A computational, biochemical, and cellular approach to explore the aggregation mechanisms underlying Immunoglobulin light chain amyloidosis

Thesis submitted in accordance with the requirements of the

University of Liverpool

for the degree of

Doctor in Philosophy

By Kieran Hand

September 2017

Abstract

To date, over 36 confirmed proteins are implicated in 50 known degenerative disorders including Alzheimer's, Parkinson's, type 2 diabetes, Huntington's and light chain associated diseases. These debilitating disorders, collectively termed the "amyloidoses" arise due to the misfolding of an otherwise native protein. Amyloid formation sees the loss of a protein's native state and adoption of aberrant conformations which result in the accumulation of insoluble fibrils that possess a highly ordered ultrastructure rich in β-sheet. The most common form of all systemic amyloidosis, which refers to the extracellular accumulation and deposition of the misfolded precursor protein at locations distant from the site of production, is Immunoglobulin (Ig) light chain (AL) amyloidosis with an incidence of 1 in every 100,000. In AL amyloidosis, patients diagnosed with a plasma cell dyscrasia show a significantly elevated population of intact or truncated Immunoglobulin light chains (LC) in their circulation. A proportion of these LC are able to adopt pathological conformations and accumulate as fibrillar aggregates causing irreversible damage to virtually all organs and tissues. Each patient presents a light chain with a unique amino acid sequence. This makes identifying a common theme that underlies AL amyloidosis challenging. A small number of publications place focus on SMA, REC and LEN, three 114 amino acid light-chain variable domains of the kappa 4 IgG family. Despite high sequence identity between these three homologs, SMA, and REC are amyloidogenic *in vivo,* but LEN adopts a stable dimer, displaying amyloidogenic properties only under destabilising conditions. In this study, we adopt a computational, biochemical, and cellular approach to explore the aggregation mechanisms

underlying Immunoglobulin light chain amyloidosis and use these three human variable domains as a model to address some of the questions surrounding light chain amyloidosis.

Previous structural analysis indicates that both SMA and LEN form dimers, however the quaternary arrangement of SMA is unknown. We report that SMA has an altered interface that is likely to be rotated 180°. Free energy calculations of the LEN dimer interface indicate that two SMA-like mutations, Q89H and Y96Q play a role in protein V_L dimer stability through destabilising the quaternary structural arrangement. In addition, structure-based pocket detection of this model reveals a cavity suitable for accommodating a small-molecule designed to inhibit dimer dissociation into aggregation prone monomers. We also combine NMR assessment of millisecond timescale protein motions with atomic-level level structural assessment to provide a rationale to how a particular mutation leads to changes in thermodynamic stability and amyloidogenicity where we show the CDR3 loop a key structural region in modulating V_L stability. Overall, this work supports the data that disruption of the dimer interface is needed for amyloid to occur, but also addresses the added complication that thermodynamic stability is not an accurate indicator of aggregation potential.

We also present here a refined and reproducible periplasmic expression and purification protocol for SMA and LEN that improves on existing methods and provides high yields of pure protein, particularly suitable for structural studies that demand highly concentrated and pure proteins. We confirm that recombinant SMA and LEN proteins have structure

and dimerization capability consistent with the native proteins and employ fluorescence to probe internalisation and cellular localisation within cardiomyocytes. Internalisation of both SMA and LEN as monomers was observed in cell culture after 24 hours of incubation where the amyloidogenic V_L SMA exhibited cytotoxic effects. We propose periplasmic expression and simplified chromatographic steps outlined within as an optimised method for production of these and other variable light chain proteins to investigate the underlying mechanisms of light chain amyloidosis.

Abbreviations

Aβ Amyloid βeta
AL Light chain amyloidosis
AFM Atomic force microscopy
ANOVA Analysis of variance
AUC Analytical ultra-centrifugation
BMRB Biological Magnetic Resonance Bank
BNP (brain natriuretic peptide),
CCS Collisional cross section
CD Circular dichroism
CDR Complementarity-determining region
C.elegans *Caenorhabditis elegans*
CMR Cardiac magnetic resonance
COSMiCS (Complex Objective Structural analysis of Multi-Component Systems)
CPHPC((2R)−1-[6-[(2R)−2 carboxypyrrolidin-1-yl]−6-oxohexanoyl pyrrolidine-2-corboxylic acid)
CV Column volume
DMSO Dimethyl sulfoxide
DMEM Dulbecco's Modified Eagle's medium
ECL Enhanced chemiluminescence
ECHO Echocardiography
EGCG (−)-epigallocatechin 3-gallate
EMA Europeans medicine agency
Fab antigen-binding fragment
FBS Fetal bovine serum
FDA food and drug administration
Fc Crystallisable fragment
FISH Fluorescent in situ hybridization
FITC Fluorescein isothiocyanate
FLC Free light chain
FR Framework region
FRET Fluorescence resonance energy transfer
FPLC Fast protein liquid chromatography
GAG Glycosaminoglycan
GAPDH Glyceraldehyde 3-phosphate dehydrogenase
GPa Gigapascal
HC Heavy chain
hCF Human Cardiac Fibroblasts (HCF)
HRP horseradish peroxidase
HSQC Heteronuclear single quantum coherence spectroscopy
Ig Immunoglobulin
IPTG Isopropyl β-D-1-thiogalactopyranoside
ISA International society of amyloidosis
LC Light chain
MALLS Multi-angle laser light scattering

MAPK Mitogen-activated protein kinases
MCWO Molecular weight cut off
MD Molecular dynamics
MDex Dexamethasone
MGUS monoclonocal gammopathy of undetermined significance
MG132 Carbobenzoxy-Leu-Leu-leucinal
MM Multiple myeloma
MS Mass spectrometry
ms millisecond
NMR Nuclear magnetic resonance
Nrf2 Nuclear-factor-E2-related-factor-2
NT-proBNP N-terminal pro-brain natriuretic peptide
PBS Phosphate buffered saline
PDB Protein data bank
PCQ Protein quality control
PCR Polymerase chain reaction
PFA paraformaldehyde
POI Protein of interest
PVDF Polyvinylidene difluoride
RMS Root mean square
RMSF Root mean square fluctuation
RP-HPLC Reversed-phase high-performance liquid chromatography
SAXS Small Angle X-ray Scattering
SAP Serum albumin protein
SDS-PAGE Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC Size exclusion chromatography
SOD1 Superoxide dismutase 1
TBS Tris buffered saline
TEM Transmission electron microscopy
T_m melting temperature
ThT Thioflavin T
TTR transthyretin
Vdw Van de Waals force
V_H Heavy chain variable domain
V_L Light chain variable domain

Contents

Acknowledgements

I am truly grateful to my supervisor Dr Jillian Madine (and for Dr David Middleton and Dr David Smith for putting in a good word!) for making the call and giving me the opportunity to complete my studies within the Amyloid team at Liverpool.

I would like to acknowledge the members of the Amyloid group stationed at the University of Liverpool institute of integrative biology. These include (past and present) Chris Longmore, James Torpey, Dr Hannah Davies, and Dr Marie Phelan. I would also like to extend this thanks to Dr Robert Gibson, who became an honorary member of the group in my Masters and first year of PhD studies.

I would also like to thank Dr Mark Wilkinson (for just about everything), Dr Edwin Yates (for long conversations on Terahertz, the phylum Tardigrada) Dr Mal Horsburgh, Dr Deborah Hogg, Dr Svetlana Antonyuk, Dr Gareth Wright, Dr Chris Hill, Dr Mike Speed, Dr Daniel Rigden, Edward Humphries, Francesca De Faveri, Dr Dominic Byrne, Dr Alistair Fielding, Professor Claire Eyers, Dr Patrick Eyers and Dr Caroline Dart (my internal assessors), Dr Ewan Blanche, Prof Roy Goodacre (for both giving me excellent advice when applying for PhD projects), Professor Paul Popelier, Dr Diana Penha (Liverpool Heart and Chest), Dr Parveen Sharma, Dr Daniel Antoine, and Dr David Jones.

I would like to thank the BBSRC for funding this project.

Some of the work presented within thesis has been published

Effect of amino acid mutations on the conformational dynamics of amyloidogenic immunoglobulin light-chains: A combined NMR and *in silico* study

Sujoy Mukherjee, Simon P. Pondaven, Kieran Hand, Jillian Madine & Christopher P. Jaroniec

Nature scientific reports, published 04 September 2017

1. Introduction

1.1. Overview

Proteins comprise a large group of biological molecules that perform an array of essential processes in every biological system (Lesk 2004). Genetic data stored in the form of nucleic acids, contain all the necessary information for the ribosome to synthesise proteins from a library of just 20 essential amino acids (Hooper 2000). This seemingly insignificant number nevertheless lays the foundation for nature to engineer an overwhelming number of proteins with unique three-dimensional shapes. The shape, or tertiary structure also determines the biological function in a link termed the structure-function relationship (Lesk 2004) and makes each protein perfectly suited to perform a specific task (Bruce, Alexander et al. 2002). Such diversity has also made proteins important targets for therapeutics, with a large proportion of all drugs currently available targeting just one unique class of transmembrane proteins (Garland 2013, Cvicek, Goddard et al. 2016).

Over the past decade, an area of research that has gained much interest is the study of protein folding. This interest stems from the link between protein misfolding, aggregation and many conformational disorders that include Parkinson’s, Alzheimer's, Huntington’s, and Type II diabetes (Soto 2001). The proteins implicated in these disorders have demonstrated the ability to depart from a soluble conformation, and adopt insoluble fibrillar conformations termed amyloid (Rambaran and Serpell 2008), which have become the hallmark for many of these diseases (Selkoe 2003). Many of the 36 known amyloid proteins (Sipe, Benson et al. 2016) implicated in human

diseases display minimal sequence identity and have large differences in fold (Stefani 2004) yet the ultrastructural and histological properties of amyloid fibrils, formed at the final stage of the fibrillation reaction are remarkably similar. Specifically, they share a common cross-β sheet organisation (Sipe and Cohen 2000) and the unique ability to bind the amyloidophillic dye Congo red (Westermark GT 1999). Both have become the characteristic features of amyloid and are used in their diagnosis. Advances in biochemical and biophysical instrumentation have been able to shed light on the complex aggregation process, revealing that misfolding follows an ordered multistep pathway involving monomer, oligomer, protofibrillar and mature fibrillar species (Sarroukh, Goormaghtigh et al. 2013, Woods, Radford et al. 2013, Karamanos, Kalverda et al. 2015, Verma, Vats et al. 2015, Wright and Dyson 2015, Herranz-Trillo, Groenning et al. 2017). However, many prefibrillar intermediates that populate this pathway remain largely uncharacterised due to their dynamic nature (Hubin, van Nuland et al. 2014) in solution and as a consequence, the cytotoxic potential and structure of such invisible conformers are unknown. Extensive research into α-synuclein and amyloid β (Aβ), the two precursor proteins implicated in the neurodegenerative diseases Parkinson's and Alzheimer's respectively (Wirths and Bayer 2003, Murphy and LeVine 2010) have established that an abundance of factors including, the composition of different molecular surfaces (Zhu, Souillac et al. 2002, Meng, Fink et al. 2008, Galvagnion, Buell et al. 2015), along with molecular (Beyer and Ariza 2013, Barrett and Timothy Greenamyre 2015), and environmental factors (Uversky, Li et al. 2001) are able to modulate, and trigger the aggregation process (Emamzadeh 2016). For other diseases such a light

chain (AL) amyloidosis, a fatal, and currently incurable condition with widespread organ damage caused by the misfolding and deposition of Immunoglobulin fragments, such an extensive set of factors has not been characterised to date. In AL amyloidosis each patient possesses a light chain with a unique amino acid composition that governs its stability. Collectively, these factors significantly complicate the study of AL amyloidosis. A better understanding of the transformation process, the cytotoxic potential and structure of pre fibrillar intermediates, including the factors that are able to initiate and modulate their aggregation is pertinent to understanding disease initiation, progression, and rational drug design with the ultimate goal of treating a wealth of conformational disorders that extend beyond AL.

This chapter first describes the basis of protein folding, protein self-association, and the characteristics of amyloid fibrils. The chapter then moves to the topic of AL including diagnosis, current and novel therapeutic strategies, before finally focussing on the light chain variable domains that are at the basis of this investigation.

1.2. Protein folding

Successful protein folding has traditionally been considered a requirement for a protein to be functional (Martin, Fau et al. 1998). In this process, chains emerging from the ribosome sample vast conformational space before they fold into a compact and stable three-dimensional organisation termed the native state (Lesk 2004). For much of the proteome, this highly dynamic event occurs cotranslationally (Fedorov and Baldwin 1997, Cabrita, Hsu et al. 2009). Techniques such as Fluorescence Resonance Energy Transfer (FRET) Microscopy and Nuclear Magnetic Resonance (NMR) spectroscopy have been

applied to study this fascinating process in real time (Holtkamp, Kokic et al. 2015, Cassaignau, Launay et al. 2016). Yet, despite such advances, the precise mechanisms of the folding process remain largely enigmatic. Indeed, this has been widely termed the “Protein folding problem” since the early 1960s (Dill, Ozkan et al. 2008, Dill and MacCallum 2012). Over time, several models of protein folding have emerged and each have built upon the early studies conducted by Christian B Anfinsen (Anfinsen and Haber 1961, Anfinsen 1973) and Cyrus Levinthal (Levinthal 1968). Both Anfinsen and Levinthal concluded that all the necessary information required for a protein to correctly assume a three-dimensional structure resides within the amino acid sequence, and that limitations on conformational freedom imposed by the main chain allowed for folding to occur on a biological timescale (Rose, Fleming et al. 2006).

Some of the traditional views of protein folding have however, been challenged with the prospect that a large proportion of the eukaryotic proteome may fulfil vital biological roles in an intrinsically disordered state (Wright and Dyson 2015, Vincent and Schnell 2016). The 140-amino acid protein α-synuclein for example, lacks any secondary structural elements yet has been suggested to play important presynaptic roles (Bendor, Logan et al. 2013, Burre 2015).

1.2.1. Protein folding - Funnel Energy landscapes

Two recent reviews by F. Ulrich Hartl *et al*. and Tuomas P. J. Knowles *et al*. detail the present understanding of protein folding, with regards to the complex energy landscape model, the formation of protein aggregates, and the importance of molecular chaperones in maintaining a functional proteome

(Hartl, Bracher et al. 2011, Knowles, Vendruscolo et al. 2014). F. Ulrich Hartl first highlights that, for a proportion of small protein molecules, folding is able to take place without energy cost or folding aids, relying solely, on the amino acid sequence to dictate the native state (Dobson, Šali et al. 1998, Rose, Fleming et al. 2006). It would seem that, for these proteins, a two-state folding model such as the idealised smooth energy landscape first proposed by Ken Dill (illustrated in Figure 1A) reflects the folding landscape for these proteins (Dill and Chan 1997, Horwich 2002) where point I is the unfolded state and point II is the folded state, and the protein does not adopt any intermediates between these two points. However, it has become clear that for much of the proteome, many proteins are unable to reach the native state without the presence of molecular chaperones. These observations, in combination with significant leaps made in theoretical methods of predicting protein folding has led to the rugged energy folding funnel depicted in Figure 1B (Bryngelson, Onuchic et al. 1995, Onuchic, Luthey-Schulten et al. 1997, Horwich 2002, Onuchic and Wolynes 2004, Dill and MacCallum 2012, Chen, Lu et al. 2017) to be the most widely accepted theory used to describe protein folding. In this model, an ensemble of unfolded polypeptides possessing both high energy and high hydration (Robinson and Cho 1999, Jerson and Debora 2009, Chen, Lu et al. 2017) collapse into the energy landscape shaped as a tapering funnel. As the unfolded chains proceed down multiple parallel pathways towards the native state, located at the bottom of the energy funnel, the flexible chain samples many different conformations and become progressively more organised (Rose, Fleming et al. 2006, Dill, Ozkan et al. 2008). The formation of favourable intramolecular contacts leads to secondary

structure formation such as the α-helix and β-sheet, which restricts the sampling space, and lowers the free energy (Srinivasan and Rose 1999, Chikenji, Fujitsuka et al. 2006, Hamelryck, Kent et al. 2006). This is indicated by a downwards arrow and narrowing of the energy funnel. Eventually, the chain packs into a compact, unstrained arrangement that possess both low free energy, and low hydration. This particular conformation is termed the native state, and represents a number of conformational sub-states rather than one particular conformation (Campioni, Monsellier et al. 2010). This process is often termed the “on-pathway” route of protein folding (Clark 2004) (Figure 1B left) and the entire process usually occurs on the microsecond to second timescale (Chen, Lu et al. 2017). The native fold is maintained and stabilised by many co-operative interactions that include hydrogen bonds, electrostatic interactions, van der Waals forces (VdW) and disulfide bonds (Lesk 2004, Yasuda, Yoshidome et al. 2010).

