2012) Principles for designing ideal protein structures. *Nature* **491**, 222-227
41.   Baldwin, R. L. (2005) Weak interactions in Protein Folding: Hydrophobic free energy, van der Waals interactions, Peptide Hydrogen Bonds, and Peptide Solvation. in *Protein Folding Handbook* (Buchner, J., and Kiefhaber, T. ed.), Wiley-VCH Verlag GmbH & Co. KGaA. pp
42.   Baldwin, R. L. (2013) Properties of hydrophobic free energy found by gas-liquid transfer. *Proc Natl Acad Sci U S A* **110**, 1670-1673
43.   Dill, K. A. (1990) Dominant forces in protein folding. *Biochemistry* **29**, 7133-7155
44.   Kauzmann, W. (1959) Some Factors in the Interpretation of Protein Denaturation. in *Advances in Protein Chemistry* (Anfinsen, C. B., and John, T. E. ed.), Academic Press. pp 1-63
45.   Huang, K. (2005) *Lectures on Statistical Physics and Protein Folding*, World Scientific Publishing Co. Pte. Ltd.
46.   Nemethy, G., and Scheraga, H. A. (1962) Structure of water and hydrophobic bonding in proteins. I. A model for thermodynamic properties of liquid water. *J. Chem. Phys.*, 3382–3400
47.   Chandler, D. (2005) Interfaces and the driving force of hydrophobic assembly. *Nature* **437**, 640-647
48.   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
49.   Vetri, V., Librizzi, F., Leone, M., and Militello, V. (2007) Thermal aggregation of bovine serum albumin at different pH: comparison with human serum albumin. *European biophysics journal : EBJ* **36**, 717-725
50.   Lopez de la Paz, M., and Serrano, L. (2004) Sequence determinants of amyloid fibril formation. *Proc Natl Acad Sci U S A* **101**, 87-92
51.   Chiti, F., Stefani, M., Taddei, N., Ramponi, G., and Dobson, C. M. (2003) Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* **424**, 805-808
52.   Krebs, M. R., Devlin, G. L., and Donald, A. M. (2007) Protein particulates: another generic form of protein aggregation? *Biophys J* **92**, 1336-1342
53.   Fernandez, A., Kardos, J., Scott, L. R., Goto, Y., and Berry, R. S. (2003) Structural defects and the diagnosis of amyloidogenic propensity. *Proc Natl Acad Sci U S A* **100**, 6446-6451
54.   Fernandez, A., and Scheraga, H. A. (2003) Insufficiently dehydrated hydrogen bonds as determinants of protein interactions. *Proc Natl Acad Sci U S A* **100**, 113-118
55.   Baldwin, R. L. (2007) Energetics of protein folding. *J Mol Biol* **371**, 283-301
56.   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
57.   Dill, K. A., and Chan, H. S. (1997) From Levinthal to pathways to funnels. *Nature structural biology* **4**, 10-19
58.   Leopold, P. E., Montal, M., and Onuchic, J. N. (1992) Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc Natl Acad Sci U S A* **89**, 8721-8725
59.   Wolynes, P. G., Onuchic, J. N., and Thirumalai, D. (1995) Navigating the folding routes. *Science* **267**, 1619-1620

60.   Dobson, C. M., and Karplus, M. (1999) The fundamentals of protein folding: bringing together theory and experiment. *Curr Opin Struct Biol* **9**, 92-101

61.   Yon, J. M. (2001) Protein folding: a perspective for biology, medicine and biotechnology. *Brazilian journal of medical and biological research = Revista brasileira de pesquisas medicas e biologicas / Sociedade Brasileira de Biofisica ... [et al.]* **34**, 419-435

62.   Ptitsyn, O. B., Bychkova, V. E., and Uversky, V. N. (1995) Kinetic and equilibrium folding intermediates. *Philos Trans R Soc Lond B Biol Sci* **348**, 35-41

63.   Ptitsyn, O. B. (1995) Structures of folding intermediates. *Curr Opin Struct Biol* **5**, 74-78

64.   Morris, E. R., and Searle, M. S. (2012) *Overview of protein folding mechanisms: experimental and theoretical approaches to probing energy landscapes.*,

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

66.   Dill, K. A., and MacCallum, J. L. (2012) The protein-folding problem, 50 years on. *Science* **338**, 1042-1046

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

68.   Lansbury, P. T., Jr. (1999) Evolution of amyloid: what normal protein folding may tell us about fibrillogenesis and disease. *Proc Natl Acad Sci U S A* **96**, 3342-3344

69.   Arai, M., and Kuwajima, K. (2000) Role of the molten globule state in protein folding. *Adv Protein Chem* **53**, 209-282

70.   Ptitsyn, O. B. (1995) Molten globule and protein folding. *Adv Protein Chem* **47**, 83-229

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

72.   Thies, M. J., Kammermeier, R., Richter, K., and Buchner, J. (2001) The alternatively folded state of the antibody C(H)3 domain. *J Mol Biol* **309**, 1077-1085

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

74.   Feige, M. J., Simpson, E. R., Herold, E. M., Bepperling, A., Heger, K., and Buchner, J. (2010) Dissecting the alternatively folded state of the antibody Fab fragment. *J Mol Biol* **399**, 719-730

75.   Frauenfelder, H., Parak, F., and Young, R. D. (1988) Conformational substates in proteins. *Annual review of biophysics and biophysical chemistry* **17**, 451-479

76.   Jahn, T. R., and Radford, S. E. (2005) The Yin and Yang of protein folding. *FEBS J* **272**, 5962-5970

77.   Fersht, A. (1999) *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*, W.H. Freeman and Company.

78.   Liu, Z., and Chan, H. S. (2005) Desolvation is a likely origin of robust enthalpic barriers to protein folding. *J Mol Biol* **349**, 872-889

79.   Dyer, R. B. (2007) Ultrafast and downhill protein folding. *Curr Opin Struct Biol* **17**, 38-47

80.   Simpson, E. R. (2006) *Exploring the Folding Pathway of Ubiquitin.* Ph.D., Dissertation submitted to the University of Nottingham

81.   Pace, C. N. (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol* **131**, 266-280

82.   Matthews, C. R. (1993) Pathways of protein folding. *Annu Rev Biochem* **62**, 653-683

83.   Tanford, C. (1968) Protein denaturation. *Adv Protein Chem* **23**, 121-282

84.   Jha, S. K., and Marqusee, S. (2014) Kinetic evidence for a two-stage mechanism of protein denaturation by guanidinium chloride. *Proc Natl Acad Sci U S A* **111**, 4856-4861

85.   Schellman, J. A. (2002) Fifty years of solvent denaturation. *Biophys Chem* **96**, 91-101

86.   Moglich, A., Krieger, F., and Kiefhaber, T. (2005) Molecular basis for the effect of urea and guanidinium chloride on the dynamics of unfolded polypeptide chains. *J Mol Biol* **345**, 153-162

87.   Soto, C. (2003) Unfolding the role of protein misfolding in neurodegenerative diseases. *Nat Rev Neurosci* **4**, 49-60

88.   Soto, C., Estrada, L., and Castilla, J. (2006) Amyloids, prions and the inherent infectious nature of misfolded protein aggregates. *Trends Biochem Sci* **31**, 150-155

89.   Ross, C. A., and Poirier, M. A. (2004) Protein aggregation and neurodegenerative disease. *Nat Med* **10 Suppl**, S10-17

90.   Blancas-Mejia, L. M., and Ramirez-Alvarado, M. (2013) Systemic amyloidoses. *Annu Rev Biochem* **82**, 745-774

91.   Falk, R. H., Comenzo, R. L., and Skinner, M. (1997) The systemic amyloidoses. *N Engl J Med* **337**, 898-909

92.   Fink, A. L. (1998) Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Fold Des* **3(1)**, R9-23

93.   Vetri, V., Canale, C., Relini, A., Librizzi, F., Militello, V., Gliozzi, A., and Leone, M. (2007) Amyloid fibrils formation and amorphous aggregation in concanavalin A. *Biophys Chem* **125**, 184-190

94.   Jahn, T. R., and Radford, S. E. (2008) Folding versus aggregation: polypeptide conformations on competing pathways. *Arch Biochem Biophys* **469**, 100-117

95.   Uversky, V. N., Li, J., and Fink, A. L. (2001) Evidence for a partially folded intermediate in alpha-synuclein fibril formation. *J Biol Chem* **276**, 10737-10744

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

97.   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 U S A* **96**, 3590-3594

98.   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

99.   Haass, C., Hung, A. Y., and Selkoe, D. J. (1991) Processing of beta-amyloid precursor protein in microglia and astrocytes favors an internal localization over constitutive secretion. *J Neurosci* **11**, 3783-3793

100.   Dispenzieri, A., Gertz, M. A., and Buadi, F. (2012) What do I need to know about immunoglobulin light chain (AL) amyloidosis? *Blood reviews* **26**, 137-154

101.   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

102.   Ross, C. A., Poirier, M. A.. (2004) Protein aggregation and neurodegenerative disease. *Nat Med*, 10 Suppl:S10-17.

103.   Dobson, C. M. (2004) Principles of protein folding, misfolding and aggregation. *Semin Cell Dev Biol* **15**, 3-16

104.   Chiti, F., and Dobson, C. M. (2006) Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* **75**, 333-366

105.   Fandrich, M., and Dobson, C. M. (2002) The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation. *EMBO J* **21**, 5682-5690

106.   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

107.   So, M., Yagi, H., Sakurai, K., Ogi, H., Naiki, H., and Goto, Y. Ultrasonication-dependent acceleration of amyloid fibril formation. *J Mol Biol* **412**, 568-577

108.   Ferrone, F. (1999) Analysis of protein aggregation kinetics. *Methods Enzymol* **309**, 256-274

109.   Ionescu-Zanetti, C., Khurana, R., Gillespie, J. R., Petrick, J. S., Trabachino, L. C., Minert, L. J., Carter, S. A., and Fink, A. L. (1999) Monitoring the assembly of Ig light-chain amyloid fibrils by atomic force microscopy. *Proc Natl Acad Sci U S A* **96**, 13175-13179

110.   McParland, V. J., Kad, N. M., Kalverda, A. P., Brown, A., Kirwin-Jones, P., Hunter, M. G., Sunde, M., and Radford, S. E. (2000) Partially unfolded states of beta(2)-microglobulin and amyloid formation in vitro. *Biochemistry* **39**, 8735-8746

111.   Nielsen, L., Khurana, R., Coats, A., Frokjaer, S., Brange, J., Vyas, S., Uversky, V. N., and Fink, A. L. (2001) Effect of environmental factors on the kinetics of insulin fibril formation: elucidation of the molecular mechanism. *Biochemistry* **40**, 6036-6046

112.   Yang, M., Lei, M., Bruschweiler, R., and Huo, S. (2005) Initial conformational changes of human transthyretin under partially denaturing conditions. *Biophys J* **89**, 433-443

113.   Nelson, R., Sawaya, M. R., Balbirnie, M., Madsen, A. O., Riekel, C., Grothe, R., and Eisenberg, D. (2005) Structure of the cross-beta spine of amyloid-like fibrils. *Nature* **435**, 773-778

114.   Jimenez, J. L., Guijarro, J. I., Orlova, E., Zurdo, J., Dobson, C. M., Sunde, M., and Saibil, H. R. (1999) Cryo-electron microscopy structure of an SH3 amyloid fibril and model of the molecular packing. *EMBO J* **18**, 815-821

115.   Harper, J. D., and Lansbury, P. T., Jr. (1997) Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins. *Annu Rev Biochem* **66**, 385-407

116.   Sipe, J. D., Benson, M. D., Buxbaum, J. N., Ikeda, S., Merlini, G., Saraiva, M. J., and Westermark, P. Amyloid fibril protein nomenclature: 2010 recommendations from the nomenclature committee of the International Society of Amyloidosis. *Amyloid* **17**, 101-104

117.   Westermark, G. T., Johnson, K. H., and Westermark, P. (1999) Staining methods for identification of amyloid in tissue. *Methods Enzymol* **309**, 3-25

118.   LeVine, H., 3rd. (1993) Thioflavine T interaction with synthetic Alzheimer's disease beta-amyloid peptides: detection of amyloid aggregation in solution. *Protein Sci* **2**, 404-410

119.   Khurana, R., Coleman, C., Ionescu-Zanetti, C., Carter, S. A., Krishna, V., Grover, R. K., Roy, R., and Singh, S. (2005) Mechanism of thioflavin T binding to amyloid fibrils. *J Struct Biol* **151**, 229-238

120.   Bhak, G., Choe, Y. J., and Paik, S. R. (2009) Mechanism of amyloidogenesis: nucleation-dependent fibrillation versus double-concerted fibrillation. *BMB Rep* **42**, 541-551

121.   Standford hospital and clinic, Amyloid division. .

122.   Williams, A. F., and Barclay, A. N. (1988) The immunoglobulin superfamily--domains for cell surface recognition. *Annu Rev Immunol* **6**, 381-405

123.   Harpaz, Y., and Chothia, C. (1994) Many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to a new structural set which is close to that containing variable domains. *J Mol Biol* **238**, 528-539

124.   Pancer, Z., and Cooper, M. D. (2006) The evolution of adaptive immunity. *Annu Rev Immunol* **24**, 497-518

125.   Xu, Z., Zan, H., Pone, E. J., Mai, T., and Casali, P. (2012) Immunoglobulin class-switch DNA recombination: induction, targeting and beyond. *Nat Rev Immunol* **12**, 517-531

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

127.   Halaby, D. M., Poupon, A., and Mornon, J. (1999) The immunoglobulin fold family: sequence analysis and 3D structure comparisons. *Protein Eng* **12**, 563-571

128.   Chan, A. C., and Carter, P. J. (2010) Therapeutic antibodies for autoimmunity and inflammation. *Nat Rev Immunol* **10**, 301-316

129.   Carter, P. J. (2006) Potent antibody therapeutics by design. *Nat Rev Immunol* **6**, 343-357

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

131.   Sliwkowski, M. X., and Mellman, I. (2013) Antibody therapeutics in cancer. *Science* **341**, 1192-1198

132.   Jefferis, R. (2009) Glycosylation as a strategy to improve antibody-based therapeutics. *Nat Rev Drug Discov* **8**, 226-234

133.   Schiffer, M. (1996) Molecular anatomy and the pathological expression of antibody light chains. *Am J Pathol* **148**, 1339-1344

134.   Jung, D., Giallourakis, C., Mostoslavsky, R., and Alt, F. W. (2006) Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annu Rev Immunol* **24**, 541-570

135.   Tonegawa, S. (1983) Somatic generation of antibody diversity. *Nature* **302**, 575-581

136.   Rajewsky, K., Forster, I., and Cumano, A. (1987) Evolutionary and somatic selection of the antibody repertoire in the mouse. *Science* **238**, 1088-1094

137.   McIntosh, R., Watson, P., and Weetman, A. (1998) Somatic hypermutation in autoimmune thyroid disease. *Immunol Rev* **162**, 219-231

138.   Schroeder, K., Herrmann, M., and Winkler, T. H. (2013) The role of somatic hypermutation in the generation of pathogenic antibodies in SLE. *Autoimmunity* **46**, 121-127

139.   Poshusta, T. L., Sikkink, L. A., Leung, N., Clark, R. J., Dispenzieri, A., and Ramirez-Alvarado, M. (2009) Mutations in specific structural regions of immunoglobulin light chains are associated with free light chain levels in patients with AL amyloidosis. *PLoS One* **4**, e5169