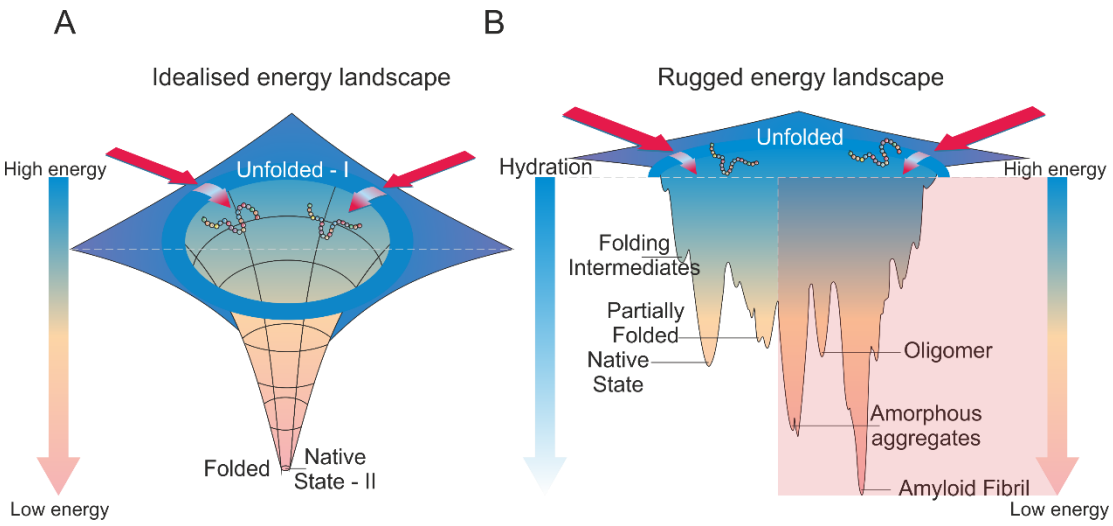


Figure 1 Schematic illustration of idealised (A) energy landscape and (B) rugged energy landscape models of protein folding. At the top of the funnel exists an ensemble of unfolded conformations which have both high free energy and high hydration (shown on the vertical axis of the funnel). Collapse of the polypeptide chain into the funnel shaped energy landscape is energetically driven, and as favourable intramolecular contacts form, the free energy drops. The idealised energy funnel (A) represents a two-state folding process during which unfolded chains rapidly adopt the native structure that represents that of the lowest energy located at the bottom of the funnel. In the rugged energy funnel model (B) the landscape is decorated with deep energy troughs termed local minima. Proteins can become kinetically trapped in this state which delays the time taken to reach the native state. Proteins can also access a region of the landscape that has low energy (highlighted red) which can lead to the formation of amorphous aggregates and amyloid assemblies. Folding energy funnel is based on that illustrated by F. Ulrich Hartl *et al.* and Arthur Howich (Horwich 2002, Hartl, Bracher et al. 2011).

Unique to the rugged energy model are secondary funnels that decorate the landscape (Figure 1B). These funnels represent deep energy minima, and are able to trap the protein in an intermediate molten globule state (existing between the unfolded and folded conformation of a protein) (Horwich 2002, Jahn and Radford 2008, Milanesi, Waltho et al. 2012). In this form, regions of the protein are more exposed to the solvent than in the native, folded state.

The exposure of short complementary amyloidogenic segments, hydrophobic patches, or particular residues that favour to β-sheet formation (aromatic and hydrophobic) (Del Pozo-Yauner, Becerril et al. 2015, Chen, Lu et al. 2017, Graña-Montes, Pujols-Pujol et al. 2017) which would normally be buried in the globular state, here, may become solvent-exposed (Goldschmidt, Teng et al. 2010, Graña-Montes, Pujols-Pujol et al. 2017). This can lead to undesirable intermolecular interactions (Turoverov, Kuznetsova et al. 2010, Brummitt, Andrews et al. 2012) where the protein is vulnerable to forming aggregated species that possess extremely low global free energy (Figure 1B shaded red) (Brummitt, Andrews et al. 2012, Knowles, Vendruscolo et al. 2014, Graña-Montes, Pujols-Pujol et al. 2017). These protein aggregates (illustrated in Figure 1B right side) are normally both non-functional but can be divided into two categories based on their morphology. The first, is amorphous aggregates that lack order and are generally considered non-toxic (Bieschke, Russ et al. 2010, Stefani and Rigacci 2013, Kumar, Namsechi et al. 2015). In stark contrast, proteins can form highly ordered pathogenic amyloid structures that are resistant to proteolytic degradation (Rambaran and Serpell 2008). This process of aggregation is often termed the “off-pathway” route of folding (Clark 2004).

As illustrated in Figure 1B, the amyloid structure has lower free energy than the native state of the protein. This is due to the greater stability that the cross β-sheet architecture (found within amyloid) provides (Graña-Montes, Pujols-Pujol et al. 2017). Yet, proteins generally avoid this state. The physico—chemical features of the protein chain and the presence of accessory molecules such as molecular chaperones (Del Pozo-Yauner,

Becerril et al. 2015, Graña-Montes, Pujols-Pujol et al. 2017) allow for proteins to fold correctly and reach their native state (Anfinsen 1973, Hartl, Bracher et al. 2011). An example of a small number of these accessory components include heat shock proteins 40, 60, 70, 90, and 100, co-chaperones, protein isomerases, a number of chaperone related-proteases, and the proteasome complex. Together, these comprise the highly coordinated cellular protein quality control (PQC) mechanism. (Gilbert 1997, Hinault, Ben-Zvi et al. 2006, Douglas, Summers et al. 2009, Chakraborty, Chatila et al. 2010, Voisine, Pedersen et al. 2010, Hartl, Bracher et al. 2011, Del Pozo-Yauner, Becerril et al. 2015). Dysfunction in anyone of these tightly regulated systems, as a result of age, or specific conditions where the rate of protein synthesis overwhelms the system, such in the case of plasma cell dyscrasias (later documented in this thesis) can lead to the undesirable intramolecular interactions that promote self-association, and the formation of toxic aggregates that can have devastating cellular consequences observed in amyloid diseases (Del Pozo-Yauner, Becerril et al. 2015).

1.3. Amyloid and Amyloidosis

Historically, the name amyloid, meaning “starch-like” was first used by botanist Matthias Jakob Schleiden in 1838 to name plant sections that demonstrated blue staining in a reaction using a solution of iodine and sulfuric acid (Rambaran and Serpell 2008). In 1854 Rudolph Virchow observed a similar staining pattern in the human nervous system (Cohen 1986). Both interpreted the positive reaction to suggest that the observed material was carbohydrate in composition (Sipe and Cohen 2000). A similar procedure later conducted by Friedrich and Kekulé on tissue from a liver biopsy disproved these early

observations, and found that the deposits were actually made of protein (Eberth 1881, Kisilevsky, Raimondi et al. 2016). Despite these initial inaccuracies, “amyloid” is used today to refer to the insoluble deposits of misfolded protein that have accumulated in the extracellular and, the more recently recognised; intracellular space of tissue (Selkoe 2003, Stefani 2004, Rambaran and Serpell 2008). The defining features of amyloid is its unique ability to bind both Congo red, a commonly used amyloidophilic dye, originally created for the textile industry (Kisilevsky, Raimondi et al. 2016) that presents an apple-green birefringence when observed under cross polarised light (Westermark GT 1999, Westermark 2005), and a cross β-spine arrangement when examined by X-ray fibre diffraction (Nelson, Sawaya et al. 2005, Eisenberg and Jucker 2012). The accumulation of amyloid in fibrillar deposits typically consists of the precursor protein in large quantities (95%), and a number of coagulation agents including glycosaminoglycans (GAGs), and serum albumin protein (SAP) (Sipe and Cohen 2000) (the remaining 5%) the removal of which, remain promising therapeutic strategies (Fikrle, Paleček et al. 2013, Gillmore and Hawkins 2013). Identification of amyloid fibrils in this state serves as the pathological hallmark for many well-known protein-misfolding disorders collectively termed “amyloidoses” (Lee, Lim et al. 2011).

From a historical perspective the ability for a protein to form amyloid was first thought to be accessible, only, to a small percentage of the proteome (Selkoe 2003). However, it is now clear that accessing the amyloid state is not limited merely by the amino acid composition of the precursor protein rather; adopting the β-sheet conformation, in which amyloid structures are enriched in, is a generic structure which is accessible to all proteins, and is mediated by

specific intermolecular backbone interactions formed by short complementary amino acid stretches within the protein rather than the entire protein chain (Esteras-Chopo, Serrano et al. 2005, Eisenberg and Jucker 2012, Graña-Montes, Pujols-Pujol et al. 2017). It is even suggested that, under specific conditions that favour partially unfolded states all proteins are capable of amyloid formation, and all that is necessary for this to occur is the exposure of complementary segments that allow for the formation of the amyloid spine (Eisenberg and Jucker , Brumshtein, Esswein et al. 2014, Graña-Montes, Pujols-Pujol et al. 2017).

Many proteins possess segments of high aggregation propensity and to date, the International Society of Amyloidosis recognises 36 proteins to be implicated in 50 known conformational disorders (Sipe, Benson et al. 2016). Amyloid deposition can be categorised as localised, affecting a specific tissue, or systemic, in which it affects multiple tissues and organs (Westermark 2012, Lin, Seldin et al. 2015). For an example, a selection of the known conformational disorders and related precursor protein are documented below in Table 1.

Table 1 Examples of amyloid related disease, the associated precursor protein and the site of deposition. The protein or precursor listed refers to the wild type. Information was adopted from amyloid nomenclature 2016 (Sipe, Benson et al. 2016)

DISEASE	PRECURSOR PROTEIN / PEPTIDE	SYSTEMIC / LOCAL
Parkinson's disease	α-synuclein	Local
Alzheimer's disease	amyloid beta (Aβ) wild-type	Local
AL (primary) amyloidosis	Immunoglobulin light chain	Systemic & Local
AA (secondary) amyloidosis	Immunoglobulin heavy chain	Systemic & Local
Type II diabetes	Amylin (Islet amyloid)	Local
Haemodialysis-related amyloidosis	β2 microglobulin	Local
Creutzfeldt-Jakob disease	Prion protein wild type	Local
Cystic Fibrosis	Transthyretin wild type	Systemic
Familial Amyloid Polyneuropathy	Transthyretin wild-type	Systemic

1.3.1. The Mechanisms of Amyloid Formation

The widely accepted hypothesis of amyloid formation suggests that conversion of the native precursor protein (found at the start of the reaction) into β-stranded fibrils (found at the end of the fibrillation reaction) occurs in a multi-step process (Figure 2) consisting of monomer, oligomer, protofibrillar and mature fibrillar species (Rochet and Lansbury 2000, Blancas-Mejía, Horn et al. 2015, Blancas-Mejia, Misra et al. 2017). Interestingly, many independent studies have revealed that despite the enormous differences in sequence identity between many precursor proteins, there are common similarities in some of the protein conformers that are found along the pathway *en route* to the formation of mature amyloid fibrils (Verma, Vats et al. 2015). There is also the possibility for the precursor protein to progress down "off-pathway" routes that result in the formation of amorphous aggregations rather than the formation of amyloid structures (Blancas-Mejía, Horn et al. 2015) (previously described in Figure 1B and illustrated below in Figure 2).

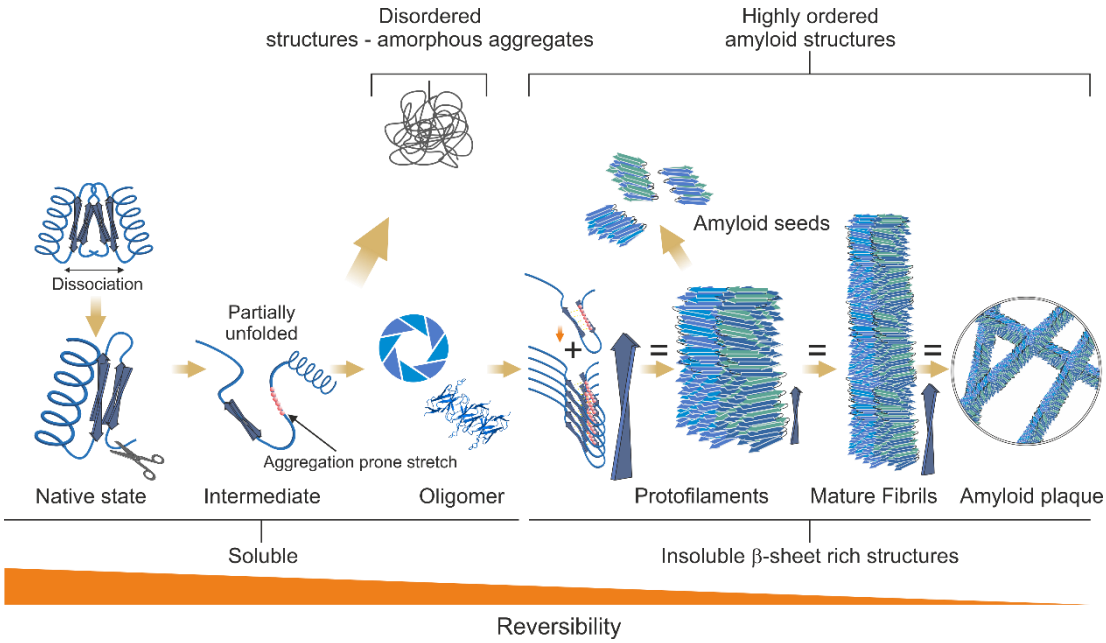


Figure 2 Schematic illustration of a simplified linear pathway of protein aggregation in amyloid formation with a focus on light chain aggregation. Native proteins undergo conformational alterations that ultimately result in mature fibrils possessing highly ordered β-sheet rich quaternary structure. Partially unfolded intermediates expose regions of the protein that are crucial for formation of the amyloid spine (highlighted red), where much of the protein is peripheral to the fibril core. These intermediates associate forming structured and unstructured oligomers of diverse morphologies that lead into the formation of β-sheet rich protofibrils. At the end of the pathway are mature fibrils. Shown are possible mechanisms that trigger the fibrillation reaction which includes dissociation of amyloid resistant quaternary structures such as Transthyretin (TTR), or cleavage of the native protein such as Aβ.

As protein folding usually occurs on the microsecond to second timescale (Chen, Lu et al. 2017), many of the transient, non-native species that populate the folding, and misfolding pathways remain uncharacterised. Many conventional spectroscopies are simply unable to directly detect transient lowly populated conformers on this short timescale. Given the implication of some of these species in misfolding diseases, there is on-going international effort by

the scientific community to understand this process and characterise the different protein states using many different approaches. David P Smith and co-workers harnessed the ability of electrospray ionisation - ion mobility mass spectrometry to first, discriminate between the many co-populated conformational states of β2-microglobulin during the early stages of its aggregation and later, to identify oligomers of different morphologies by analysis of their collision cross-section, separating species by their mass to charge ratio, and shape (Smith, Radford et al. 2010, Smith, Woods et al. 2011). There have also been a number of investigations using NMR spectroscopy, and small angle X-ray scattering (SAXS) coupled with novel software COSMiCS (Complex Objective Structural analysis of Multi-Component Systems) to characterise transient, (Herranz-Trillo, Groenning et al. 2017) intermediate species, of low population number even in structurally heterogeneous samples (Karamanos, Kalverda et al. 2015, Karamanos, Pashley et al. 2016, Herranz-Trillo, Groenning et al. 2017). Collectively these investigations, are amongst growing evidence which highlight that misfolding does not proceed along a linear pathway rather, many structurally diverse oligomers, are able to proceed through multiple pathways towards fibril formation (Eichner and Radford 2011).

1.3.2. Formation of the amyloid fibril

The kinetics of amyloid formation is heavily dependent on protein concentration, and a number of intrinsic and extrinsic factors that include the physicochemical properties of the precursor protein, pH, ionic strength, temperature, metal ion content, different surface compositions (such as the difference in phospholipid head groups) tissue composition (elastin, collagen

types) and the presence of other co-factors (Hane and Leonenko 2014, Marin-Argany, Guell-Bosch et al. 2015, Kastritis and Dimopoulos 2016, Graña-Montes, Pujols-Pujol et al. 2017). Under specific experimental conditions, protein aggregation can be induced by destabilising the native state, enabling a range of modulating factors to be examined *in vitro*. Advances in techniques such as two-colour super-resolution microscopy (Pinotsi, Buell et al. 2014), and total internal reflection fluorescence microscopy (Ban and Goto 2006) provide high resolution methods of monitoring amyloid formation and fibril growth, but the most widely accessible method of quantifying fibrillation kinetics *in vitro* is Thioflavin T (ThT) fluorescence (Biancalana and Koide 2010). ThT is a benzothiazole dye containing two coplanar aromatic rings (Figure 3), (benzothiazole and N,N-dimethylaniline) that in solution, are in free rotation around the axial bond (Taniguchi, Shimizu et al. 2016). The exact mechanisms by which ThT binds are unclear, but evidence proposes a cavity, present only within fibrillar species that possess a cross-β-sheet spine are able to quench the aromatic rotation that gives rise to an increase in fluorescence upon photo-irradiation when compared to an unbound ThT molecule (Khurana, Coleman et al. 2005, Biancalana and Koide 2010, Qin, Sun et al. 2017).

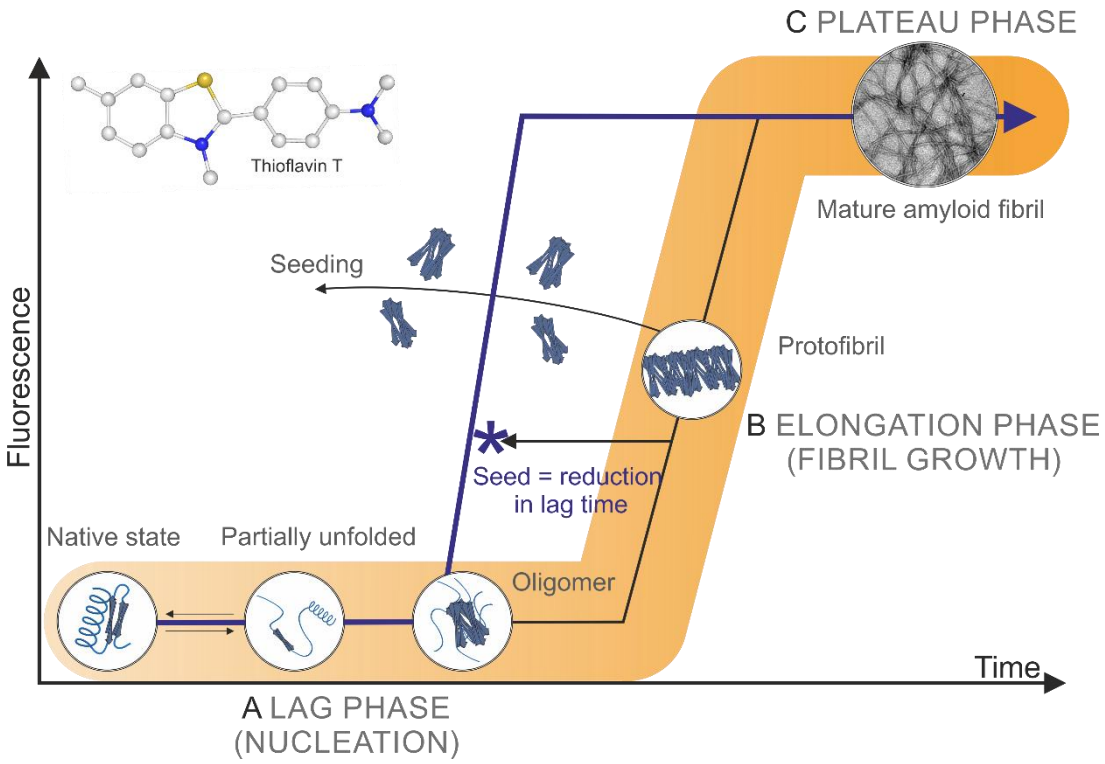


Figure 3 Three-stage sigmoidal growth curve representing a ThT fluorescence experiment. Curve is used to illustrate both ThT reactive and nonreactive species in the aggregation pathway (A) lag phase (B) elongation phase (C) plateau phase. The presence of fibrillar species that contain cross-β architecture results in an increase in ThT fluorescence (Y axis). Preformed seeds are able to reduce the lag phase (highlighted, blue). Schematic based on information from (Wilson, Yerbury et al. 2008)

The most widely accepted concept for the i*n vitro* formation of amyloid proceeds through a nucleation dependent polymerisation procedure (Lee, Nayak et al. 2007, Kumar and Walter 2011, Blancas-Mejia, Misra et al. 2017). The kinetics of this reaction, which can be observed in a typical ThT fluorescence assay fit a sigmoidal growth function (illustrated in Figure 3) and can be divided into three distinct stages (Lee, Nayak et al. 2007): the lag phase (A), the elongation phase (B) and the plateau phase (C). At the start of the fibrillation reaction, a solution that contains proteins in their monomeric state can associate to form an oligomeric nucleus that is enriched in β-sheets

(Baden, Sikkink et al. 2009, Arosio, Knowles et al. 2015, Blancas-Mejia, Misra et al. 2017). The lag phase (illustrated in Figure 3 as phase A) is used to represent the time that is taken for the gradual formation of this amyloid nucleus. This is the rate-limiting step of the reaction, and can be accelerated by the presence of pre-formed fibrillar aggregates from the same precursor protein (self-seeding) or proteins that have an entirely different amino acid composition (cross-seeding) (illustrated in Figure 3 blue line) (Merlini and Bellotti 2003, Wall, Kennel et al. 2012, Arosio, Knowles et al. 2015) (Ghosh, Vaidya et al. 2016). Once the amyloid nucleus is established, free monomers are able to bind to complementary ends of the aggregates, which elongates them (Phase B). As illustrated in Figure 3, this occurs much more rapidly than the formation of the amyloid nucleus (indicated by time x axis)(Kumar and Walter 2011). Fibrils may also break up into smaller fragments, that extend the number of complementary ends where monomers can bind (Knowles, Vendruscolo et al. 2014, Arosio, Knowles et al. 2015). In the final phase, which is referred to as the plateau phase, available monomers have been consumed during the fibrillation reaction, and so there is no further increase in ThT signal (Scheinost, Boldt et al. 2012)

1.3.3. Structure of the amyloid fibril

Amyloid fibrils are the stationary, end stage assembly component for all amyloid aggregation pathways. As a result, there is arguably a more developed understanding of their structural properties in comparison to the transient species mentioned earlier. Historically, the protein found within amyloid deposits was first thought to be amorphous, that is, lacking in any of the highly ordered structural elements that are described here (Ross and

Poirier 2005, Yoshimura, Lin et al. 2012). This view was altered in 1935 where, X-ray diffraction analysis of aggregated lysozyme revealed a cross β diffraction pattern (Astbury, Dickinson et al. 1935). Over the years, the study of fibrils generated from many amyloid proteins and synthetic fibrils have revealed that all share a similar ultrastructure despite low sequence identity (Eanes and Glenner 1968, Geddes, Parker et al. 1968, Rambaran and Serpell 2008).

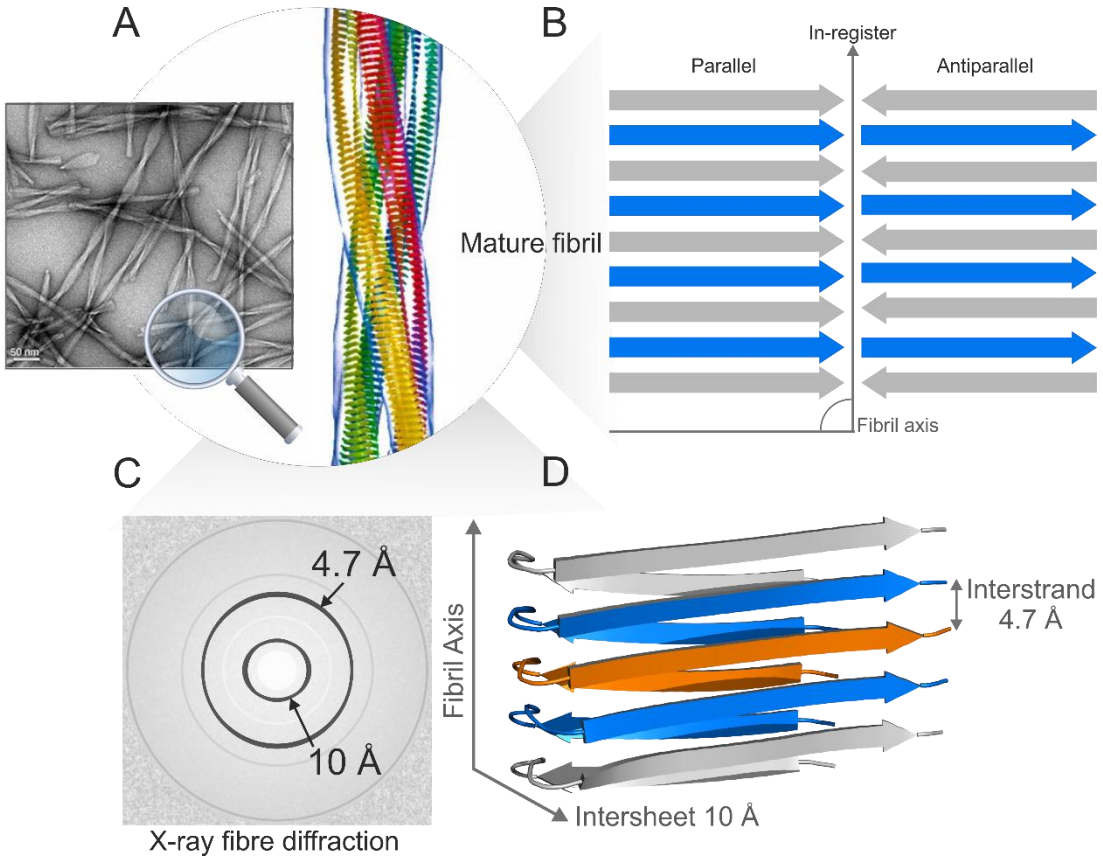


Figure 4 Illustration of fibril architecture. A) Molecular model of mature fibril adopted from (Jiménez, Nettleton et al. 2002) with negative stain electron micrograph image of Transthyretin synthetic peptide (residues 105-115 termed TTR1) inset adopted from (Gras, Waddington et al. 2011). B) Illustration of parallel or antiparallel running β-strands that stack in-register, and are orientated perpendicular to the fibril axis. D) Section of fibril taken from Aβ shows that monomers adopt a β-turn, and form individual β-strands that stack in register to form β-sheets which run parallel to the fibril axis. (C) Cross-β spacing can be measured by X-ray fibre diffraction that gives rise to distinct bands at 4.7 Å and 10 Å.

On a microscopic level, electron and atomic force microscopy have revealed that mature fibrils are often straight, elongated, unbranched rope-like structures (Figure 4A) (Serpell 2000), that are ~ 100-200 Å in height and around 1-10 microns in length (Serpell 2000, Rambaran and Serpell 2008, Xu, Paparcone et al. 2010, Fitzpatrick, Debelouchina et al. 2013). At a molecular level, amyloid fibrils are built from thousands of individual self-associated monomers, that during the course of the fibrillation reaction (previously illustrated in Figure 2) stack on top of one another to form arrays of elongated β-sheets (Figure 4B) (Riek and Eisenberg 2016). These sheets are positioned parallel or anti-parallel (less common) to each other, (Rambaran and Serpell 2008, Toyama and Weissman 2011, Arosio, Knowles et al. 2015) but are orientated perpendicular against the main axis of the fibre (Figure 4 B). The characteristic feature of the amyloid fibril is a cross β diffraction that has meridional and equatorial reflections of 4.7 Å and 10 Å respectively (Figure 4C) (Rambaran and Serpell 2008, Fandrich, Schmidt et al. 2011). The origin of these reflections is shown in Figure 4D, where a segment of the Aβ fibre is illustrated. Here, individual monomers have adopted a β-turn-β motif with two opposite running β- sheets (Toyama and Weissman 2011) The inter- β strand and intra- β sheet spacings (highlighted in Å) which arise due to hydrogen bond and side chain arrangements (Toyama and Weissman 2011) correspond to the aforedescribed reflections acquired in a typical X-ray diffraction experiment (Del Pozo-Yauner, Becerril et al. 2015). In the core of the fibrils, side chains of neighbouring monomers point inwards, and interdigitate to create an anhydrous zipper-like arrangement (Sawaya, Sambashivan et al. 2007, Toyama and Weissman 2011, Graña-Montes, Pujols-Pujol et al. 2017)

which, along with the extensive hydrogen bonding network of the cross-β conformation contributes to the incredible rigidity of the fibril (Sawaya, Sambashivan et al. 2007, Toyama and Weissman 2011). Analysis of microcrystals using small synthetic peptides that possess this structure has revealed tensile strengths in the Gigapascal (GPa) range (Smith, Knowles et al. 2006, Knowles, Fitzpatrick et al. 2007, Fitzpatrick, Park et al. 2013, Knowles, Vendruscolo et al. 2014), which are comparable to that of steel and spiders silk (Fitzpatrick, Debelouchina et al. 2013).

Sets of β-sheets assemble to form protofilaments, which can be described as “sub fibrils” that wind around one another in a helical-like arrangement. This often leaves the mature fibril with a hollow tube-like core (Serpell 2000, Toyama and Weissman 2011, Fitzpatrick, Debelouchina et al. 2013). The number of associated protofilaments (documented to be 2-4-6 (Jimenez, Nettleton et al. 2002, Jiménez, Nettleton et al. 2002, Fandrich, Schmidt et al. 2011) ultimately leads to different levels of periodicity (the twist of the full amyloid fibril) (Adamcik, Jung et al. 2010) which is observed when using electron or atomic force microscopy (Figure 4A).

1.4. Immunoglobulin light-chain amyloidosis

1.4.1. Overview and clinical presentation

AL amyloidosis (previously termed primary amyloidosis) is the most frequent systemic conformational disorder, with an incidence of 1:100,000 (Dispenzieri, Buadi et al. 2015). It is currently incurable, and requires a rapid diagnosis followed by appropriate treatment for any increase in life expectancy. The median survival for a patient with advanced-stage cardiac damage (which is

common clinical manifestation in this disease) is just 6 months (Grogan, Dispenzieri *et al.* 2017). AL amyloidosis has some similarities to multiple myeloma (MM). Both disorders are associated with an underlying plasma cell dyscrasia (Martin, Williams et al. 2017) where plasma B cells proliferate abnormally, and synthesise enormously high levels of light chain (Figure 5A) (Geddes, Parker et al. 1968, Wei and Juneja 2003). In MM, the light chain is not amyloidogenic, instead forming amorphous aggregates which commonly leads to cast nephropathy (formation of urinary casts, within the renal tubules) (Korbet and Schwartz 2006). This disorder is mainly characterised by the aggressive plasma cell clone that infiltrates the bone marrow, forming lytic bone lesions that lead to the osteoporosis and fracturing (Kristinsson, Minter et al. 2011, Tosi 2013). By comparison, the plasma cell clone in AL is less aggressive (lower tumour burden), instead, light chains can be amyloidogenic (Sanchorawala 2006). Amyloidogenic LCs that evade the tightly regulated PQC mechanisms of the plasma B cell (Cooley, Ryno et al. 2014) enter the bloodstream and can form amyloid fibrils (Rajkumar, Dispenzieri et al. 2006, Martin and Ramirez-Alvarado 2010).

Deposition of the amyloidogenic LC can occur in virtually every organ, including circumventricular organs (tissue distribution illustrated in Figure 5) (Schroder and Linke 1999), the light chain is however unable to traverse the blood brain barrier (Martin and Ramirez-Alvarado 2010). Sites of deposition vary between patients, but is most commonly found in the heart (75%), liver (75%) and kidney (65%) (Baden, Sikkink et al. 2009, Melmed 2009, Martin and Ramirez-Alvarado 2010, Basnayake, Stringer et al. 2011, Palladini and Merlini 2016). The exact reasons to why there are differences in organ involvement

between patients is not entirely clear, but several lines of evidence point towards the amino acid composition of each light chain as the crucial determinant to organ tropism (Comenzo, Zhang et al. 2001). For example, light chains of the IGLV6S1/Vλ6a group has been previously associated with renal involvement (Ramirez-Alvarado, De Stigter et al. 2007). Light chains of the Vλ and λII 2b2 have shown to be predominantly cardiac involvement, and λIII 3r germ-line gene has been correlated with soft tissue involvement (Comenzo, Wally et al. 1999, Abraham, Geyer et al. 2003).

AL is largely non-symptomatic until the latter stages of disease progression. There are recent indications to suggest that the underlying plasma B cell clone can be active, synthesising large quantities of light chain for years without detection (Chaulagain & Comenzo 2015). As a result, many patients display severe organ dysfunction when AL is diagnosed, and unfortunately, 30% of patients die within just 12 months following this diagnosis (Dispenzieri and Merlini 2016, Grogan, Dispenzieri et al. 2017). The presenting features of AL are mostly dependent on the levels of amyloid deposition in the patient, and the organ involved (Melmed 2009). Some of the common symptoms of AL (seen in 90% of patients) include weight loss, stiff macroglossia (tongue), edema (swelling of limbs), fatigue and light-headedness. These are however, commonly mistaken for more benign conditions and are not a clear indication of AL amyloidosis (Baker and Rice 2012).

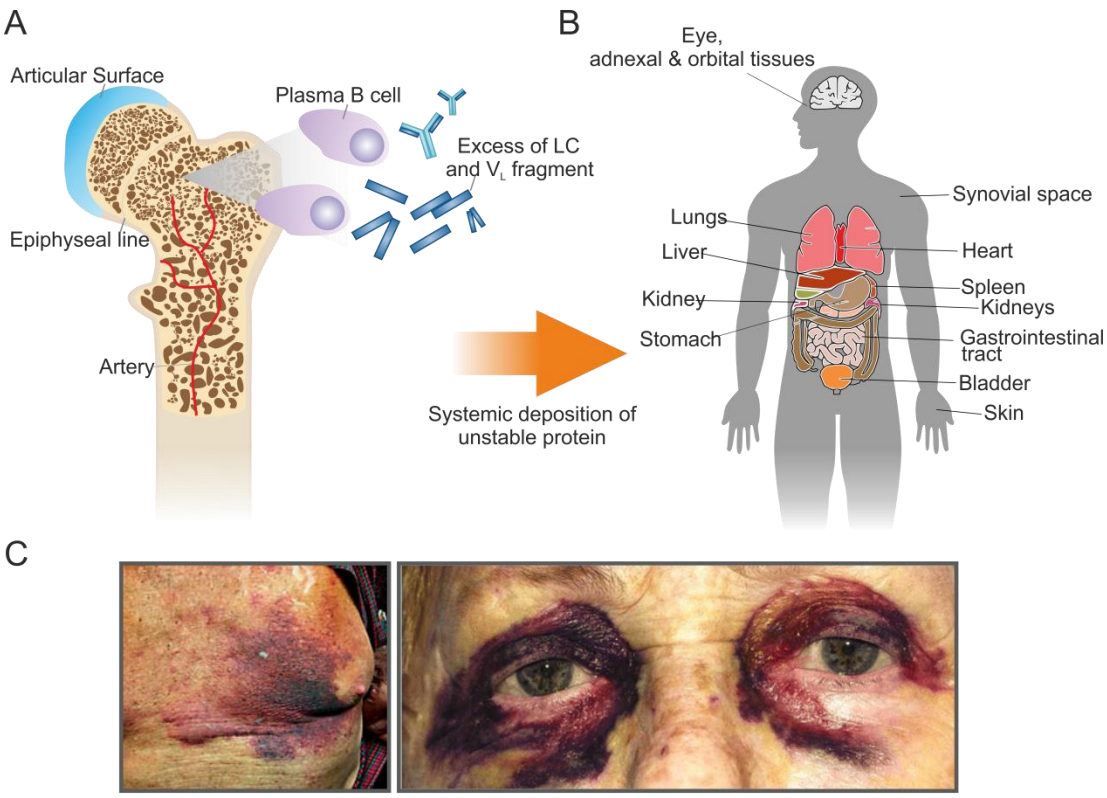


Figure 5 Origin and distribution of Immunoglobulin free light chain. (A) The small, indolent plasma B cell clone responsible for secretion of excess of full LC and truncated fragments is located within the bone marrow (B). The self-assembly and deposition of unstable LC is systemic, occurring in virtually all organs and tissues outside of the central nervous system and the blood brain barrier. Frequent sites of deposition include heart, kidneys, liver, gastrointestinal tract, and the peripheral nervous system. (C) Shows hematoma on chest well in patient diagnosed with AL (left panel) adopted from (Muchtar, Buadi et al. 2016) where right panel shows example of 73 year old patient diagnosed with AL with periorbital purpura (adopted from (Eder and Bitterman 2007).