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

141.   Feige, M. J., Groscurth, S., Marcinowski, M., Shimizu, Y., Kessler, H., Hendershot, L. M., and Buchner, J. (2009) An unfolded CH1 domain controls the assembly and secretion of IgG antibodies. *Mol Cell* **34**, 569-579

142.   Hendershot, L., Bole, D., Kohler, G., and Kearney, J. F. (1987) Assembly and secretion of heavy chains that do not associate posttranslationally with immunoglobulin heavy chain-binding protein. *J Cell Biol* **104**, 761-767

143.   Coleclough, C. (1983) Chance, necessity and antibody gene dynamics. *Nature* **303**, 23-26

144.   Abraham, R. S., Katzmann, J. A., Clark, R. J., Bradwell, A. R., Kyle, R. A., and Gertz, M. A. (2003) Quantitative analysis of serum free light chains. A new marker for the diagnostic evaluation of primary systemic amyloidosis. *Am J Clin Pathol* **119**, 274-278

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

146.   Kyle, R. A., and Gertz, M. A. (1995) Primary systemic amyloidosis: clinical and laboratory features in 474 cases. *Semin Hematol* **32**, 45-59

147.   Merlini, G., and Palladini, G. (2008) Amyloidosis: is a cure possible? *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* **19 Suppl 4**, iv63-66

148.   Buxbaum, J., and Gallo, G. (1999) Nonamyloidotic monoclonal immunoglobulin deposition disease. Light-chain, heavy-chain, and light- and heavy-chain deposition diseases. *Hematol Oncol Clin North Am* **13**, 1235-1248

149.   Pozzi, C., D'Amico, M., Fogazzi, G. B., Curioni, S., Ferrario, F., Pasquali, S., Quattrocchio, G., Rollino, C., Segagni, S., and Locatelli, F. (2003) Light chain deposition disease with renal involvement: clinical characteristics and prognostic factors. *Am J Kidney Dis* **42**, 1154-1163

150.   Stevens, F. J. (2000) Four structural risk factors identify most fibril-forming kappa light chains. *Amyloid* **7**, 200-211

151.   Enqvist, S., Sletten, K., Stevens, F. J., Hellman, U., and Westermark, P. (2007) Germ line origin and somatic mutations determine the target tissues in systemic AL-amyloidosis. *PLoS One* **2**, e981

152.   Enqvist, S., Sletten, K., and Westermark, P. (2009) Fibril protein fragmentation pattern in systemic AL-amyloidosis. *J Pathol* **219**, 473-480

153.   Stevens, P. W., Raffen, R., Hanson, D. K., Deng, Y. L., Berrios-Hammond, M., Westholm, F. A., Murphy, C., Eulitz, M., Wetzel, R., Solomon, A., and et al. (1995) Recombinant immunoglobulin variable domains generated from synthetic genes provide a system for in vitro characterization of light-chain amyloid proteins. *Protein Sci* **4**, 421-432

154.   Wall, J., Schell, M., Murphy, C., Hrncic, R., Stevens, F. J., and Solomon, A. (1999) Thermodynamic instability of human lambda 6 light chains: correlation with fibrillogenicity. *Biochemistry* **38**, 14101-14108

155.   Kim, Y., Wall, J. S., Meyer, J., Murphy, C., Randolph, T. W., Manning, M. C., Solomon, A., and Carpenter, J. F. (2000) Thermodynamic modulation of light chain amyloid fibril formation. *J Biol Chem* **275**, 1570-1574

156.   Raffen, R., Dieckman, L. J., Szpunar, M., Wunschl, C., Pokkuluri, P. R., Dave, P., Wilkins Stevens, P., Cai, X., Schiffer, M., and Stevens, F. J. (1999) Physicochemical consequences of amino acid variations that contribute to fibril formation by immunoglobulin light chains. *Protein Sci* **8**, 509-517

157.   Omtvedt, L. A., Husby, G., Cornwell, G. G., 3rd, Kyle, R. A., and Sletten, K. (1997) The amino acid sequence of the glycosylated amyloid immunoglobulin light chain protein AL MS. *Scand J Immunol* **45**, 551-556

158.   Holm, E., Sletten, K., and Husby, G. (1986) Structural studies of a carbohydrate-containing immunoglobulin-lambda-light-chain amyloid-fibril protein (AL) of variable subgroup III. *Biochem J* **239**, 545-551

159.   Connors, L. H., Jiang, Y., Budnik, M., Theberge, 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 kappa1 immunoglobulin light chains. *Biochemistry* **46**, 14259-14271

160.   Stevens, F. J., and Argon, Y. (1999) Pathogenic light chains and the B-cell repertoire. *Immunol Today* **20**, 451-457

161.   Bosman, F. T., and Stamenkovic, I. (2003) Functional structure and composition of the extracellular matrix. *J Pathol* **200**, 423-428

162.   Nelson, S. R., Lyon, M., Gallagher, J. T., Johnson, E. A., and Pepys, M. B. (1991) Isolation and characterization of the integral glycosaminoglycan constituents of human amyloid A and monoclonal light-chain amyloid fibrils. *Biochem J* **275 (Pt 1)**, 67-73

163.   Jiang, X., Myatt, E., Lykos, P., and Stevens, F. J. (1997) Interaction between glycosaminoglycans and immunoglobulin light chains. *Biochemistry* **36**, 13187-13194

164.   Ohishi, H., Skinner, M., Sato-Araki, N., Okuyama, T., Gejyo, F., Kimura, A., Cohen, A. S., and Schmid, K. (1990) Glycosaminoglycans of the hemodialysis-associated carpal synovial amyloid and of amyloid-rich tissues and fibrils of heart, liver, and spleen. *Clin Chem* **36**, 88-91

165.   McLaughlin, R. W., De Stigter, J. K., Sikkink, L. A., Baden, E. M., and Ramirez-Alvarado, M. (2006) The effects of sodium sulfate, glycosaminoglycans, and Congo red on the structure, stability, and amyloid formation of an immunoglobulin light-chain protein. *Protein Sci* **15**, 1710-1722

166.   Kisilevsky, R., and Fraser, P. (1996) Proteoglycans and amyloid fibrillogenesis. *Ciba Found Symp* **199**, 58-67; discussion 68-72, 90-103

167.   Pepys, M. B., Rademacher, T. W., Amatayakul-Chantler, S., Williams, P., Noble, G. E., Hutchinson, W. L., Hawkins, P. N., Nelson, S. R., Gallimore, J. R., Herbert, J., and et al. (1994) Human serum amyloid P component is an invariant constituent of amyloid deposits and has a uniquely homogeneous glycostructure. *Proc Natl Acad Sci U S A* **91**, 5602-5606

168.   Tennent, G. A., Lovat, L. B., and Pepys, M. B. (1995) Serum amyloid P component prevents proteolysis of the amyloid fibrils of Alzheimer disease and systemic amyloidosis. *Proc Natl Acad Sci U S A* **92**, 4299-4303

169.   Souillac, P. O., Uversky, V. N., Millett, I. S., Khurana, R., Doniach, S., and Fink, A. L. (2002) Elucidation of the molecular mechanism during the early events in immunoglobulin light chain amyloid fibrillation. Evidence for an off-pathway oligomer at acidic pH. *J Biol Chem* **277**, 12666-12679

170.   Souillac, P. O., Uversky, V. N., and Fink, A. L. (2003) Structural transformations of oligomeric intermediates in the fibrillation of the immunoglobulin light chain LEN. *Biochemistry* **42**, 8094-8104