Other, organ specific symptoms include distal sensory neuropathy when there is nervous system involvement, (Sanchorawala 2006) proteinuria and reduced glomerular filtration rates, with renal involvement (Dember 2006, Basnayake, Stringer et al. 2011) and fatigue, dyspnea (breathing difficulties) and orthostaic hypertension (increase in blood pressure upon standing) (Mahmood, Palladini et al. 2014, Lin, Seldin et al. 2015) for patients with cardiac involvement.

Symptoms may also be more visual, such as the skin hematoma (Muchtar, Buadi et al. 2016), or periorbital purpura illustrated in Figure 5. This occurs in approximately 1/3rd of all patients (Mahmood, Palladini et al. 2014).

1.4.2. Diagnosis of AL amyloidosis

Several techniques are used to confirm diagnosis. Usually, this begins with the demonstration of monoclonal immunoglobulin free light chain (FLCs) in the blood or urine using protein and immunofixation electrophoresis (Dember 2006, Kastritis and Dimopoulos 2016). An example of a serum protein electrophoresis experiment is shown in Figure 6A, where elevated FLC levels are seen as a spike in the gamma globulin (also termed immune globulin) region of the densitometric trace. Following the identification of a spike in this region, which is commonly termed a paraprotein or monoclonal M spike, quantification of total FLC levels are made using patient serum (Jenner 2014). In healthy individuals, normal concentrations of free light chain are ~3.3 mg/L – 19.4 mg/L for kappa and ~5.71–26.3 mg/L for lambda isotypes (Sanchorawala 2006).

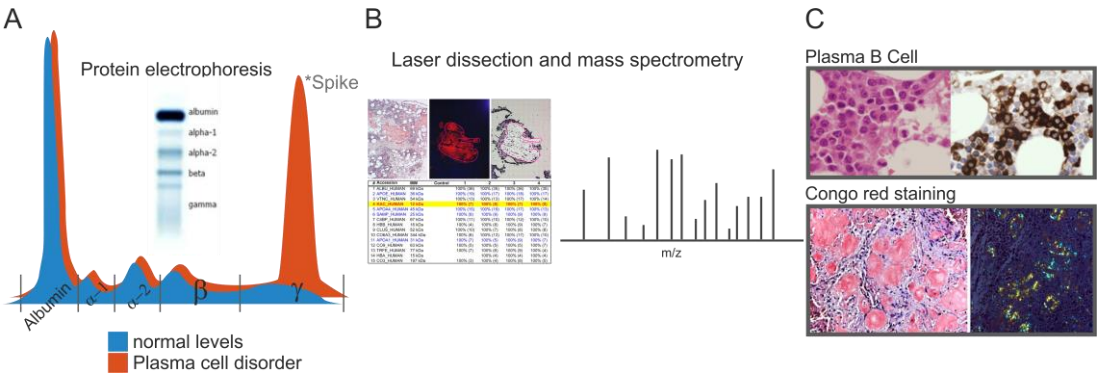


Figure 6 Simplified schematic of steps taken towards diagnosis of AL. Suspicion of AL is followed by routine electrophoresis and immunofixation of urine and serum samples. (A) Example of serum protein electrophoresis and corresponding gel shows separation of major serum proteins albumin, α 1, α 2 β and γ globulin. Normal levels are shown in blue, and patients with an underlying MGUS have a spike at gamma globulin region. (B) Accurate subtyping of amyloid is made by laser dissection of amyloid deposit. This procedure is designed to elucidate the precursor protein. Image adopted from (Vrana, Gamez et al. 2009).(C) Assessment of Plasma B cell burden by immunohistochemical stain (CD138) (top panel) is followed by confirmation of amyloid presence (bottom panel). Congo red staining shows cardiac tissue, under fluorescent and polarising light, displaying characteristic apple green birefringence (image adopted from (Mohty, Damy et al. 2013).

In patients with AL amyloidosis, immunoglobulin synthesis is dramatically increased and so, free light chain levels can be astonishingly high, documented to be in hundreds of milligrams (Melmed 2009, Kumar, Dispenzieri et al. 2010) grams (Imamura, Ogata et al. 2006, Martin and Ramirez-Alvarado 2010, Hajra and Bandyopadhyay 2016) and kilograms per litre (Knowles, Vendruscolo et al. 2014). Monitoring the ratio of free light chain is also important in recognising the disease, where, the normal κ/λ ratio of 2:1 observed in healthy individuals is altered to a λ/κ ratio of 3:1 (Gertz, Lacy et al. 2002) (Sanchorawala 2006, Baden, Owen et al. 2008, Jenner 2014, Le Bras, Molinier-Frenkel et al. 2017). Elevated FLC levels, abnormal κ/λ ratios, or the detection of monoclonal light chain by electrophoretic, or immunofixation

methods signifies any number of haematological disorders, that include the benign monoclonal gammopathy of undetermined significance (MGUS), to the more severe plasma cell dyscrasias that include multiple myeloma, Waldenström's macroglobulinemia , smouldering multiple myeloma, non-Hodgkin's lymphoma, and light chain amyloidosis (Boccadoro and Pileri 1995, Jenner 2014, Singh 2017).

To distinguish AL from other gammopathies, both immunohistochemical classification (using kappa and lambda antibodies) and laser microdissection mass spectrometry (MS) is performed on tissue taken from the symptomatic organ, salivary gland, fat aspirates, or other subcutaneous tissue (Figure 6B) (Vrana, Gamez et al. 2009, Mohty, Damy et al. 2013, Kastritis and Dimopoulos 2016, Muchtar, Buadi et al. 2016).

Another important step to ensure the correct diagnosis of AL amyloidosis is assessing the plasma cell burden. In AL amyloidosis, there is typically a much lower number of malignant plasma cells within the bone marrow than that of multiple myeloma and Waldenström's macroglobulinemia (Wei and Juneja 2003, Rajkumar, Dispenzieri et al. 2006, Sanchorawala 2006, Sanchorawala, Blanchard et al. 2006, Dinner, Witteles et al. 2013). Detection is performed by fluorescence in situ hybridisation (FISH) or CD138 (plasma cell marker) immunohistochemistry on bone marrow sections, that are usually obtained from percutaneous trephine biopsy (Figure 6C) (O'Connell, Pinkus et al. 2004, Selkoe 2013, Falk 2014). Finally, confirmation of AL also requires the presence of amyloid to be made histologically. Here, tissue sections are evaluated for Congo red positivity, where in the presence of amyloid, tissue will exhibit characteristic apple-green

birefringence when visualised under cross polarised light (Figure 6C) (Kastritis and Dimopoulos 2016).

1.4.3. Cardiovascular manifestations in AL amyloidosis

As preserving cardiac function is key to patient survival, assessing the involvement of this organ has now become a critical part in establishing diagnosis and methods for assessment became standardised at the 10th Annual Meeting of the International Society of Amyloidosis (ISA) (Gertz, Comenzo et al. 2005, Cohen and Comenzo 2010, Mohty, Damy et al. 2013). In addition to some of the clinical presentations of cardiac involvement described earlier (dyspnea and orthostaic hypertension) (Cohen and Comenzo 2010) detection of elevated presymptomatic cardiac biomarkers such as BNP (brain natriuretic peptide), N-terminal portion of its pro-hormone (NT-proBNP), troponins I or T (Palladini, Campana et al. 2003; Merlini, Seldin et al. 2011; Panagopoulou, Deftereos et al. 2013) as well as performing a number of electrophysiological, morphological and hemodynamic assessments have become part of the diagnostic procedure (Shin, Ward et al. 2012)

Typical examples of cardiac involvement include a low voltage electrocardiogram which is seen in ~45-70% of patients (Falk 2005, Chaulagain and Comenzo 2015). Cardiac magnetic resonance (CMR) and echocardiography (ECHO) are used to detect structural abnormalities and functional impartments (Cohen and Comenzo 2010, Fikrle, Paleček et al. 2013, Muchtar, Buadi et al. 2016). Typical structural abnormalities detected by ECHO imaging are exemplified in Figure 7 and include myocardium deposition of amyloid which results in a speckled myocardium, myocardial wall and

septum thickening, and ventricular hypertrophy (increased ventricular wall thickness) (Mohty, Damy et al. 2013).

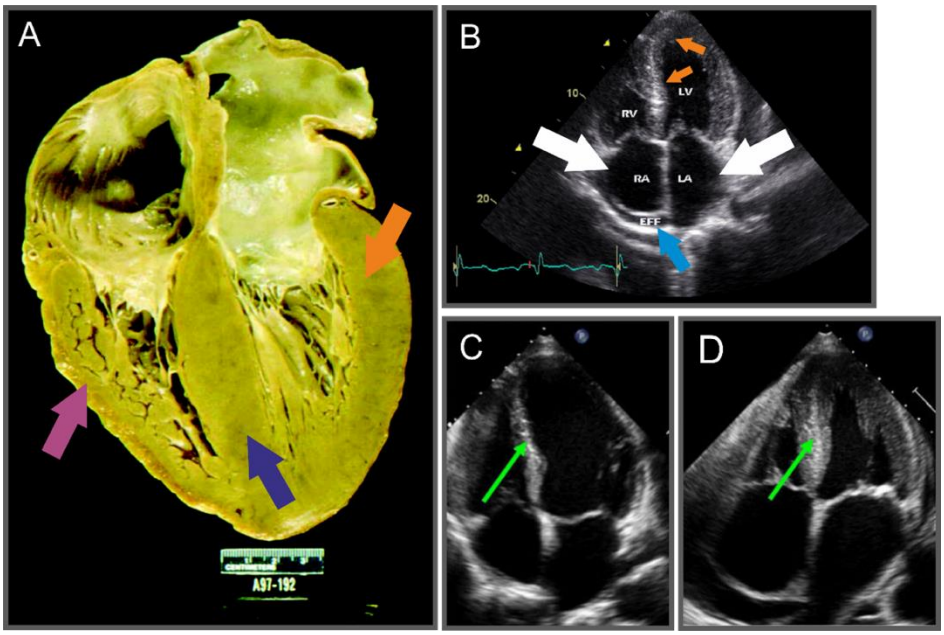


Figure 7 Example of cardiac involvement in AL amyloidosis. (A) shows cross section of heart isolated from patient diagnosed with AL. Arrows indicate enlarged right vetricular wall, with nodules, significantly enlarged ventricular septum (dark blue) and enlarged ventricular wall (orange). Image adapted from (Falk 2005). (B) Echocardiogram shows enlargement of right and left atrium (white arrow) and presence of a pericardial effusion (fluid collection around heart, light blue arrow). The speckled appearance of the septum shown in the ECG image (orange arrow) is indicative of amyloid infiltration. Image adopted from (Falk and Dubrey 2010). Echocardiogram (C) show septum thickness in healthy patient (green arrow) in comparison to patient diagnosed with AL (D green arrow) that shows enlarged atrial septum. Image adopted from (Care 2017).

1.4.4. Amyloid pathophysiology and cardiac involvement in AL

Sudden death, arrhythmias, syncope, and diastolic dysfunction are common consequences of myocardial amyloid infiltration (Brenner, Jain et al. 2004, Hassan, Al-Sergani et al. 2005, Mohty, Damy et al. 2013). Despite the urgency in preserving cardiac function to enhance patient survival, the precise mechanism in which toxic variable domains lead to fatal cardiomyopathy is uncertain. In truth, the mechanisms that underlie amyloid toxicity for all conformational disorders are still unclear.

Historically, the mature amyloid fibril was once thought to represent the cytotoxic species after numerous demonstrations of fibrillar Aβ to impair many critical cellular processes (Lorenzo and Yankner 1996, Verma, Vats et al. 2015). Early notions suggested that amyloid deposits could form a barrier around cells, and physically impair many fundamental cellular processes (Blancas-Mejía and Ramirez-Alvarado 2013). Other works have shown fibrils to disrupt the cell membrane (Ow and Dunstan 2014). For AL amyloidosis, the physical presence of amyloid fibrils is still held responsible for many of the cardiac manifestations described above. The displacement of parenchymal (functional) tissue, such as the contractile elements and muscle of the heart with rigid amyloid fibrils (Hassan, Al-Sergani et al. 2005, Grogan, Dispenzieri et al. 2017) has been previously demonstrated to restrict ventricular filling (Pepys 2006). However, differences in cardiac outcomes have been noted between patients despite very similar levels of amyloid deposition (Hassan, Al-Sergani et al. 2005, Fikrle, Paleček et al. 2013, Mishra, Guan et al. 2013). Furthermore, elevated levels of NT-proBNP (the biomarker used to indicate cardiac involvement) has also been detected in patients despite the absence

of amyloid fibril deposits (Grogan, Dispenzieri et al. 2017). A number of studies using animal models (Zebrafish, mouse and *Caenorhabditis elegans (C.elegans)* have also detected functional impairments that reflect the manifestations found in human amyloid cardiomyopathy, without any evidence of fibrillar light chain deposition (negative for Congo red staining and electron microscopy) (Shin, Ward et al. 2012, Mishra, Guan et al. 2013, Diomede, Rognoni et al. 2014, Diomede, Rognoni et al. 2014, Guan, Mishra et al. 2014). The ability of soluble light chains to internalise into primary heart cardiac fibroblasts (hCFs), cardiomyocytes and renal cells and alter cellular metabolism and cellular ultrastructure has also been documented (and is an aspect that is further explored within this thesis) (Brenner, Jain et al. 2004, Baden, Sikkink et al. 2009, Shi, Guan et al. 2010, Sikkink and Ramirez-Alvarado 2010, Lavatelli, Imperlini et al. 2015, Marin-Argany, Lin et al. 2016). Collectively, these findings suggest that in addition to insoluble mature amyloid fibrils, soluble oligomers play a significant role in disease pathology (Knowles, Vendruscolo et al. 2014). In truth, this is reflected across all the amyloidoses, however, the exact mechanisms whereby cytotoxicity is accomplished for oligomers (also referred to as prefibrillar aggregates) is still a matter of debate.

There are documented examples of annular and wreath shaped oligomers permeabilising the plasma membrane, and creating pores that disrupt cellular ion homoeostasis which leads to apoptosis (Knowles, Vendruscolo et al. 2014, Ow and Dunstan 2014). The ability to adopt a pore-like conformation is noted in a number of protein precursors including κ1 light chains (Monis, Schultz et al. 2006), alpha synuclein and Aβ suggesting there may be a common molecular mechanism of proteotoxicity (Di Scala, Yahi et al.

2016). In AL, soluble oligomers are believed to cause diastolic dysfunction by inducing oxidative stress through activation of non-canonical p38 mitogen-activated protein kinase (MAPK) pathway which also leads to cellular apoptosis (Brenner, Jain et al. 2004) (Brenner, Jain et al. 2004, Shi, Guan et al. 2010, McWilliams-Koeppen, Foster et al. 2015, Kastritis and Dimopoulos 2016, Palladini and Merlini 2016). Quite interestingly, activation of this pathway promotes NT-proBNP synthesis; the presymptomatic cardiac biomarker mentioned earlier.

1.4.5. Treatment strategies for AL amyloidosis

Light chain amyloidosis is currently incurable. To date, there is no European Medicines Agency (EMA), or Drug Administration (FDA) approved treatment that addresses AL specifically (Bayliss, McCausland et al. 2017). Treatment strategies are constantly changing based on the results of clinical trials, but a widely adopted strategy used in the treatment of multiple myeloma (Reece, Sanchorawala et al. 2009) involves the administration of high dose cyclic melphalan (an alkylating agent) in combination with autologous blood stem cell transplantation and chemotherapy (Sanchorawala 2006, Roy 2012, Mahmood, Palladini et al. 2014, Grogan, Dispenzieri et al. 2017). This is an anti-plasma cell therapy that is designed to eradicate the underlying plasma B cell dyscrasia. Suppression of the plasma cell clone has a knock-on effect, leading to a reduction in light chain synthesis, and a lower quantity of light chain in circulation. From clinical evaluation studies, patient survival under this treatment has increased dramatically, from a median of 6 months without treatment, to 5 years (Dispenzieri, Seenithamby et al. 2013, Muchtar, Buadi et al. 2016). In one study, the 40% of patients who demonstrated a complete

haematological regression following treatment (an assessment criterion that refers to no detectable plasma B cells in the bone marrow, no detectable levels of free light chain in serum and urine, and normal free light chain levels (Muchtar, Buadi et al. 2016)) a survival rate of 13.2 years was documented (Cibeira, Sanchorawala et al. 2011, Dispenzieri, Seenithamby et al. 2013, Mahmood, Palladini et al. 2014, Chaulagain and Comenzo 2015, Muchtar, Buadi et al. 2016). There are however, several downsides to this therapy. Mainly, the treatment is highly toxic. Early studies had a treatment related mortality of 15-40% (Sanchorawala 2006) and as a result, had risk adopted doses of melphalan (doses altered dependent on patient age and severity of renal and cardiac involvement) to try and limit the toxic effects (Cibeira, Sanchorawala et al. 2011, Chaulagain and Comenzo 2015). The eligibility for this treatment is also very low, accessible to only 20-25% of patients that have preserved organ function (Poshusta, Katoh et al. 2013). Patients that do not satisfy the eligibility criteria for ASCT (cardiac involvement in advanced stages) are simply too weak to survive the treatment (Fikrle, Paleček et al. 2013, Grogan, Dispenzieri et al. 2017) and instead, combination therapies that include high-dose melphalan, dexamethasone (Mdex), cyclophosphamide and bortezomib (proteasome inhibitor) are a commonly used alternatives (Wechalekar, Lachmann et al. 2008, Simpson, Herold et al. 2009, Cohen and Comenzo 2010, Mikhael, Schuster et al. 2012, Mahmood, Palladini et al. 2014, Palladini, Milani et al. 2014, Huang, Wang et al. 2016, Palladini and Merlini 2016, Le Bras, Molinier-Frenkel et al. 2017). These methods do not however, address the resident amyloid. Emerging therapies that aim to remove the

existing amyloid deposits or prevent aggregation are described below in Table 2.