171.   Davis, P. D., Raffen, R., Dul, L. J., Vogen, M. S., Williamson, K. E., Stevens, J. F., and Argon, Y. (2000) Inhibition of amyloid fiber assembly by both BiP and its target peptide. *Immunity* **13**, 433-442

172.   Khurana, R., Gillespie, J. R., Talapatra, A., Minert, L. J., Ionescu-Zanetti, C., Millett, I., and Fink, A. L. (2001) Partially folded intermediates as critical precursors of light chain amyloid fibrils and amorphous aggregates. *Biochemistry* **40**, 3525-3535

173.   Khurana, R., Souillac, P. O., Coats, A. C., Minert, L., Ionescu-Zanetti, C., Carter, S. A., Solomon, A., and Fink, A. L. (2003) A model for amyloid fibril formation in immunoglobulin light chains based on comparison of amyloidogenic and benign proteins and specific antibody binding. *Amyloid* **10**, 97-109

174.   Giehm, L., and Otzen, D. E. Strategies to increase the reproducibility of protein fibrillization in plate reader assays. *Anal Biochem* **400**, 270-281

175.   Yamamoto, K., Yagi, H., Lee, Y. H., Kardos, J., Hagihara, Y., Naiki, H., and Goto, Y. The amyloid fibrils of the constant domain of immunoglobulin light chain. *FEBS Lett*

176.   Souillac, P. O., Uversky, V. N., Millett, I. S., Khurana, R., Doniach, S., and Fink, A. L. (2002) Effect of association state and conformational stability on the kinetics of immunoglobulin light chain amyloid fibril formation at physiological pH. *J Biol Chem* **277**, 12657-12665

177.   Kim, Y. S., Cape, S. P., Chi, E., Raffen, R., Wilkins-Stevens, P., Stevens, F. J., Manning, M. C., Randolph, T. W., Solomon, A., and Carpenter, J. F. (2001) Counteracting effects of renal solutes on amyloid fibril formation by immunoglobulin light chains. *J Biol Chem* **276**, 1626-1633

178.   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

179.   Zhu, M., Han, S., Zhou, F., Carter, S. A., and Fink, A. L. (2004) Annular oligomeric amyloid intermediates observed by in situ atomic force microscopy. *J Biol Chem* **279**, 24452-24459

180.   Meng, X., Fink, A. L., and Uversky, V. N. (2008) The effect of membranes on the in vitro fibrillation of an amyloidogenic light-chain variable-domain SMA. *J Mol Biol* **381**, 989-999

181.   Yamamoto, K., Yagi, H., Lee, Y. H., Kardos, J., Hagihara, Y., Naiki, H., and Goto, Y. The amyloid fibrils of the constant domain of immunoglobulin light chain. *FEBS Lett* **584**, 3348-3353

182.   Yamamoto, S., Hasegawa, K., Yamaguchi, I., Tsutsumi, S., Kardos, J., Goto, Y., Gejyo, F., and Naiki, H. (2004) Low concentrations of sodium dodecyl sulfate induce the extension of beta 2-microglobulin-related amyloid fibrils at a neutral pH. *Biochemistry* **43**, 11075-11082

183.   Pertinhez, T. A., Bouchard, M., Smith, R. A., Dobson, C. M., and Smith, L. J. (2002) Stimulation and inhibition of fibril formation by a peptide in the presence of different concentrations of SDS. *FEBS Lett* **529**, 193-197

184.   Baden, E. M., Randles, E. G., Aboagye, A. K., Thompson, J. R., and Ramirez-Alvarado, M. (2008) Structural insights into the role of mutations in amyloidogenesis. *J Biol Chem* **283**, 30950-30956

185.   Kim, Y. S., Randolph, T. W., Stevens, F. J., and Carpenter, J. F. (2002) Kinetics and energetics of assembly, nucleation, and growth of aggregates and fibrils for an amyloidogenic protein. Insights into transition states from pressure, temperature, and co-solute studies. *J Biol Chem* **277**, 27240-27246

186.   Abraham, R. S., Geyer, S. M., Ramirez-Alvarado, M., Price-Troska, T. L., Gertz, M. A., and Fonseca, R. (2004) Analysis of somatic hypermutation and antigenic selection in the clonal B cell in immunoglobulin light chain amyloidosis (AL). *J Clin Immunol* **24**, 340-353

187.   Comenzo, R. L., Zhang, Y., Martinez, C., Osman, K., and Herrera, G. A. (2001) The tropism of organ involvement in primary systemic amyloidosis: contributions of Ig V(L) germ line gene use and clonal plasma cell burden. *Blood* **98**, 714-720

188.   Perfetti, V., Casarini, S., Palladini, G., Vignarelli, M. C., Klersy, C., Diegoli, M., Ascari, E., and Merlini, G. (2002) Analysis of V(lambda)-J(lambda) expression in plasma cells from primary (AL) amyloidosis and normal bone marrow identifies 3r (lambdaIII) as a new amyloid-associated germline gene segment. *Blood* **100**, 948-953

189.   Abraham, R. S., Geyer, S. M., Price-Troska, T. L., Allmer, C., Kyle, R. A., Gertz, M. A., and Fonseca, R. (2003) Immunoglobulin light chain variable (V) region genes influence clinical presentation and outcome in light chain-associated amyloidosis (AL). *Blood* **101**, 3801-3808

190.   Prokaeva, T., Spencer, B., Kaut, M., Ozonoff, A., Doros, G., Connors, L. H., Skinner, M., and Seldin, D. C. (2007) Soft tissue, joint, and bone manifestations of AL amyloidosis: clinical presentation, molecular features, and survival. *Arthritis Rheum* **56**, 3858-3868

191.   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

192.   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

193.   Simpson, E. R., Herold, E. M., and Buchner, J. (2009) The folding pathway of the antibody V(L) domain. *J Mol Biol* **392**, 1326-1338

194.   Worn, A., and Pluckthun, A. (2001) Stability engineering of antibody single-chain Fv fragments. *J Mol Biol* **305**, 989-1010

195.   Zahnd, C., Spinelli, S., Luginbuhl, B., Amstutz, P., Cambillau, C., and Pluckthun, A. (2004) Directed in vitro evolution and crystallographic analysis of a peptide-binding single chain antibody fragment (scFv) with low picomolar affinity. *J Biol Chem* **279**, 18870-18877

196.   Jager, M., and Pluckthun, A. (1997) The rate-limiting steps for the folding of an antibody scFv fragment. *FEBS Lett* **418**, 106-110

197.   Hoyer, W., Ramm, K., and Pluckthun, A. (2002) A kinetic trap is an intrinsic feature in the folding pathway of single-chain Fv fragments. *Biophys Chem* **96**, 273-284

198.   Rothlisberger, D., Honegger, A., and Pluckthun, A. (2005) Domain interactions in the Fab fragment: a comparative evaluation of the single-chain Fv and Fab format engineered with variable domains of different stability. *J Mol Biol* **347**, 773-789

199.   Olzscha, H., Schermann, S. M., Woerner, A. C., Pinkert, S., Hecht, M. H., Tartaglia, G. G., Vendruscolo, M., Hayer-Hartl, M., Hartl, F. U., and Vabulas, R. M. Amyloid-like aggregates sequester numerous metastable proteins with essential cellular functions. *Cell* **144**, 67-78

200.   Campioni, S., Mannini, B., Zampagni, M., Pensalfini, A., Parrini, C., Evangelisti, E., Relini, A., Stefani, M., Dobson, C. M., Cecchi, C., and Chiti, F. A causative link between the structure of aberrant protein oligomers and their toxicity. *Nat Chem Biol* **6**, 140-147

201.   Cleary, J. P., Walsh, D. M., Hofmeister, J. J., Shankar, G. M., Kuskowski, M. A., Selkoe, D. J., and Ashe, K. H. (2005) Natural oligomers of the amyloid-beta protein specifically disrupt cognitive function. *Nat Neurosci* **8**, 79-84

202.   Brenner, D. A., Jain, M., Pimentel, D. R., Wang, B., Connors, L. H., Skinner, M., Apstein, C. S., and Liao, R. (2004) Human amyloidogenic light chains directly impair cardiomyocyte function through an increase in cellular oxidant stress. *Circ Res* **94**, 1008-1010

203.   Liao, R., Jain, M., Teller, P., Connors, L. H., Ngoy, S., Skinner, M., Falk, R. H., and Apstein, C. S. (2001) Infusion of light chains from patients with cardiac amyloidosis causes diastolic dysfunction in isolated mouse hearts. *Circulation* **104**, 1594-1597

204.   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

205.   Reixach, N., Deechongkit, S., Jiang, X., Kelly, J. W., and Buxbaum, J. N. (2004) Tissue damage in the amyloidoses: Transthyretin monomers and nonnative oligomers are the major cytotoxic species in tissue culture. *Proc Natl Acad Sci U S A* **101**, 2817-2822

206.   Walsh, D. M., and Selkoe, D. J. (2004) Oligomers on the brain: the emerging role of soluble protein aggregates in neurodegeneration. *Protein Pept Lett* **11**, 213-228

207. Ward, J. E., Ren, R., Toraldo, G., Soohoo, P., Guan, J., O'Hara, C., Jasuja, R., Trinkaus-Randall, V., Liao, R., Connors, L. H., and Seldin, D. C. (2011) Doxycycline reduces fibril formation in a transgenic mouse model of AL amyloidosis. *Blood* **118**, 6610-6617

208. Miyazaki, D., Yazaki, M., Gono, T., Kametani, F., Tsuchiya, A., Matsuda, M., Takenaka, Y., Hosh, Y., 2nd, and Ikeda, S. (2008) AH amyloidosis associated with an immunoglobulin heavy chain variable region (VH1) fragment: a case report. *Amyloid : the international journal of experimental and clinical investigation : the official journal of the International Society of Amyloidosis* **15**, 125-128

209. 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 U S A* **105**, 13373-13378

210. 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. Implications for folding and interaction with the chaperone bip. *J Biol Chem* **276**, 3287-3294

211. Wetzel, R. (1997) Domain stability in immunoglobulin light chain deposition disorders. *Adv Protein Chem* **50**, 183-242

212. Baden, E. M., Owen, B. A., Peterson, F. C., Volkman, B. F., Ramirez-Alvarado, M., and Thompson, J. R. (2008) Altered dimer interface decreases stability in an amyloidogenic protein. *J Biol Chem* **283**, 15853-15860

213. Goto, Y., and Hamaguchi, K. (1987) Role of amino-terminal residues in the folding of the constant fragment of the immunoglobulin light chain. *Biochemistry* **26**, 1879-1884

214. Sambrook, J. F., and Russell, D. W. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd edition ed., Cold Spring Harbor Laboratory Press

215. Feige, M. J., Walter, S., and Buchner, J. (2004) Folding mechanism of the CH2 antibody domain. *J Mol Biol* **344**, 107-118

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

217. Schmid, F. X. (2005) *Spectroscopic techniques to study protein folding and stability, in The Protein Folding Handbook; Editor: J. Buchner, Kiefhaber, T.*, Wiley-VCH Verlag GmbH & Co. KGaA

218. Santoro, M. M., and Bolen, D. W. (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* **27**, 8063-8068

219. Lakowicz, J. R. (2006) *Principles of Fluorescence Spectroscopy*, 3rd Edition ed., Springer Science+Business Media, LLC

220. Li, H., Rahimi, F., Sinha, S., Maiti, P., and Bitan, G. (2009) Amyloids and Protein Aggregation–Analytical Methods. in *Encyclopedia of Analytical Chemistry R.A. Meyers* (Murakami, K. ed., John Wiley & Sons Ltd

221. Duy, C., and Fitter, J. (2006) How aggregation and conformational scrambling of unfolded states govern fluorescence emission spectra. *Biophys J* **90**, 3704-3711

222. Feige, M. J., Hendershot, L. M., and Buchner, J. How antibodies fold. *Trends Biochem Sci* **35**, 189-198

223. Goto, Y., Azuma, T., and Hamaguchi, K. (1979) Refolding of the immunoglobulin light chain. *J Biochem* **85**, 1427-1438

224. Hawe, A., Sutter, M., and Jiskoot, W. (2008) Extrinsic fluorescent dyes as tools for protein characterization. *Pharmaceutical research* **25**, 1487-1499

225. Cardamone, M., and Puri, N. K. (1992) Spectrofluorimetric assessment of the surface hydrophobicity of proteins. *Biochem J* **282 (Pt 2)**, 589-593

226. Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas, A. F., and Gilmanshin, R. I. (1991) Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers* **31**, 119-128

227. Booth, D. R., Sunde, M., Bellotti, V., Robinson, C. V., Hutchinson, W. L., Fraser, P. E., Hawkins, P. N., Dobson, C. M., Radford, S. E., Blake, C. C., and Pepys, M. B. (1997) Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature* **385**, 787-793

228. Naiki, H., Higuchi, K., Hosokawa, M., and Takeda, T. (1989) Fluorometric determination of amyloid fibrils in vitro using the fluorescent dye, thioflavin T1. *Anal Biochem* **177**, 244-249

229. LeVine, H., 3rd. (1999) Quantification of beta-sheet amyloid fibril structures with thioflavin T. *Methods Enzymol* **309**, 274-284

230.   Pedersen, J. S., Dikov, D., Flink, J. L., Hjuler, H. A., Christiansen, G., and Otzen, D. E. (2006) The changing face of glucagon fibrillation: structural polymorphism and conformational imprinting. *J Mol Biol* **355**, 501-523

231.   Krebs, M. R., Bromley, E. H., and Donald, A. M. (2005) The binding of thioflavin-T to amyloid fibrils: localisation and implications. *J Struct Biol* **149**, 30-37

232.   Fodera, V., Librizzi, F., Groenning, M., van de Weert, M., and Leone, M. (2008) Secondary nucleation and accessible surface in insulin amyloid fibril formation. *The journal of physical chemistry. B* **112**, 3853-3858

233.   Ohhashi, Y., Kihara, M., Naiki, H., and Goto, Y. (2005) Ultrasonication-induced amyloid fibril formation of beta2-microglobulin. *J Biol Chem* **280**, 32843-32848

234.   Chatani, E., Lee, Y. H., Yagi, H., Yoshimura, Y., Naiki, H., and Goto, Y. (2009) Ultrasonication-dependent production and breakdown lead to minimum-sized amyloid fibrils. *Proc Natl Acad Sci U S A* **106**, 11119-11124

235.   So, M., Yagi, H., Sakurai, K., Ogi, H., Naiki, H., and Goto, Y. (2011) Ultrasonication-dependent acceleration of amyloid fibril formation. *J Mol Biol* **412**, 568-577

236.   Yagi, H., Hasegawa, K., Yoshimura, Y., and Goto, Y. (2013) Acceleration of the depolymerization of amyloid beta fibrils by ultrasonication. *Biochim Biophys Acta* **1834**, 2480-2485

237.   Kitayama, H., Yoshimura, Y., So, M., Sakurai, K., Yagi, H., and Goto, Y. (2013) A common mechanism underlying amyloid fibrillation and protein crystallization revealed by the effects of ultrasonication. *Biochim Biophys Acta* **1834**, 2640-2646

238.   Yoshimura, Y., So, M.,, Yagi, H., and Goto, Y. (2013) Ultrasonication: An Efficient Agitation for Accelerating the Supersaturation-Limited Amyloid Fibrillation of Proteins. *Japanese Journal of Applied Physics* **52**, 07HA01-07HA01-08

239.   Doty, P., and Steiner, R. F. (1950) Light Scattering and Spectrophotometry of Colloidal Solutions. *J. Chem. Phys.* **18**, 1211–1220.