Table 2 Emerging strategies for the treatment of AL amyloidosis. The name, type, clinical stage and mechanism of action for each therapeutic is shown below. Information for each therapy was adopted from the following sources; NEOD001 (Gertz, Landau et al. 2016), (2R)−1-[6-[(2R)−2-carboxypyrrolidin-1-yl]−6-oxohexanoyl] pyrrolidine-2-corboxylic acid (CPHPC) (Richards, Cookson et al. 2015), Epigallocatechin-3-gallate (ECGC) (Stefani and Rigacci 2013, Pelaez-Aguilar, Rivillas-Acevedo et al. 2015, Merlini 2016), Doxycycline (Grogan, Dispenzieri et al. 2017).

Name	Stage	Mechanism	Type
NEOD001	Phase I/II	Promote amyloid clearance - resorption by macrophage	Monoclonal antibody
Anti-sap and CPHPC	Phase I	Remove co-localised SAP and promote amyloid clearance	Combined small molecule and monoclonal antibody
EGCG	Phase II	Inhibit fibrillogenesis Direct protein into off-pathway non-toxic species	Small molecule
Doxycycline	Phase II	Inhibit fibrillogenesis	Small molecule

1.4.6. Immunoglobulin structure

Structurally, the complete Immunoglobulin (Ig) molecule is a 150 kDa heterotetramer comprising of four subunits: two identical light chains (LC) of ~25 kDa in complex with two identical ~50 kDa heavy chains (HC) (Carter 2006) that are linked by an intermolecular disulphide bond which gives rise to the familiar “Y-shape” morphology depicted in Figure 8. Under normal conditions, the light and heavy chains are synthesised at near equivalent ratios within the plasma B cell, where the complete Immunoglobulin molecule is

secreted into circulation (Randles, Thompson et al. 2009). In addition to the expansion of the plasma B cell population, control of chain synthesis is lost (Brumshtein, Esswein et al. 2014). This leads to the disproportionate synthesis of unbound LC, which is subsequently secreted into the circulation. The heavy chain component is retained by the endoplasmic reticulum and degraded (Cooley, Ryno et al. 2014).

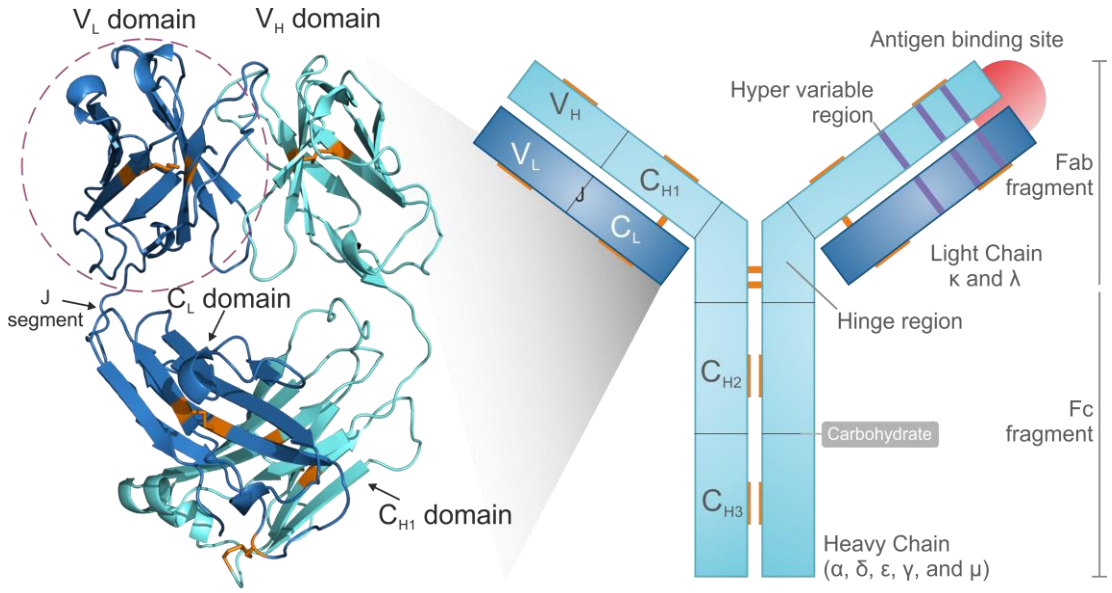


Figure 8 Diagram of single Immunoglobulin IgG molecule. The Ig molecule is a ~150 kDa heterotetrameric complex composed of four main components; two identical heavy chains (HC) covalently linked to two identical light chains (LC). The HC contains a single variable domain (V_H) and three constant domains termed C_H1 C_H2, and C_H3. The expanded region (Left) shows the V_L (circled) and C_L (blue) bound to the V_H and C_H1 domain (cyan). Five heavy chain isotypes (α, δ, ε, γ, and μ) denote which five classes the Ig will belong to following synthesis; IgG, IgA, IgM, IgD and IgE respectively (Schroeder and Cavacini 2010). Disulfide bridges are indicated in orange. Regions that comprise the fragment antigen-binding (Fab) and fragment crystallisable (Fc) are highlighted. The antigen binding site (red) and carbohydrate binding site are also highlighted. Each domain contains the highly conserved Ig fold, a homologous β-barrel (Ramirez-Alvarado 2011).

The light chain component spans 214-220 residues and consists of an N-terminal variable domain (V_L) (residues 1-108), a junction (J) and a constant domain (C_L) spanning residues 109-214 (often displayed as V_L-J-C_L) which pair non-covalently to the heavy (V_H and C_H1) (Figure 8 expanded, left) (Baden, Owen et al. 2008). This LC can be of lambda (λ) or kappa (κ) family, and is formed from a total of 33 λ and 40 κ available germlines genes (Ramirez-Alvarado 2012). A collection of studies have revealed that in AL amyloidosis, there is an overrepresentation of germline genes: VκI O18/O8, λ 1, Vλ II 2a2, Vλ III 3r, and λ VI 6a, where λ VI 6a and VκI account for the majority of all monoclonal light chains found in patients with AL amyloidosis studied (Comenzo, Zhang et al. 2001, Abraham, Geyer et al. 2003, Prokaeva, Spencer et al. 2007, Poshusta, Sikkink et al. 2009). Astonishingly a possible 3000 different light chain sequences (Ramirez-Alvarado 2012) can be generated, which is further diversified by somatic hypermutations, a process that occurs in the maturation stage of the antibody to confer high antigen specificity (Ara Celi DiCostanzo 2011, Kastritis and Dimopoulos 2016). This means that every LC isolated from a patient has a unique amino acid sequence, and is a factor that several complicates the study of AL amyloidosis (Ara Celi DiCostanzo 2011).

The full-length LC (consisting of components V_L-J-C_L) has been found in a number of patients diagnosed with AL, yet, mass spectrometric analysis of fat aspirates, has revealed that the V_L (circled in figure 8) domain has been the main fibrillar component for the majority (85%) of patients (Glenner, Cuatrecasas et al. 1969, Olsen, Sletten et al. 1998). As a result of these

findings, the majority of all studies to date focus on the light chain variable domain of the full-length LC (Randles, Thompson et al. 2009, Ara Celi DiCostanzo 2011, Blancas-Mejía, Horn et al. 2015).

1.4.6. Structure of the light chain variable domain

Structurally, the V_L is a Greek key β-sandwich that comprises nine tightly packed β-strands (A, B, C, C', C'', D, E, F, G) in a four and five arrangement (Figure 9) (Simpson, Herold et al. 2009, Hernández-Santoyo, del Pozo Yauner et al. 2010). These strands are split into two anti-parallel β-sheets, and are stabilised be an inter-chain disulphide bond which is formed between β strands B and F (Randles, Thompson et al. 2009, Ramirez-Alvarado 2012). Strand A corresponds to the N-terminus and strand G corresponds to the C-terminus (illustrated in Figure 9). Strands C, C', F and G usually form interactions with the variable domain of the heavy chain which together comprise the antigen binding site of the full immunoglobulin molecule (Figure 8) (Ramirez-Alvarado, De Stigter et al. 2007).

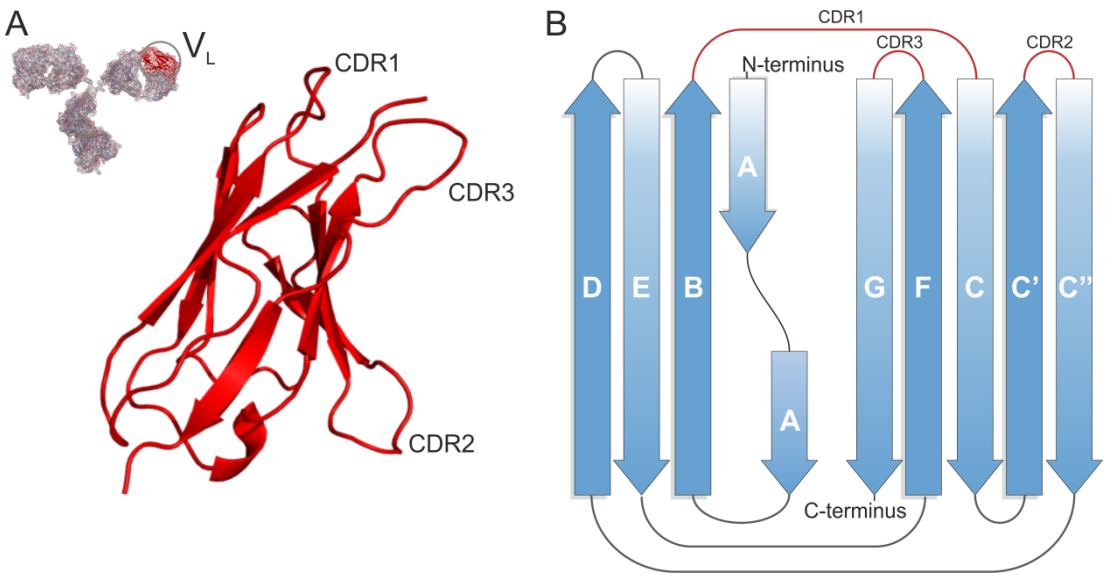


Figure 9 Structure and schematic of variable domain. Immunoglobulin molecule and expanded monomeric V_L is shown in ribbon format (A) with the complementarity determining regions highlighted. Schematic of β-strand arrangement (B) shows 4-5 anti-parallel sheet arrangement. Arrows indicate direction of chain in each strand.

The V_L also harbours four highly conserved framework regions (FR) (β-sheet structure shown in figure 9) that act as scaffolds for three unstructured hypervariable loop regions (Koenig, Lee et al. 2017) referred to as the complementarity-determining regions (CDR). The CDR regions are located between strands C'-C'' (CDR2 residues 23-34) strands B-C (CDR1 residues 50-56) and strands F-G (CDR3 residues 89-95) (Blancas-Mejia, Tischer et al. 2014) (highlighted Figure 9) (Ara Celi DiCostanzo 2011). The CDR regions are responsible for antigen binding (Randles, Thompson et al. 2009), and so, accumulate antigen-driven somatic mutations much more than the FR regions which are more highly conserved (Baden, Sikkink et al. 2009).

1.5. Variable domains pair to form homodimers

In solution V_L coexist as monomers and as stable V_L-V_L homodimers, termed Bence Jones proteins (Arosio, Owczarz et al. 2012), (Baden, Owen et al. 2008). Many of the homodimers that have been isolated from patients diagnosed with either AL, or multiple myeloma have been crystallised in a structural arrangement similar to the light and heavy variable domains (V_L and V_H) of the antigen-binding fragment (Figure 8 and Figure 10A) (Padlan 1994, Stevens and Schiffer 1995, Peterson, Baden et al. 2010). In this arrangement, both V_L monomers are orientated in the same direction. This is termed the canonical dimer arrangement and is illustrated in Figure 10A (Brumshtein, Esswein et al. 2014).

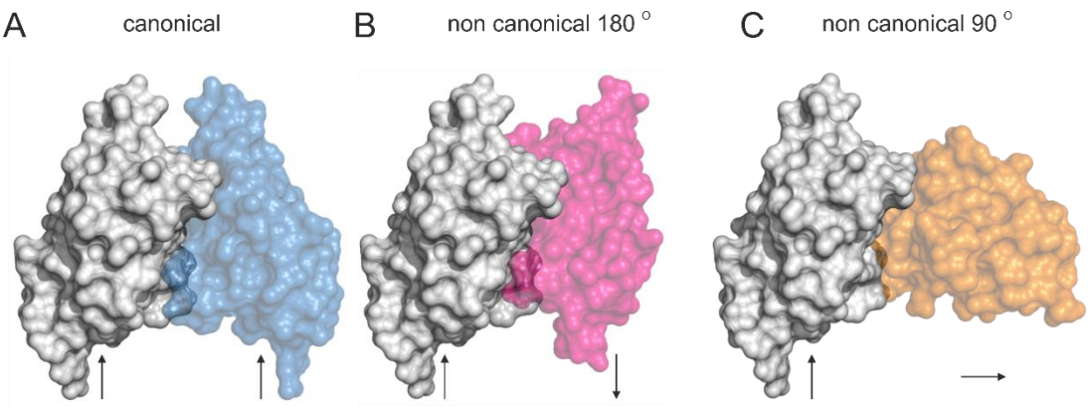


Figure 10 Alternative quaternary structure arrangements of V_L homodimer. Surface representation of (A) LEN (PDB ID: 1LVE), (B) AL-T05 (PDB ID: 5T93) and (C) AL-09 (PDB ID: 2Q1E) show canonical and non-canonical dimer interfaces with 180 ^{o} and 90 ^{o} orientation respectively. Arrows indicate orientation of the two monomers which can also be referred to as parallel, antiparallel and perpendicular, arrangements respectively.

Several lines of evidence have shown that the introduction of somatic mutations to the dimer interface can destabilise the canonical dimer, inducing its dissociation (Baden, Owen et al. 2008, Randles, Thompson et al. 2009, DiCostanzo, Thompson et al. 2012, Brumshtein, Esswein et al. 2014) or, they can cause the dimer to adopt a number of different quaternary structural arrangements, that are less stable than the canonical counterpart (Brumshtein, Esswein et al. 2014). These dimers are termed non-canonical. It is also interesting to note that these alterations are achieved without any large rearrangements to the tertiary structure of the V_L.

Alterations to the quaternary state can be quite minor, seen in the 1.9 Å structure of the amyloidogenic λ6a V_Ls Wil (PDB ID: 2CD0), which adopts a dimer that has just an 11 ^{o} rotation in comparison to the non-amyloidogenic homologus V_L Jto (PDB ID: 1CDO). Alternatively, these changes can be much more pronounced, as depicted in Figure 10B and C, where the individual subunits relative to one another are twisted into anti-parallel and perpendicular orientations. Figure 10B shows the arrangement observed within the crystal structure of the amyloidogenic V_L AL-T05 (PDB ID: 5T93) that belongs to the Vλ1 1b (IGLV 1-51) family. An identical arrangement is also seen in the amyloidogenic V_L domain REC (PDB ID: 1EK3) that belongs to the κIV family. Here, each monomer is twisted 180 ^{o} in comparison to the canonical dimer (Figure 10A). The last arrangement (Figure 10C) is seen in the amyloidogenic V_L domain AL-09, where the individual monomers are twisted 90 ^{o} relative to one another (Peterson, Baden et al. 2010). Through site-directed mutagenesis based methods, such dramatic dimer arrangements have been shown to arise from just a single point mutation. In AL-09, Luis M Blancas-Mejia *et al* revealed

that the single mutant H87Y located within the dimer interface was able to dictate the orientation of this dimer (Peterson, Baden *et al.* 2010). For the non-amyloidogenic kIV light chain LEN, single mutants K30T, Q38E, Q89A and Q89L have also been able to cause flipped dimer arrangements (Pokkuluri, Huang et al. 1998, Pokkuluri, Cai et al. 2000). New evidence provided by solution state NMR suggests that these dimer arrangements may not be static, and instead interconvert between the different conformations described above (Figure 10). As this has only been demonstrated for the V_L κI O18/O8 that harbours a single mutation (Y87H) it is not clear if this is a common feature of all V_L domains, or if this is just an isolated case (Peterson, Baden et al. 2010)

1.6. Mechanisms of light chain aggregation

Due to the different quaternary states of the variable domain that are present in solution (monomer, canonical dimer and non-canonical dimer, Figure 10), there has been some uncertainty to which species is responsible for triggering the aggregation cascade in AL amyloidosis. Recently, Boris Brumstein and colleagues have demonstrated that the aggregation in AL amyloidosis may follow similar steps to the aggregation of transthyretin (TTR), where the native tetramer has to dissociate into the monomer for aggregation to occur (Saelices, Johnson et al. 2015). Using the light chain variable domain of Mcg (PDB ID: 3MCG) in both its native dimer state, and a dimer reinforced with disulfide bonds, the authors demonstrated in a series of denaturation experiments that, dissociation of the dimer into a monomer was a necessary prerequisite for fibrillation to occur (schematic of this process illustrated in Figure 11).