240.   Cole, J. L., Lary, J. W., Moody, T. P., and Laue, T. M. (2008) Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium. *Methods in cell biology* **84**, 143-179

241.   Brown, P. H., and Schuck, P. (2006) Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation. *Biophys J* **90**, 4651-4661

242.   Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* **78**, 1606-1619

243.   Dworkin, L. D., and Cooper, C. J. (2009) Clinical practice. Renal-artery stenosis. *N Engl J Med* **361**, 1972-1978

244.   Lashuel, H. A., and Wall, J. S. (2005) Molecular electron microscopy approaches to elucidating the mechanisms of protein fibrillogenesis. *Methods Mol Biol* **299**, 81-101

245.   Muller, D. J., Buldt, G., and Engel, A. (1995) Force-induced conformational change of bacteriorhodopsin. *J Mol Biol* **249**, 239-243

246.   Czajkowsky, D. M., Iwamoto, H., and Shao, Z. (2000) Atomic force microscopy in structural biology: from the subcellular to the submolecular. *Journal of electron microscopy* **49**, 395-406

247.   Schabert, F. A., Henn, C., and Engel, A. (1995) Native Escherichia coli OmpF porin surfaces probed by atomic force microscopy. *Science* **268**, 92-94

248.   Gosal, W. S., Myers, S. L., Radford, S. E., and Thomson, N. H. (2006) Amyloid under the atomic force microscope. *Protein Pept Lett* **13**, 261-270

249.   Sattler, M., Schleucher, J., and Griesinger, C.,. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy* **34**, 93–158

250.   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* **59**, 687-696

251.   Meller, J. (2001) Molecular Dynamics. Nature Publishing Group

252.   Karplus, M., and Kuriyan, J. (2005) Molecular dynamics and protein function. *Proc Natl Acad Sci U S A* **102**, 6679-6685

253.   Karplus, M., and McCammon, J. A. (2002) Molecular dynamics simulations of biomolecules. *Nature structural biology* **9**, 646-652

254.   Scheraga, H. A., Khalili, M., and Liwo, A. (2007) Protein-folding dynamics: overview of molecular simulation techniques. *Annual review of physical chemistry* **58**, 57-83

255.   Guex, N., Diemand, A., and Peitsch, M. C. (1999) Protein modelling for all. *Trends Biochem Sci* **24**, 364-367

256.   Case, D. A., Darden. T. A., Cheatham, T. E., Simmerling, C., Wang, J., Duke, R. E., Luo, R., Walker, R. C., Zhang, W., Merz, K. M., Roberts, B., Hayik, S., Roitberg, A., Seabra, G., Swails, J., Goetz, A. W., Kolossvary, I., Wong, K. F., Paesani, F., Vanicek, J., Wolf, R. M., Liu, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M. J., Cui, G., Roe, D. R., Mathews, D. H., Seetin, M. G., Salomon- Ferrer, R., Sagui ,C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A., and Kollman, P. A. . (2012) Amber12. University of California, San Francisco,

257.   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

258.   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. *The method. J. Comput. Chem* **13**, 1011-1021

259.   Kodandapani, R., Veerapandian, B., Kunicki, T. J., and Ely, K. R. (1995) Crystal structure of the OPG2 Fab. An antireceptor antibody that mimics an RGD cell adhesion site. *J Biol Chem* **270**, 2268-2273

260.   Chiti, F., Taddei, N., Baroni, F., Capanni, C., Stefani, M., Ramponi, G., and Dobson, C. M. (2002) Kinetic partitioning of protein folding and aggregation. *Nature structural biology* **9**, 137-143

261.   Johnson, G., and Wu, T. T. (2000) Kabat database and its applications: 30 years after the first variability plot. *Nucleic Acids Res* **28**, 214-218

262.   Lefranc, M. P. (2003) IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res* **31**, 307-310

263.   Chacko, S., Padlan, E. A., Portolano, S., McLachlan, S. M., and Rapoport, B. (1996) Structural studies of human autoantibodies. Crystal structure of a thyroid peroxidase autoantibody Fab. *J Biol Chem* **271**, 12191-12198

264.   Chiba, T., Hagihara, Y., Higurashi, T., Hasegawa, K., Naiki, H., and Goto, Y. (2003) Amyloid fibril formation in the context of full-length protein: effects of proline mutations on the amyloid fibril formation of beta2-microglobulin. *J Biol Chem* **278**, 47016-47024

265.   Faber, C., Shan, L., Fan, Z., Guddat, L. W., Furebring, C., Ohlin, M., Borrebaeck, C. A., and Edmundson, A. B. (1998) Three-dimensional structure of a human Fab with high affinity for tetanus toxoid. *Immunotechnology : an international journal of immunological engineering* **3**, 253-270

266.   Wu, T. T., and Kabat, E. A. (1970) An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J Exp Med* **132**, 211-250

267.   Mukherjee, S., Pondaven, S. P., and Jaroniec, C. P. (2011) Conformational flexibility of a human immunoglobulin light chain variable domain by relaxation dispersion nuclear magnetic resonance spectroscopy: implications for protein misfolding and amyloid assembly. *Biochemistry* **50**, 5845-5857

268.   Pace, C. N. (1990) Measuring and increasing protein stability. *Trends Biotechnol* **8**, 93-98

269.   Billings, K. S., Best, R. B., Rutherford, T. J., and Clarke, J. (2008) Crosstalk between the protein surface and hydrophobic core in a core-swapped fibronectin type III domain. *J Mol Biol* **375**, 560-571

270.   Benitez-Cardoza, C. G., Stott, K., Hirshberg, M., Went, H. M., Woolfson, D. N., and Jackson, S. E. (2004) Exploring sequence/folding space: folding studies on multiple hydrophobic core mutants of ubiquitin. *Biochemistry* **43**, 5195-5203

271.   Ewert, S., Huber, T., Honegger, A., and Pluckthun, A. (2003) Biophysical properties of human antibody variable domains. *J Mol Biol* **325**, 531-553

272.   Bellotti, V., Mangione, P., and Merlini, G. (2000) Review: immunoglobulin light chain amyloidosis--the archetype of structural and pathogenic variability. *J Struct Biol* **130**, 280-289

273.   Jones, S., Smith, D. P., and Radford, S. E. (2003) Role of the N and C-terminal strands of beta 2-microglobulin in amyloid formation at neutral pH. *J Mol Biol* **330**, 935-941

274.   Richardson, J. S., and Richardson, D. C. (2002) Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc Natl Acad Sci U S A* **99**, 2754-2759

275.   Monsellier, E., and Chiti, F. (2007) Prevention of amyloid-like aggregation as a driving force of protein evolution. *EMBO Rep* **8**, 737-742

276.   Nowak, M. (2004) Immunoglobulin kappa light chain and its amyloidogenic mutants: a molecular dynamics study. *Proteins* **55**, 11-21

277. Dudgeon, K., Rouet, R., Kokmeijer, I., Schofield, P., Stolp, J., Langley, D., Stock, D., and Christ, D. (2012) General strategy for the generation of human antibody variable domains with increased aggregation resistance. *Proc Natl Acad Sci U S A* **109**, 10879-10884

278. Zanetti, M., and Capra, D. (ed) (2005) *The Antibodies*, Vol. 5, harwood academic publishers

279. Canet, D., Last, A. M., Tito, P., Sunde, M., Spencer, A., Archer, D. B., Redfield, C., Robinson, C. V., and Dobson, C. M. (2002) Local cooperativity in the unfolding of an amyloidogenic variant of human lysozyme. *Nature structural biology* **9**, 308-315