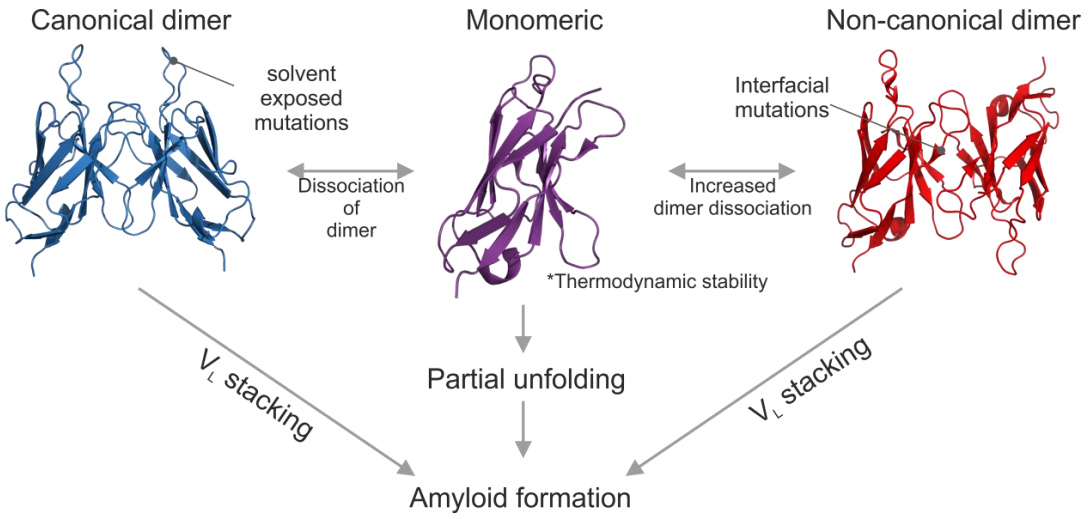


Figure 11 Schematic of amyloid formation. Diagram shows possible mechanisms of amyloid formation described by Brumshtein et al., (Brumshtein, Esswein et al. 2014). Early hypotheses suggested that dimers with different quaternary structural arrangements (canonical blue (left) and non-canonical (red)) could stack to form amyloid fibrils. A second hypothesis suggested that dissociation into the monomer was a requirement for aggregation to occur. It is now believed that dimers perform a protective role, and dissociation in the monomeric form is key to triggering aggregation. Mutations that occur within the dimer interface are able to dictate the orientation, or weaken the dimer leading to its dissociation into the aggregation prone monomer. In the monomeric state, these same mutations can lower the thermodynamic stability of the monomer, and allow it to access the partially unfolded states which favour aggregation.

Based on these findings we can now consider that the aggregation cascade in AL occurs in a multi-step process, and the somatic mutations can contribute to the aggregation potential of the V_L in a number of different ways (Brumshtein, Esswein et al. 2014). Firstly, the V_L homodimer is much more stable than the monomeric state, and dissociation of this amyloid resistant structure is required for aggregation to occur. Mutations that occur within the dimer interface may weaken the dimer and lead to its dissociation into

individual V_L monomers (mentioned earlier) (Figure 11) (Brumshtein, Esswein et al. 2014). In the monomeric state, somatic mutations are able to globally destabilise the V_L, rendering it thermodynamically unstable and more liable to aggregation. Although the monomer is far less stable than the dimer, aggregation does not occur while the monomer retains its native Ig fold; neither does it occur when the protein is completely unfolded. Rather, partially unfolded conformations, that exist somewhere between the native and unfolded state are responsible for triggering the fibrillation reaction in AL amyloidosis. To support this hypothesis, a number of studies using the V_L domains AL-09, 6aJL2 and SMA of the κIV family have shown that fibrillation can be induced in conditions (acidic pH, presence of chemical denaturants) which favour the dissociation of the dimer and unfolding of the V_L (Blancas-Mejia, Tellez et al. 2009, Hernández-Santoyo, del Pozo Yauner et al. 2010, Blancas-Mejía and Ramirez-Alvarado 2013, Brumshtein, Esswein et al. 2014). Mutations that render a variable domain thermodynamically unstable (routinely assessed by equilibrium unfolding experiments) are more liable to unfolding (requires less energy input to unfold) (Ramirez-Alvarado 2012, Marin-Argany, Guell-Bosch et al. 2015) and sample the partially unfolded states which favour aggregation than a protein that is thermodynamically stable. The thermodynamic stability of a protein refers to its physical stability, and is defined as the free energy between the folded (native) and unfolded state (Ohage, Graml et al. 1997).

This new-found evidence contrasts the early notions of light chain aggregation which proposed a possible stacking mechanism (Figure 11). (Brumshtein, Esswein et al. 2014). Now, the dimer is seen to perform a

protective role (Baden, Owen et al. 2008), and stabilising this arrangement is highly regarded as a target for therapeutic intervention. This follows a similar mechanism to the current treatment of wild type transthyretin, where the drug Tafamidis (Pfiezer) stabilises the amyloid resistant tetramer preventing its dissociation. In a follow up study to this investigation, Brumshtein and co-workers employed the use of ThT, Electron Microscopy (EM), Analytical ultra-centrifugation (AUC), crystallographic and ligand binding experiments, and highlighted the ability of two small molecules; the palindromic molecule methylene blue, and the anti-rheumatic drug sulfasalazine (from an original library of 27) to stabilise the homodimer Mcg, and inhibit its dissociation into the aggregation prone monomer (Brumshtein, Esswein et al. 2015). This is an area that is investigated within this thesis.

Recently however, there have been some slight exceptions to the thermodynamic stability and increased aggregation propensity relationship. Tanya L. Poshusta found that the variable domain AL-T03, quite unusually, is too unstable to form amyloid fibrils, instead forming non-toxic amorphous aggregates (Poshusta, Katoh et al. 2013). This is also reflected in a study conducted by Marta Marin-Argany and co-workers who found that the introduction of certain mutations into the kI family V_L AL-12, were able to reduce the thermodynamic stability of the protein too significantly that it actually precluded the domain from amyloid formation. In these scenarios it is believed that partially unfolded intermediates, the species that are key for triggering aggregation are sampled too infrequently as the mutations cause the V_L domains to populate mostly unfolded states (Ramirez-Alvarado, De Stigter et al. 2007, Del Pozo Yauner, Ortiz et al. 2008, Blancas-Mejia, Tellez et al.

2009, Randles, Thompson et al. 2009, Ramirez-Alvarado 2012, Poshusta, Katoh et al. 2013, Marin-Argany, Guell-Bosch et al. 2015).

Another layer of complexity is added to the study of V_L aggregation with the finding that mutations can have a local structural effect, causing subtle conformational alterations that lead to enhanced aggregation propensity in a mechanism that is completely independent to alterations in thermodynamic stability (Del Pozo Yauner, Ortiz et al. 2008, Blancas-Mejia, Tellez et al. 2009, Marin-Argany, Guell-Bosch et al. 2015). A growing body of evidence actually suggests that the quantity or intrinsic property of a mutation actually plays a less significant role in dictating aggregation propensity than the specific structural location to which it is introduced (Stevens 2000, Poshusta, Sikkink et al. 2009, Marin-Argany, Guell-Bosch et al. 2015). An example of this can be seen in two mutations of the V_L LEN, P40L and Q89H (which are studied later in this thesis). These mutations were able to independently destabilise the domain by +0.7 kcal/mol and 1.0 kcal/mol respectively, yet only P40L formed amyloid (Raffen, Dieckman et al. 1999, Blancas-Mejia, Tellez et al. 2009). It is suggested that a loss in stabilising contacts between neighboring loops were the source of the enhanced aggregation potential (Davis, Raffen et al. 2000). Assessing the consequence of a mutation that exceeds merely the changes to stability and other physicochemical properties is however, not a trivial task. The Ig-like fold of the variable domain is highly conserved and so circular dichroism (CD) and intrinsic fluorescence measurements of mutants harbouring somatic mutations often do not display any significant alterations to the protein backbone (Marin-Argany, Guell-Bosch et al. 2015). Higher resolution techniques such as X-ray crystallography and solution state NMR

have been used to pick out changes that range from minor side chain alterations (Randles, Thompson et al. 2009; DiCostanzo, Thompson et al. 2012) to the rather dramatic alterations in dimer arrangement we describe earlier (Figure 11).

The amino acid composition of the pathogenic V_L also varies considerably between patients, and assessing the result of each and every mutation by such methods as those described above are quite challenging. Several studies have turned to bioinformatics in order to identify trends amongst known light chain sequences. Stevens and co-workers analysed 121 κI light chains and found that the loss or gain of particular mutations could lead to specific sequence changes that rendered a light chain more or less aggregation prone. Specifically, the introduction of an isoleucine at position 27 or substitution of any residue at position 31 to an aspartic acid could increase aggregation propensity. In addition, a loss of proline residues in certain β-turns and the introduction of a glycosylation site were also contributing factors to increased aggregation (Stevens 2000, Ramirez-Alvarado, De Stigter et al. 2007, Ramirez-Alvarado 2012, Blancas-Mejía and Ramirez-Alvarado 2013). A more recent study of a similar nature was conducted by *Poshusta et al.,* who found, through the analysis of 141 κ and λ AL light chain sequences that accumulation of non-conservative mutations in β-strand A, loop C–C′ (commonly referred to as the proline 40 loop) are crucial mediators in aggregation potential (Poshusta, Sikkink et al. 2009, Ramirez-Alvarado 2012).

Such a detailed analysis is particularly useful in uncovering alterations that confer amyloidogenicity, however despite the 1000's of possible light chain combinations; the outcome of a particular mutation has been studied at

a structural level in only a handful of V_Ls. In addition, very few studies assess the outcome of a mutation on a molecular level, and take into account the structural changes can alter both tertiary and quaternary structural arrangements (Baden, Owen et al. 2008).

In summary, it is important to consider the multi-step process that lead to V_L aggregation and the interplay of multiple factors that together determine protein stability and aggregation potential. Not only can a somatic mutation destabilise the dimer, leading to its dissociation into aggregation prone monomers, but the aggregation of the V_L can be as a result of enhanced thermodynamically instability and a loss in structural integrity. Cataloguing the outcome of a mutation, and understanding how it confers enhanced aggregation propensity requires both assessment of changes to the folding (folded to unfolded state) and binding (dissociation of dimer to monomer) energies in addition to any structural changes. Such values are important in highlighting trends between previously uncharacterised V_Ls, and are an area that we investigated within this thesis using structural bioinformatics.

1.7. SMA, REC and LEN as a model system

In this study, we place focus on three homologous light chain variable domains termed SMA REC and LEN (referring to the patient initials). These three 114 amino acid (108 in kabat nomenclature) variable domains belong to the the kappa κIV family of IgG light chains (Meng, Fink et al. 2008). SMA was originally extracted post-mortem as amyloid fibrils from the lymph node of a patient suffering from AL amyloidosis (Pras, Schubert et al. 1968). Patient-REC was diagnosed with AL, where the V_L was isolated from the urine (Stevens, Raffen et al. 1995). The V_L LEN was isolated from the urine as a

Bence Jones protein of a patient diagnosed with multiple myeloma. No incidence of neuropathy or amyloid deposition was reported for LEN, despite significantly elevated levels of circulating light chain (50 g/L) in the urine (Solomon 1985). In addition to these original *in vivo* observations, characterisation of these proteins *in vitro* (largely performed by three key studies which are referred to and built upon throughout this work) reveals that SMA and REC are significantly less stable than LEN (Stevens, Raffen et al. 1995, Raffen, Dieckman et al. 1999, Davis, Raffen et al. 2000), displaying enhanced fibrillation kinetics and lower stability under destabilising conditions. Such dramatic differences arise only from a few amino acid substitutions, where SMA is altered by 8 residues (S29N, K30R, P40L, Q89H, T94H, Y96Q, S97T and I106L) and REC 14 (L15P, S27aN, V27bL, Y27dD, S27eA, N28F, S29D, K30T, Y32T, T53S, Y96P, S97T, Q100G, and L104V) (Figure 12) in comparison to LEN, which differs from the kIV germline protein by just one position (N29S) (Stevens, Raffen et al. 1995).

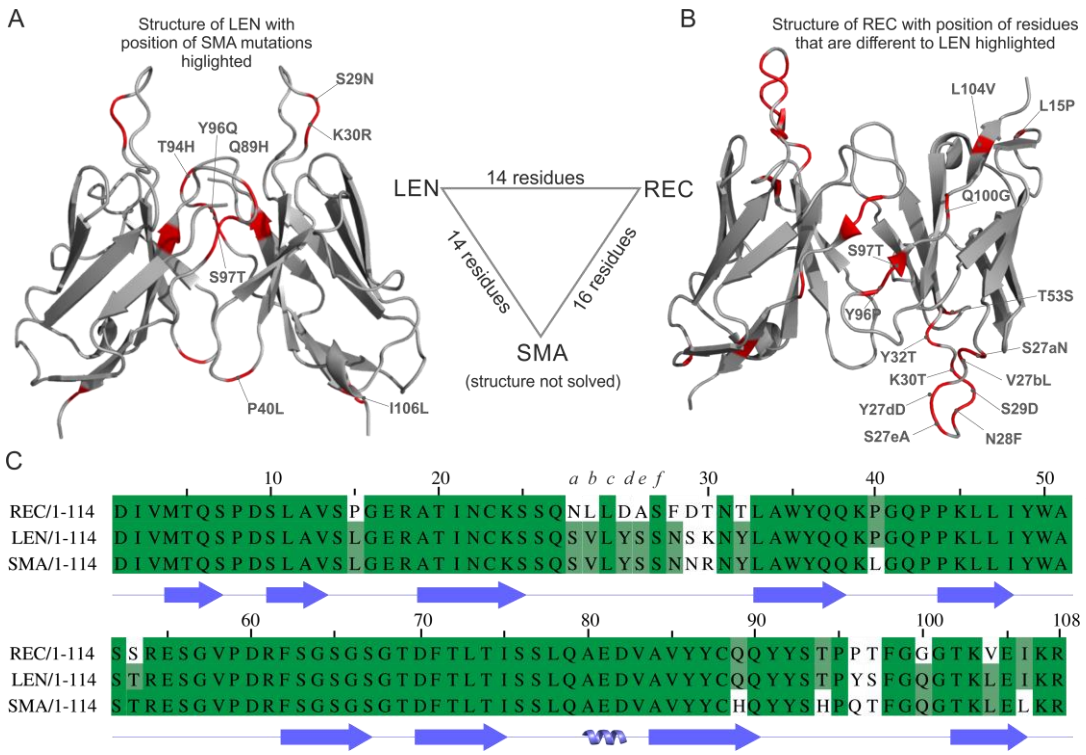


Figure 12 Structural overview and multi-sequence alignments between the pathogenic V_L proteins REC, SMA and the non-pathogenic LEN. A) X-ray crystal structure of LEN (PDB: 1LVE) is used to illustrate the locations of the SMA like mutations. B) X-ray crystal structure of REC dimer (PDB ID: 1EK3) shows an altered dimer interface. The positions of residues that differ from LEN are highlighted (red). C) Multiple amino acid sequence alignment of LEN, SMA and REC show a difference of 8 and 14 residues to LEN respectively. Conserved residues are highlighted by solid green backgrounds. Residue changes indicating somatic mutations are in white. Secondary structure elements are indicated; β-pleated sheet arrows (➡) and α-helices as coils (〰). Total residue differences are shown (centre triangle). Residue numbering, are labelled to those as defined by Kabat *et al*. Uniprot sequence P01625 for LEN and sequences noted by Stevens *et al*. for SMA and REC were used and aligned using the program clustalW (Larkin, Blackshields et al. 2007). The structure of SMA has not been experimentally determined, and is not presented.

1.8. Aims of the present investigation

In this present work, we use a combination of computational, biochemical and cellular methods to build upon the existing knowledge of AL amyloidosis and investigate two main topics. The first is the impact a mutation has on the structural integrity and thermodynamic stability of a normally non-amyloidogenic light chain variable domain. The second is the ability of $V_L\kappa$ domains SMA and LEN to internalise into cardiomyocytes, and affect cell viability.

To date, no crystallographic information of SMA is present, and so the orientation of the dimer (canonical or non-canonical example in Figure 10) is currently unknown. The first aim was to decipher the orientation of SMA using the crystal structure of known homologs as a reference structure, and employing structural informatics to assess the favourability of each model. The first experimental chapter (Chapter 2) documents this process and also includes the use of X-ray screening conditions and mass spectrometry in the attempt to experimentally verify our findings. As the V_L-V_L homodimer confers structural stability and precludes amyloid formation, the second aim was to assess the druggability of REC using the crystal structure and the model of SMA generated by computational methods to see if a cavity suitable of accommodating a small molecule stabiliser is present. This process is also documented in chapter 2.

Single mutations have demonstrated the ability to destabilise one particular V_L dimer conformation, leading to others to be favoured (Figure 10). We hypothesised that introducing SMA or REC-like symmetry mutants into the structure of dimeric LEN would destabilise the native quaternary state of this

protein, and in turn stabilise LEN that had been posed in a non-native (and thus normally unfavourable) altered 180 ^{o} quaternary state. We hypothesised that the more favourable conformation would be reflected, positively in the energetic and geometrical properties of each interface allowing us to decipher the outcome of each mutation.

Several lines of evidence suggest that AL proteins are less stable than their non-amyloidogenic counterparts due to the accumulation of somatic mutations in CDR and FR regions of the V_L. Mutations are able to stabilise/ destabilise the monomer, but they can also independently stabilise/destabilise the dimer leading to its dissociation into aggregation prone monomers. Changes to the binding free energies for each SMA-like and REC-like mutation are largely undetermined by experimental methods. This motivated us to assess the ability of a number of digital platforms to recognise changes to the intermolecular (binding) free energies. While any calculated changes cannot be benchmarked against any experimental methods, co-ordinance between the multiple programs used and providing a structural link to these mutations, i.e assessing their ability to destabilise any known highly conserved intermolecular contacts would add confidence to findings.

The next aim was to take again an informatics approach, and assess the accuracy of a number of digital platforms to calculate changes to the folding free energy as a consequence of a SMA-like or REC-like mutation (chapter 2, 3, 4). Firstly, the existing experimental data available for SMA (previously conducted by (Raffen, Dieckman et al. 1999)) would be compared to the digitally calculated values (presented within this thesis). Agreement between experimental and computational data would give validity to the

computational methods, and allow for the remaining REC-like mutations that have not been assessed by experimental methods to be calculated using this approach.