280. Hammarstrom, P., Jiang, X., Hurshman, A. R., Powers, E. T., and Kelly, J. W. (2002) Sequence-dependent denaturation energetics: A major determinant in amyloid disease diversity. *Proc Natl Acad Sci U S A* **99 Suppl 4**, 16427-16432

281. Wetzel, R. (1994) Mutations and off-pathway aggregation of proteins. *Trends Biotechnol* **12**, 193-198

282. O'Donnell, C. W., Waldispuhl, J., Lis, M., Halfmann, R., Devadas, S., Lindquist, S., and Berger, B. (2011) A method for probing the mutational landscape of amyloid structure. *Bioinformatics* **27**, i34-42

283. Alberti, S., Halfmann, R., King, O., Kapila, A., and Lindquist, S. (2009) A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell* **137**, 146-158

284. Lei, H., Wu, C., Wang, Z., and Duan, Y. (2006) Molecular dynamics simulations and free energy analyses on the dimer formation of an amyloidogenic heptapeptide from human beta2-microglobulin: implication for the protofibril structure. *J Mol Biol* **356**, 1049-1063

285. Zheng, J., Ma, B., Tsai, C. J., and Nussinov, R. (2006) Structural stability and dynamics of an amyloid-forming peptide GNNQQNY from the yeast prion sup-35. *Biophys J* **91**, 824-833

286. Reddy, G., Straub, J. E., and Thirumalai, D. (2009) Dynamics of locking of peptides onto growing amyloid fibrils. *Proc Natl Acad Sci U S A* **106**, 11948-11953

287. Valleix, S., Gillmore, J. D., Bridoux, F., Mangione, P. P., Dogan, A., Nedelec, B., Boimard, M., Touchard, G., Goujon, J. M., Lacombe, C., Lozeron, P., Adams, D., Lacroix, C., Maisonobe, T., Plante-Bordeneuve, V., Vrana, J. A., Theis, J. D., Giorgetti, S., Porcari, R., Ricagno, S., Bolognesi, M., Stoppini, M., Delpech, M., Pepys, M. B., Hawkins, P. N., and Bellotti, V. (2012) Hereditary systemic amyloidosis due to Asp76Asn variant beta2-microglobulin. *N Engl J Med* **366**, 2276-2283

288. Bouchard, M., Zurdo, J., Nettleton, E. J., Dobson, C. M., and Robinson, C. V. (2000) Formation of insulin amyloid fibrils followed by FTIR simultaneously with CD and electron microscopy. *Protein Sci* **9**, 1960-1967

289. Pedersen, J. S., Christensen, G., and Otzen, D. E. (2004) Modulation of S6 fibrillation by unfolding rates and gatekeeper residues. *J Mol Biol* **341**, 575-588

290. Chow, M. K., Ellisdon, A. M., Cabrita, L. D., and Bottomley, S. P. (2004) Polyglutamine expansion in ataxin-3 does not affect protein stability: implications for misfolding and disease. *J Biol Chem* **279**, 47643-47651

291. Haspel, N., Zanuy, D., Ma, B., Wolfson, H., and Nussinov, R. (2005) A comparative study of amyloid fibril formation by residues 15-19 of the human calcitonin hormone: a single beta-sheet model with a small hydrophobic core. *J Mol Biol* **345**, 1213-1227

292. Lipfert, J., Franklin, J., Wu, F., and Doniach, S. (2005) Protein misfolding and amyloid formation for the peptide GNNQQNY from yeast prion protein Sup35: simulation by reaction path annealing. *J Mol Biol* **349**, 648-658

293. Michelitsch, M. D., and Weissman, J. S. (2000) A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions. *Proc Natl Acad Sci U S A* **97**, 11910-11915

294. Eisenberg, D., and Jucker, M. (2012) The amyloid state of proteins in human diseases. *Cell* **148**, 1188-1203

295. Sambashivan, S., and Eisenberg, D. (June 2006) Recent advances in decoding the structure of amyloid fibril. *MOLECULAR STRUCTURE as published in BTi*

296. Schroeder, H. W., Jr. (2006) Similarity and divergence in the development and expression of the mouse and human antibody repertoires. *Dev Comp Immunol* **30**, 119-135

297. Kihara, M., Chatani, E., Sakai, M., Hasegawa, K., Naiki, H., and Goto, Y. (2005) Seeding-dependent maturation of beta2-microglobulin amyloid fibrils at neutral pH. *J Biol Chem* **280**, 12012-12018

298. Strub, C., Alies, C., Lougarre, A., Ladurantie, C., Czaplicki, J., and Fournier, D. (2004) Mutation of exposed hydrophobic amino acids to arginine to increase protein stability. *BMC biochemistry* **5**, 9

299.   Pace, C. N., Grimsley, G. R., and Scholtz, J. M. (2009) Protein ionizable groups: pK values and their contribution to protein stability and solubility. *J Biol Chem* **284**, 13285-13289

300.   Trevino, S. R., Scholtz, J. M., and Pace, C. N. (2007) Amino acid contribution to protein solubility: Asp, Glu, and Ser contribute more favorably than the other hydrophilic amino acids in RNase Sa. *J Mol Biol* **366**, 449-460

301.   Gitlin, I., Carbeck, J. D., and Whitesides, G. M. (2006) Why are proteins charged? Networks of charge-charge interactions in proteins measured by charge ladders and capillary electrophoresis. *Angewandte Chemie* **45**, 3022-3060

302.   Shi, J., Guan, J., Jiang, B., Brenner, D. A., Del Monte, F., Ward, J. E., Connors, L. H., Sawyer, D. B., Semigran, M. J., Macgillivray, T. E., Seldin, D. C., Falk, R., and Liao, R. (2010) Amyloidogenic light chains induce cardiomyocyte contractile dysfunction and apoptosis via a non-canonical p38alpha MAPK pathway. *Proc Natl Acad Sci U S A* **107**, 4188-4193

303.   Keeling, J., Teng, J., and Herrera, G. A. (2004) AL-amyloidosis and light-chain deposition disease light chains induce divergent phenotypic transformations of human mesangial cells. *Lab Invest* **84**, 1322-1338

304.   Teng, J., Turbat-Herrera, E. A., and Herrera, G. A. (2014) An animal model of glomerular light-chain-associated amyloidogenesis depicts the crucial role of lysosomes. *Kidney Int* **86**, 738-746

305.   Diomede, L., Rognoni, P., Lavatelli, F., Romeo, M., del Favero, E., Cantu, L., Ghibaudi, E., di Fonzo, A., Corbelli, A., Fiordaliso, F., Palladini, G., Valentini, V., Perfetti, V., Salmona, M., and Merlini, G. (2014) A Caenorhabditis elegans-based assay recognizes immunoglobulin light chains causing heart amyloidosis. *Blood* **123**, 3543-3552

PUBLICATION

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. *JOURNAL OF BIOLOGICAL CHEMISTRY* **VOL. 289, NO. 39**, pp. 26829–26846

ACKNOWLEDGEMENTS

This doctoral thesis would not have been possible without the contribution of many great scientists. Therefore I would like to express my gratitude to the following people:

First and most importantly I would like to thank my supervisor Prof. Dr. Johannes Buchner for the exciting and stimulating antibody project, for providing continued support and guidance throughout the project and thesis preparation.

For their respective contribution to this work, I would like to thank Prof. Dr. Bernd Reif and his graduate student Manuel Hora for conducting the NMR experiments, Prof. Dr. Martin Zacharias, of the Department of Physics, Technische Universität München for performing the MD simulations, and Prof. Dr. Yuji Goto and his postdoc Dr. Hisashi Yagi, of the Protein Research Institute, Osaka University, Japan, for performing the amyloid induction assays by ultrasonication. Collaborating with you all was very fruitful and a great experience. I also would like to thank Bettina Richter for her tremendous assistance with the electron microscopy measurements.

I wish to express my sincere gratitude to the formal and current members of the Antibody subgroup and the Chair of Biotechnology who I got to know. Working with you all taught me a lot over the last years. To my Bachelor and Masters students, most especially Susanne Pettinger, I say thank you for the time spent and excellent work done.

I am also very grateful to the International Max Planck Research School for Life Sciences (IMPRS-LS) Munich, through which I joined the Buchner-lab, for their excellent curricular and extracurricular activities. I am thankful to the IMPRS-LS coordinators and to my Thesis Advisory Committee members; Prof. Dr. Michael Sattler, Dr. Marc Schmidt-Supprian and Prof. Dr. Bernd Reif for their support and guidance through this work.

Good science with great scientist alone would not have been sufficient for the successful accomplishment of this thesis. Many friends both from and outside the scientific environment, in many different ways, contributed enormously to this success. To the great friends I got to meet via the IMPRS network; Dr. Manal Chatila-Schiffer and husband Dr. Niclas Schiffer, Shoh Asano and wife Dr. Ida Pavlichenko, Nikolas Schrod (Niki) and wife Alexa Schrod, Dhawal Jain,...and the list continues, the numerous "Drinks" times and other social activities helped relax the brain all along for the good science – thank you and you will forever be in my memory.

Special thanks also to my many other admirable friends and well-wishers; in alphabetical order: Dr. Babila Tachu ("boy", for always being ready to help + all the good times in Berlin and around Germany), Ernest Ekwende ("Ernesto"…ahaha, no! u catch? for the good times men), Dr. Daniel Richart ("Mr. D", for the memorable and the great times in Munich), Delphine Nnane ("Able Grande soeur"…ahaha, for the good times and meals), George E. Nkumbe ("Georges"… u know nooh…list too long..ahaha), Ivo Sumelong ("Grand Ivoooo…", for the great times, inspirational discussions, arguments and many more (a long list)), Justus Metuge ("Grand Jus", for encouragement), Ngome Njume ("Grand Njums") and wife Lynda Ekundime-Njume ("Lyndos", for the delicious meals that sometimes provided the required energy for survival), Omar Ekine ("O'boy"…u know nooh,..ahah for all the enjoyable moments), Roland Ediage ("Ma Grand"…drinks…ahaha + the good times). To some

on the list, out of sight is not out of mind, at some point during this journey you played a very instrumental role, which was highly appreciated and will never be forgotten.

I will also like to express my gratefulness to Katalin Liebert who made it easy to come and study here in Germany and to Prosper E. Nokwe, for initiating this, added to his moral support all along.

Special thanks to my girlfriend Dr. Giulia Ossato (Sweetie) for her patience, love and assistance.

Finally, I want to express my love and deepest gratitude to my parents for their constant, unconditional, heartfelt and enduring support in all life circumstances. I am sure it was one of your greatest wishes and now the pride to see your "little A'Ngi" at this level.

CURRICULUM VITAE

Cardine NOKWE NKUMBE	
	Date and Place of Birth: January 11, 1983, Bangem
	Nationality: German
	Address: Ungererstrasse. 19, D-80802 Munich, Germany

EDUCATION

05/2010 – present	**Dissertation/Ph.D. project** International Max Planck Research School for Life sciences /Technische Universität München (TUM) - Munich Department of Chemistry, chair of Biotechnology Supervisor: Prof. Dr. Johannes Buchner Project: Molecular determinants and mechanisms of antibody light chain amyloidosis Award of Ph.D. degree: expected December 2014
10/2007 – 11/2009	**M.Sc. Studies of Molecular Medicine** Charité-Universitätsmedizin – Berlin Faculty of Medicine, Department of Molecular Medicine Award of M.Sc. degree in Molecular Medicine
10/2004 – 10/2007	**B.Sc. Studies of Applied Biology** Bonn-Rhein-Sieg University of Applied sciences – Rheinbach/Bonn Faculty of Applied Natural Sciences Award of B.Sc. degree in Applied Biology
10/2006 – 03/2007	**Exchange scholar within the ERASMUS exchange program** Jagiellonian University - Cracow, Poland Faculty of Biochemistry, Biophysics and Medical Biotechnology

ACADEMIC AND INDUSTRIAL RESEARCH EXPERIENCE

02/2010	**Research internship** Charité-Universitätsmedizin – Berlin Department of Glycodesign and Glycoanalysis Supervisor: Dr. Markus Berger Project: Optimization of an *N*-linked glycoanalytical scheme
10/2009 – 01/2010	**Research internship** Max Planck Institute for Colloids & Interfaces (MPIKG) –Berlin Department of Biomolecular systems Supervisors: Prof. Dr. Peter H. Seeberger/Dr. Faustin Kamena Project: *In vitro* glycosylation of IL-6 protein by semi-enzymatic method
04/2009 – 10/2009	**Master's thesis** Max Planck Institute for Infection Biology (MPI-IB) – Berlin Molecular Immunology Group Supervisor: Dr. Hedda Wardemann Project: Reactivity Analysis of Murine Peritoneal Cavity B cells

Focusing on the Canonical VH11/Vκ9 BCR

10/2007 – 08/2009	**Part-time employment**
	Biopharma-ProBioGen AG – Berlin
	Department of Quality Control (QC), Cells and Tissue Service (CaTS)
	Task: In charge of glycosylation analyses of therapeutic antibodies, under GMP guidelines
04/2007 – 10/2007	**Bachelor's thesis**
	Biopharma-ProBioGen AG – Berlin
	Department of Quality Control (QC), Cells and Tissue Service (CaTS)
	Supervisor: Prof. Ulrike Bartz /Dr. Stefan Zietze
	Project: Optimisation of a Scheme for *N*-Linked Glycosylation Analysis
10/2006 – 03/2007	**Research Scholar within the ERASMUS exchange program**
	Jagiellonian University - Cracow, Poland
	Faculty of Biochemistry, Biophysics and Medical Biotechnology
	Supervisor: Dr. Joanna Bereta
	Project: production of monoclonal antibodies by hybridoma technique

SCIENTIFIC RELATED ACTIVITIES

Memberships	German Pharmaceutical Society (DPhG) (2007 – 2010
	German Society of Biochemistry and Molecular Biology (GBM) 2010-213
Conference attendances with presentation	▪ Annual meeting of the German Consortium for amyloidosis (GERAMY), Kiel, Germany. (Nov 07 – 08, 2014). **Talk:** Mechanisms of antibody light chain amyloid formation
	▪ Annual meeting of the German Consortium for amyloidosis (GERAMY), Munich, Germany. (Nov 29 – 30, 2013). **Talk:** Mutations influencing antibody V_L domain stability and amyloid formation
	▪ Protein folding symposium (Faltertage), Regensburg, Germany, (Oct 18 – 20, 2013). **Talk:** Pivotal role of residue 2 for the integrity of antibody variable domains
	▪ Endoplasmic Reticulum (ER) conference, University of Warwick, Coventry, United Kingdom. (April 25 – 27, 2013). **Poster**: Restoring stability and preventing amyloid formation in an antibody V_L domain
	▪ Annual meeting of the German Pharmaceutical Society (DPhG), Bonn, Germany, (Oct 08 – 11, 2008). **Poster:** Examination of glycoengineering efficiencies of a model glycoprotein in two different bioreactor systems

PUBLICATION

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. *JOURNAL OF BIOLOGICAL CHEMISTRY,* VOL. 289, NO. 39, pp. 26829–26846

DECLARATION

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

Hiermit erkläre ich, Cardine NOKWE NKUMBE, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Die Arbeit wurde bisher keiner Prüfungskommission vorgelegt. Teile dieser Arbeit wurde in wissenschaftlichen Journalen veröffentlicht.

München, den 15.10.2014   _____
                                                            Cardine NOKWE NKUMBE