Recently, the ability of V_L from the λ family to internalise into cardiomyocytes and induce toxic effects has been confirmed by imaging methods and toxicity assays (respectively). The purpose of chapter 5 is to build upon the existing work of Immunoglobulin LC by *E.coli* synthesis, and provide a refined and a reproducible protocol of SMA and LEN expression that provides high yields suitable for experiments that are demanding of high protein concentrations. Using these recombinant V_Ls, the goal was to then assess the ability of V_Lκ4 domains to internalise which is currently unknown.

2. Predicting the dimer arrangement of the light chain variable domains

2.1. Introduction

The process of light chain formation includes the recombination of multiple-gene segments (Bruce, Alexander et al. 2002) and somatic mutations (Baden, Owen et al. 2008) that give rise to a highly diverse repertoire of proteins, where each patient presents a V_L with a unique sequence (Marin-Argany, Guell-Bosch et al. 2015). It has become increasingly evident that the variations in amino acid composition between many of the V_Ls studied to date, largely equate for their observed differences in aggregation propensity (Ramirez-Alvarado 2012, Blancas-Mejía and Ramirez-Alvarado 2013, Marin-Argany, Guell-Bosch et al. 2015). Mutations occurring within the V_L-homodimer interface are able to induce altered quaternary structural arrangements, and also lead to weakened dimer affinities and as a result, increase the population of the aggregation prone monomer in solution (Baden, Owen et al. 2008, Poshusta, Sikkink et al. 2009, Bhavaraju and Hansmann 2015). The accumulation of non-conservative mutations can also alter the thermodynamic stability of the monomer, defining its aggregation potential by allowing the protein to access partially folded conformations which are believed to be key in initiating the aggregation process (Qin, Hu et al. 2007, Randles, Thompson et al. 2009, Ara Celi DiCostanzo 2011, Blancas-Mejia, Tischer et al. 2014).

It is important to consider that not all mutations destabilise the dimer, and not all mutations destabilise the monomer. In fact, some mutations can increase V_L stability, and compensate for the destabilising effects of others

(Baden, Sikkink *et al.* 2009; Del Pozo-Yauner, Becerril *et al.* 2015). Acquiring detailed information on changes for every mutation is no trivial task, and routinely employs the use of site directed mutagenesis, recombinant protein expression, equilibrium unfolding and refolding experiments (Raffen, Dieckman et al. 1999, Blancas-Mejía and Ramirez-Alvarado 2013, Poshusta, Katoh et al. 2013), analytical ultracentrifugation in combination with fibrillation based assays and high resolution techniques such as X-ray crystallography and, NMR (Raffen, Dieckman et al. 1999, Ramirez-Alvarado, De Stigter et al. 2007, Baden, Owen et al. 2008, Baden, Randles et al. 2008).

However, these methods have limitations. Many of the mutations studied can substantially destabilise the protein to a point where it can be toxic to the bacterial host, or lead to the formation of insoluble inclusion bodies that require many difficult steps to acquire soluble, pure protein suitable for analysis (Redler, Das et al. 2016). The high resolution techniques described above are low-throughput, and often require significant quantities of protein for analysis to be performed. This means that rationalising observed changes in aggregation potential with protein stability and structural alterations is time consuming and is presumably why the number of studies that provide a complete structural – stability relationship for a complete set of mutations observed between amyloidogenic and non-amyloidogenic V_Ls are scarce.

There is evidence to suggest that crystallographic analysis of V_Ls may not be completely suitable where, tertiary and quaternary conformations are becoming trapped in states that would not necessarily be populated in solution (Peterson, Baden et al. 2010, Ahlstrom, Vorontsov et al. 2017). Due to these shortcomings, we are employing a computational approach to study how a

somatic mutation can affect both intermolecular and intermolecular energies that would otherwise be unresolved by routinely available experimental methods. Understanding the consequences of each somatic mutation and identifying trends amongst all known V_L sequences has particular importance in a disease where each patient presents a V_L with a unique amino acid composition. The raison d'être for studying the impact of each mutation is however, to highlight a putative therapeutic target, with the aim of preventing aggregation.

Recently, the ability of a repurposed anti-rheumatic drug named sulfasalazine was shown to occupy a conserved hydrophobic cavity that is present within the lambda Mcg V_L homodimer (PDB 5ACL), and stabilise the protein in a mechanism similar to that of wildtype TTR (O'Dell , Haire et al. 1996, Brumshtein and Esswein 2015). It is yet unclear if such a mechanism is applicable to other V_L that have unique interfacial residues and dimer arrangements such as those highlighted in Figure 10.

Accordingly, the present chapter aims to first determine the orientation of the SMA dimer by assessing the energetic features of multiple dimer conformations known to be populated by other, closely related V_Ls. Subsequent chapters then go on to evaluate the effect of each SMA-like and REC-like (amyloidogenic) mutations to independently affect intradomain (folding) and interdomain (binding) energies of the non-amyloidogenic protein LEN. To rationalise any observed changes to experimentally derived values, and increased aggregation propensity we manually annotate, at atomic-level detail, changes in the tertiary and quaternary structure.

The scheme of this work is summarised below (Figure 13). Firstly, dimers of LEN and REC are mutated using two independent methods (Rosetta & PyMol) (Figure 13A) so they contain the amino acid sequence that corresponds to SMA. Resulting structures are then energy minimised and their overall stereo chemical quality assessed (Figure 13B) before the energetic features of each interface are calculated, and used to decipher the most plausible arrangement of the SMA dimer (Figure 13C). In addition, changes in binding and unfolding energies of LEN as a result of a single REC-like or SMA-like mutation is calculated using a number of digital platforms (Figure 13C) in order to assess the outcome of each mutation by its ability to perturb binding or unfolding free energies.

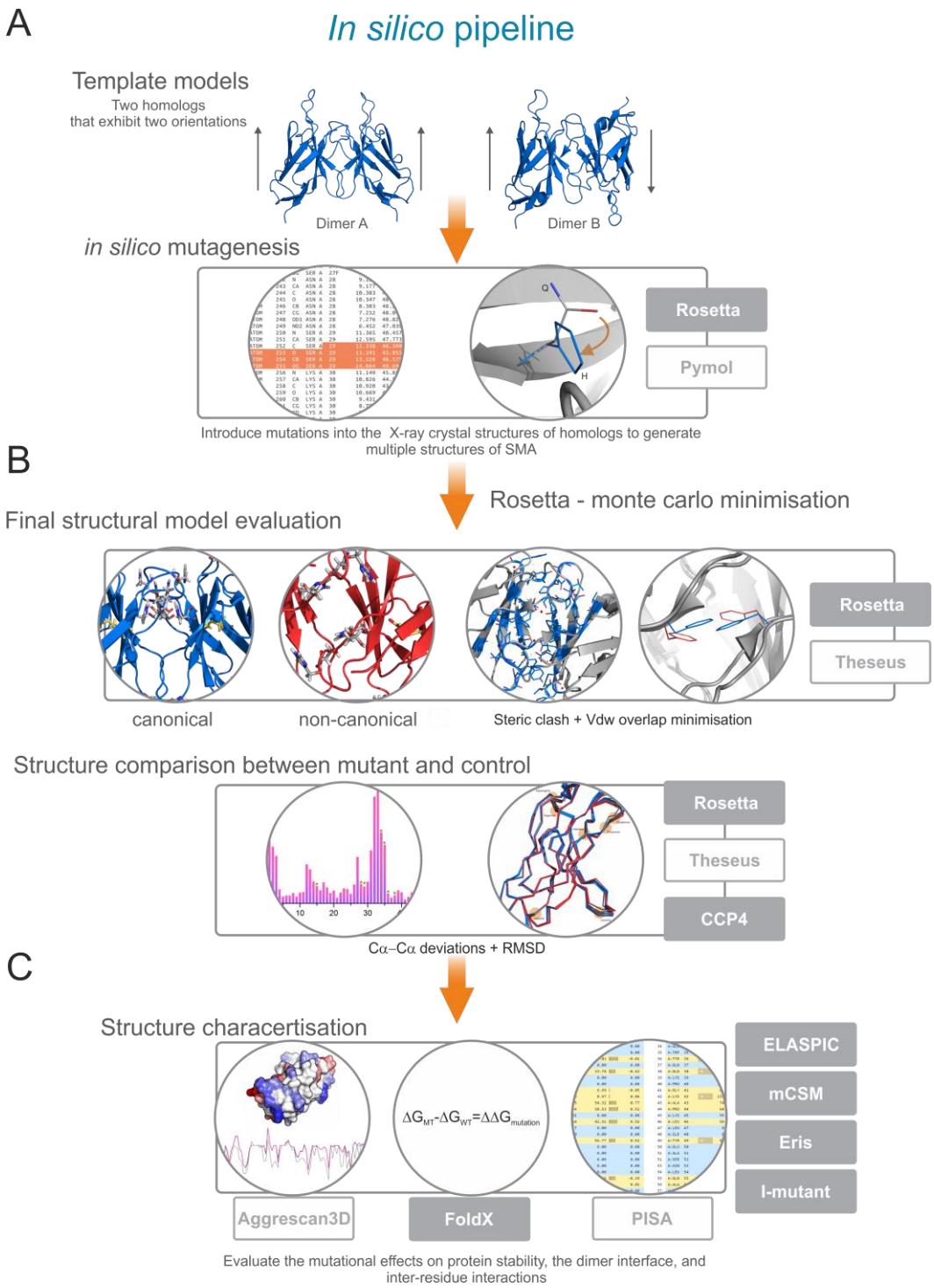


Figure 13 Computational analysis pipeline. (A) Dimers of LEN (Dimer A) and REC (Dimer B) were mutated so they contained the primary amino acid sequence corresponding to SMA using two methods (Rosetta and PyMol). All structures were subject to energy minimisation using the macromolecular suite Rosetta. (B) The stereochemical quality of the most optimal models was assessed before the features of the dimer interfaces were acquired using PISA. (C) Multiple platforms were used to assess changes in both binding and folding energies as a consequence of mutation.

2.2. Methods

The methods documented here are applicable for chapters 2, 3 and 4.

2.2.1. *In silico* mutagenesis as a method of constructing models of SMA

The structures of LEN and REC were retrieved from the PDB (PDB ID: 1LVE and 1EK3 respectively) as their biological assemblies and used as template structures for the 3D structure prediction of SMA. LEN and REC are crystallised as dimers, with resolutions of 1.95 Å and 1.90 Å respectively. To date, REC has no publication associated with the deposition. For both template models, Yasara (Van Durme, Delgado et al. 2011) was used to remove kabat nomenclature, re-numbering the amino acid sequence to 1-114 (from 108), and relabel chains of the dimer to A and B when necessary. Next, *in silico* mutagenesis was performed using two methods (Rosetta and PyMol). This was used as an internal validation check to identify any differences that were caused by the method of mutagenesis, and did not arise as a consequence of the residue substitution.

1) The PyMOL (Schrodinger 2015) mutagenesis tool was used to introduce the eight SMA-like mutations (S29N, K30R, Q89H, T94H, Y96Q, S97T P40L and 106L) into the structure of LEN. In an identical manner, REC was mutated to SMA by altering residues at the following positions; P15L, N27aS, L27bV, D27dY, A27eS, F28N, D29N, T30R, P40L, T32Y, S53T, P96Q, G100Q, V104L and I106L. Where steric clashes arose as a result of substitutions, rotamers were chosen from the Dunbrack and Cohen backbone-dependent rotamer library

implemented in PyMol (Dunbrack and Cohen 1997). These were chosen in the effort to minimise steric clash rather than the frequency in which they occur within proteins in nature, otherwise the most prevalent was favoured. As the backbone atoms of the templates are defined from crystallographic studies, rotamers were all selected with backbone dependency.

2) The second method (entirely Rosetta based) involved side chain replacement by editing the coordinate file (PDB format). For each residue to be modified, all associated side chain atomic coordinates were removed (for example Cβ, Cγ, Cδ, Cε and Nζ) leaving only the complete set of heavy backbone atoms (N, Cα, C and O). Introduction of the mutation was performed by altering the associated 3 letter amino acid type in the main chain to the 3-letter code of the new residue.

2.2.2. Energy minimisation

Mutated structures were subject to global energetic minimisation using Rosetta's Abrelax function. Backbone and side chain symmetry of the dimer was preserved during the simulation by inclusion of symmetry definition files (DiMaio, Leaver-Fay et al. 2011). Symmetry was checked by matching the coordinates of randomly chosen residues of chains A to chain B. Run time for each V_L dimer (228 amino acids in total) was approximately 90 minutes on a system equipped with a Xeon E3-1200 v3 processor clocked at 3.4GHz, and 16 gigabyte random accessory memory running at a speed of 1600MHz running Ubuntu v14.04.5.

2.2.3. Assessing the quality of generated models

In total, 100 decoy models (equalling one cluster) were generated for each mutated structure. To assess the quality of these models, the lowest 10 scoring structures from each cluster were superimposed using a maximum likelihood (ML) method as part of the program THESEUS (Theobald and Wuttke 2006). Here, models were checked to see if they converge to a similar end point, indicating that the models preserve sampling density, and do not represent a high energy conformation (Bradley, Misura et al. 2005). The lowest scoring model, representing the most optimal, was taken for further analysis. Superpose (Krissinel and Henrick 2004), as part of the comprehensive computing suite for protein crystallography (CCP4) (Winn, Ballard et al. 2011) package was used to calculate Cα-Cα deviations between template PDB and generated structures. The stereo chemical quality of the lowest energy scoring models was assessed using RAMPAGE (Lovell, Davis et al. 2003). As a control, the unaltered template structures were subject to the same energy minimisation process as those harbouring the mutation in order to ensure that any changes within each of the structures were as a result of mutation and not merely a result of movement away from the crystal contacts.

2.2.4. Interfacial residue analysis

The Proteins, Interfaces, Structures and Assemblies (PISA) service (Krissinel and Henrick 2005, Krissinel and Henrick 2007) was used to discriminate between significant and insignificant interfaces of each protein dimer, and retrieve energetic values of the highest scoring models. When using PISA to acquire values using “No-crystal” analysis of the V_L dimers, a CRYST1 card

containing unit cell dimensions (1.000 1.000 1.000 90.00 90.00 90.00 P 1 1), a remark containing details on unit cell dimensions was included in line with PISA documentation.

2.2.5. Calculating changes in binding free energy as a result of mutation

Changes in the binding free energy ($\Delta\Delta G_{bind}$) as a consequence of a single mutation were calculated by *FoldX* (Schymkowitz, Borg et al. 2005), *BeAtMuSic* (Dehouck, Kwasigroch et al. 2013), *ELASPIC (Witvliet, Strokach et al. 2016), MutaBind* (Li, Simonetti et al. 2016), *and mCSM-PPI* (Pires, Ascher et al. 2014). The energy minimised LEN crystal structure was used as the query structure for all analyses. The $\Delta\Delta G_{bind}$ in this case is the change in binding free energy between the wildtype protein dimer (LEN) and the dimer harbouring a single SMA-like or REC-like mutation. The $\Delta\Delta G_{bind}$ is calculated by the formula presented below (Eq.1).

$$\Delta\Delta Gbind = (\Delta G mutant - \Delta G wildtype) \quad \text{(Eq.1)}$$

Results are expressed as $\Delta\Delta G$ kcal/mol, where negative values ($\Delta\Delta G<0$ kcal/mol) indicate favourable interactions and are stabilising, and mutations that result in positive values ($\Delta\Delta G>0$ kcal/mol) are destabilising, (Peng, Norris et al. 2016). For *FoldX,* Rosetta energy minimised dimer of LEN (PDB ID: crystal structure) was subject to the “RepairPDB” process In line with the *FoldX* steps detailed previously in the following publications (Schymkowitz, Borg et al. 2005, Tokuriki, Stricher et al. 2007, Tokuriki, Stricher et al. 2008,

Van Durme, Delgado et al. 2011). Invoking the “BuildModel” function within *FoldX* enables a structure to be generated harbouring a single mutation, where the software then calculates the ΔΔG (Yu, Wang et al. 2015).

2.2.6. Calculating changes in folding free energy as a result of mutation

Changes in the free energy of unfolding ($\Delta\Delta G_{stability}$) as a consequence of a single mutation were calculated by *FoldX*, *DUET* (that encompassed *mCSM* and *SDM*) (Pires, Ascher et al. 2014), *ERIS (*$Ερισ$*)* (Yin, Ding et al. 2007) and the structure based version of *I-mutant 3.0* (Capriotti, Fariselli et al. 2008). Protein thermodynamic stability changes between the wildtype (LEN homodimer) and structure harbouring a single SMA-like or REC-like mutation is calculated in the formula presented below (Eq.2).

$$\Delta\Delta Gstability\ (change) = (\Delta Gmutant - \Delta Gwildtype) \qquad \text{(Eq.2)}$$

For *FoldX,* all structures were subject to the software's internal minimisation function (RepairPDB) to correct for poor stereo chemical quality. Single mutations were introduced using the “mutate residues” command with parameters (5 runs, temperature (310 K), ionic strength (0.05 M) and pH (7) that best matches the experimental conditions described previously (Raffen, Dieckman et al. 1999). For *ERIS*, substitutions were made using the medusa forcefield and flexible backbone options as previously described (Yin, Ding et al. 2007, Redler, Das et al. 2016). Other programs were used under default

parameters. For each mutation, the location (CDR or FR), and the solvent accessibility (calculated by mCSM) is presented.

2.2.7. Structure rendering

All protein structures were rendered using PyMol v1.7.5.0 (Schrodinger 2015), and imported into CorelDRAW X6 (Corel) for final processing.

2.2.8. Assessing the druggability of 3-dimensional structures

Identification of pockets that may be capable of binding small-molecule stabilisers were assessed by DogSiteServer (Volkamer, Kuhn et al. 2012). This was performed using V_L homodimers. All structures were energy minimised using the Rosetta abrelax algorithm (both wildtype crystal and mutant computationally generated structures) before assessment. The druggability score given by DogSiteServer provides an estimation of how druggable a potential binding cavity is based on a number of characteristics. Here, the scoring method evaluates geometric and physico-chemical properties that include the volume, surface, and chemical properties (lipophilic nature and overall hydrophobicity of each individual pocket. A machine learning method specifically, a support vector machine (which is evaluated to be 88% on a druggability dataset database of 1069 structure) (Volkamer, Kuhn et al. 2012) is used to discriminate each detected binding pocket between druggable and non-druggable where a score between 0 and 1 is provided. A higher druggability score (nearer 1) for a predicted pocket is estimated to be more druggable than a pocket scoring 0 (Zhang, Martiny et al. 2014). The geometric properties of each highlighted pocket were also extracted from the

server. Sites are presented mapped onto the query structure and colour coded.

2.2.8. Molecular dynamics

Simulations of V_L homodimers (both wildtype and mutant structures) were performed using either CABS-FLEX (Jamroz, Kolinski et al. 2013) or where indicated, the AMBER99SB force field with the TIP3P water model (Jorgensen, Chandrasekhar et al. 1983, Hornak, Abel et al. 2006) in Gromacs v5.1.4 (Abraham, Murtola et al. 2015). Sufficient ions were added to neutralise the system before dimers were subject to energy minimisation (500 steps), equilibrated using both NVT dynamics at a temperature of 310 K (37 ^{o}C), or 400 K (126.85 ^{o}C) (indicated in results) and NPT dynamics at an atmospheric pressure of 1.0 bar, each for 100.0 picoseconds. Full atom production simulations were performed for 150 ns (75000000 steps). Snapshots of structure trajectory and root mean square fluctuations (RMSF) were extracted from molecular dynamics simulations. Run time for each V_L dimer was approximately 13 days (11.5 ns/day) on a system equipped with an Intel Xeon E3-1240 v3 processor clocked at 3.40GHz or approximately 6.8 days (25.996 ns/day) on a system equipped with an Intel I7 5960X processor clocked at 3.0 GHz, and 16 gigabyte random accessory memory at a speed of 1600 MHz. Both systems use Ubuntu v14.04.5 as the operating system.

2.2.9. Prediction of protein aggregation

Location of aggregation hotspots were identified by a number of computational platforms.

1. **Aggregscan3D.** The Aggrescan server calculates the aggregation propensity of the query structure (3D). Query structures were assessed using a 5 Å radius, specifically for the assessment of the contribution of individual acids to aggregation prone regions (Zambrano, Jamroz et al. 2015).
2. **AmylPred2.** AmylPred2 combines a number of different algorithms (Aggrescan2D, AmyloidMutants, Amyloidogenic Pattern, Average Packing Density, Beta-strand contiguity, Hexapeptide Conformational Energy, NetCSSP, PaFig, SecStr, Tango, and Waltz) to perform a consensus prediction of amyloidogenic regions within a queried sequence (Tsolis, Papandreou et al. 2013). Agreement between 5 or more programs for a specific sequence is defined as a hit and is presented. Detailed descriptions for each of the software used are presented in the following link (http://aias.biol.uoa.gr/AMYLPRED2).
3. **FISH amyloid.** Recognises amyloidogenic segments in proteins using a machine learning approach. A detailed description for this software is presented in the following publication (Gasior and Kotulska 2014)
4. **PASTA 2.0.** The PASTA algorithm predicts which portion of the query sequence will participate in stabilising the cross β-strand of the amyloid fibril. A detailed description for this software is presented in the following publication (Walsh, Seno et al. 2014)

2.2.9. X-ray crystallography screens

Purified SMA was buffer exchanged into 20 mM sodium phosphate 20 mM NaCl pH 7.4 and concentrated to either 10 mg/mL or 30 mg/mL using a 10 kDa MWCO filter (Merck). Crystallisation screens were performed using, PEGRx (Hampton), PACT PREMIER, MIDAS, JCSG+, MORPHEUS (Molecular Dimensions), Wizard classic, and Wizard Cryo (Rigaku) screens. Crystallisation trays were stored and imaged in a Rigaku crystal mation Intelli-plate system at either 4 ^{o}C or at 20 ^{o}C. Crystals were cryoprotected with a 10% glycerol solution and stored in liquid nitrogen for transportation. Collection of diffraction data was attempted at the beamline of SOLEIL synchrotron.

2.2.10. Mass spectrometry

Collision cross-section determination of recombinant SMA was performed by mass spectrometry. Protein samples were dialysed into 100 mM ammonium acetate using HPLC grade water and spectra acquired using a Waters SYNAPT G2-S*i* HDMS. (Waters synapt, Manchester, U.K) All Mass spectrometry was performed by Dr Matthias Vonderach in the Centre for Proteome research located at the University of Liverpool. Determination of collisional section (CCS) was performed using the software MassLynx and the exact hard sphere scattering model.

2.3. Results

2.3.1. Comparative modelling: Generating SMA dimers using *In silico* mutagenesis

The deposited crystal structures of LEN (PDB: 1LVE) and REC (PDB 1EK3) show different dimer assemblies. LEN crystallised as a dimer possessing the canonical quaternary structural arrangement, whereas the amyloidogenic REC crystallised with a non-canonical altered dimer conformation. In this state, the interface of REC is twisted by 180 °, where each monomer sits in an antiparallel orientation, as illustrated previously (Figure 10). Presumably, owing to the high solubility of SMA at high concentrations (Qin, Hu et al. 2007) high resolution crystallisation data of this particular V_L have not been successful to date. As a result, information on the dimer arrangement and the interfacial residues of this structure are unknown. The absence of such information for SMA was the motivation here to see if the orientation of this V_L could be predicted computationally by generating two models of SMA, using both LEN and REC dimers in their different orientations as templates, followed by a detailed examination of each putative interface. The hypothesis here was that the most plausible orientation of SMA could be inferred by statistical analysis of the most energetically favourable out of the two likely alternatives.

As the light chains variable domains in this study contain only minor amino acid differences between them (SMA has 8 residue difference to LEN and REC has 14) (sequence alignment performed in *FoldX*), and the Ig fold is highly conserved (structure of V_L) a homology modelling approach (typically requiring only ~30% identity (Xiang 2006) was deemed unnecessary. Instead, the X-ray crystal structures of LEN and REC were mutated so they contained

the amino acid sequence of SMA. Mutagenesis was followed by an energy-based refinement step as part of the Rosetta molecular modelling suite (DiMaio, Leaver-Fay et al. 2011). The first approach was to create V_Ls in a monomeric form, where dimers could be assembled by superimposing individual monomers onto the template structures which would act as a scaffold. However, a steric clash was encountered in all structures exhibiting a canonical dimer interface where, Phenylalanine 104 (Figure 14) of the neighbouring monomers would show considerable van der Waals overlap. Of course, a clash would mean a particular dimer orientation would be sterically implausible; however, this particular observation was not a result of mutation. The clash arises due to the enhanced freedom of the residue in a monomer, where it can access a rotamer that was not accessible in the dimer. Attempts to optimise the geometry were unsuccessful, and also resulted in a loss of homodimer symmetry (Figure 14).

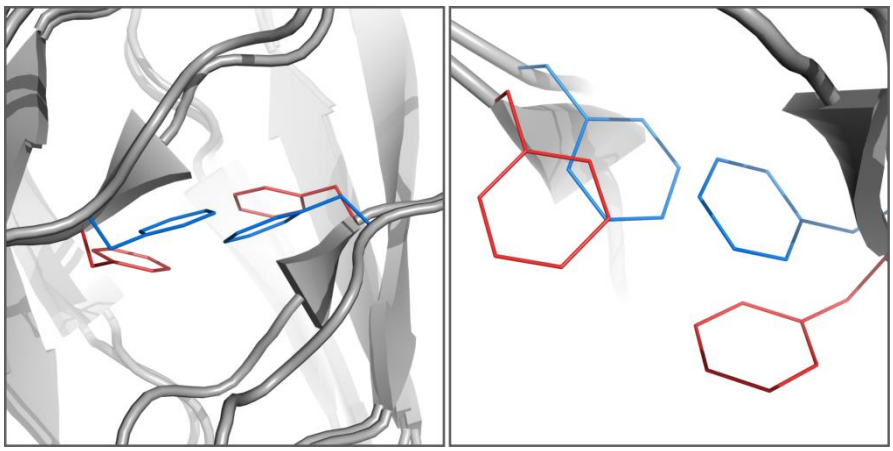


Figure 14 Steric clash example of SMA modelled on REC dimers. REC (red) and modelled SMA) (blue) side chains (minus hydrogens) of phenylalanine 104 (F96 in kabat nomenclature) reveals the close proximity of aromatic side chain that presents a steric clash.

To overcome this issue, the crystallographic structures of REC and LEN as their biological assemblies (dimer state) were extracted from the PDB, where they were each mutated to match the sequence of SMA and minimised under symmetry constraints. Here, mutagenesis was performed by two different methods to ensure that the procedure had no effect on the end structure. The first was performed by removing the atomic coordinates of the residue to be mutated side chain and relabelling the backbone to the desired residue (see materials and methods). The second was performed using the PyMOL *in silico* mutagenesis tool. Mutated structures were subject to Rosetta energy-based refinement where, to ensure convergence, 100 models were generated. For clarity, SMA models generated from directly editing the PDB files of LEN and REC are termed *SMA.1lveCLUS1,* and *SMA.1ek3CLUS2* respectively, where 1lve and 1ek3 refer to the template model PDB ID. Those mutated by the PyMOL *in silico* mutagenesis tool are termed *SMA.1lveCLUS3,* and *SMA.1ek3CLUS4*. From these models, a series of validation checks were performed before any analysis was made towards predicting the orientation of the SMA dimer. Firstly, we assessed if the generated models converged to a similar end point. Here, the ten best-ranking models from each cluster were superimpositioned using a maximum likelihood method as part of the software THESEUS. The finalised alignment was visualised in PyMOL (v 1.7.0.0).

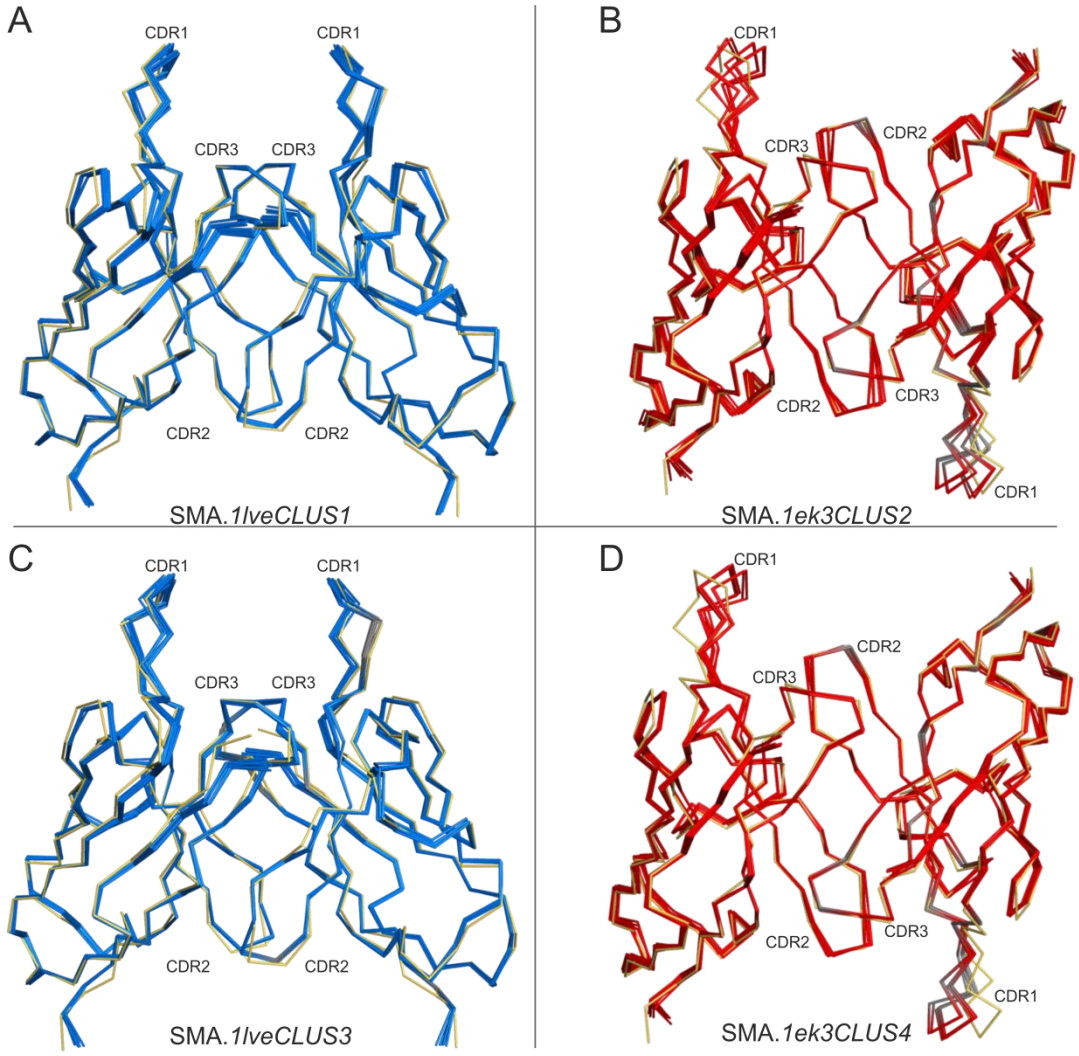


Figure 15 Conformational ensemble of lowest energy models. Each cluster represents the 10 best ranking models obtained from the Rosetta Abrelax function. Each model is a C-alpha atom trace and each cluster is coloured based on the template it was modelled from (LEN is blue, REC is RED). Panel A and B (CLUS1, and CLUS2) show models generated by directly editing the structure coordinate file, while Panels C and D show those edited using PyMOL mutagenesis tool. Figure is used to illustrate the conformation that is most populated, and see if structures during minimisation process converge to a common end point. Each coloured ensemble (Panel A-D) is aligned to the 10 best-ranking starting structure (Yellow) using a maximum likelihood method calculated in the program THESEUS (version 3.3.0). The location of mutated residues is shaded (coloured spheres). Shown also, are the locations of each CDR region.

As evidenced from the low energy ensembles displayed in Figure 15 the structures using LEN as their template show consistent movement away from the starting structure (Yellow). This suggests that the observed topology is likely to be the most native low energy structure. For models based on REC, the lowest 3 energy scoring structures do ultimately converge to a similar end point (indicated Figure 15) which suggests there are sufficient sampling steps to reach the most likely native structure. Some models do however, favour alternative conformations along the way. This is most noticeable in the CDR1 loop (indicated Figure 15) of *SMA.1ek3clus2* and *clus4 which* contains a number of SMA-like somatic mutations. It is interesting to note that, performing the same relaxation protocol for the crystal structure that act as the template structures moved in the same way as the models harbouring the mutations (not displayed). This suggests that the mutations have not triggered any major structural changes in the main chain, which is consistent with experimental data in a number of separate studies (Randles, Thompson et al. 2009, Bhavaraju and Hansmann 2015, Blancas-Mejia and Ramirez-Alvarado 2016) that have consistently shown low backbone deviations when superimposing the structures of V_L monomers. This also indicates that, any differences that could lead to the way in which the dimers forms or alterations in stability for SMA are likely to be as a result of small differences in side chain rearrangements rather than main chain alterations. This possibility is investigated further on within this thesis.

Regarding differences in the mutagenesis procedure, a comparison made between structures from each method (PyMol and Rosetta) do reveal minor differences that are mostly situated in the loop regions that have

undergone a change in sequence. We would expect that under the Relax application, Rosetta would repack all side chains into the most favourable rotamer, regardless of the mutagenesis procedure and starting point of the structure which appears not to be the case. However, the results of stereochemical quality assessment of each of the final models; *SMA.1lveCLUS1,* (96.4% residues in allowed, 3.6% in favoured and zero in disallowed regions) *SMA1ek3CLUS2* (96.4% residues in allowed, 3.6% in favoured and zero in disallowed regions) *SMA.1lveCLUS3, (*95.5% residues in allowed, 4.5% in favoured and zero in disallowed regions) *SMA.1ek3CLUS4*. (94.6% residues in allowed, 5.4% in favoured and zero in disallowed regions) which is comparable to the template structure LEN (96.4% residues in allowed, 3.6% in favoured and 0.0% in disallowed regions) and REC (96.9% residues in allowed, 3.1% in favoured and 0.0% in disallowed regions), assessed by RAMPAGE server (Lovell, Davis et al. 2003) (software used to generate a Ramachandran plot and assess ϕ,ψ angles of each residue within the structure), indicate that neither method can be considered more accurate. Instead, the lowest scoring model (best ranking) from each of the four clusters, suggesting the most optimal was taken for further analysis.

2.3.2. Can bioinformatics predict which kind of dimers can and cannot form by analysis of modelled dimer interfaces?

The mechanisms that dictate V_L homodimer orientation are complex, and appear to be defined by a number of protein-protein interactions that include hydrogen bond number, shape complementarity, electrostatic interactions, hydrophobic interactions, that each contribute to the binding energy of the

interface. (Kaplan, Livneh et al. 2011, Sowmya and Ranganathan 2015) Disruption in any number of such interactions could lead to one particular arrangement being more favourable than another. Commonly occurring mutations and identifying trends between already characterised V_L can be used to infer the likely structural consequences of another novel mutation (Stevens 2000, Poshusta, Sikkink et al. 2009). This vastly narrows down the evaluation of structural consequences. However, considering that the sequence similarity and identity between many different patient isolated V_Ls is low, and the structural consequence of many mutations us unknown, analysis of all variables for each case is somewhat time consuming and unwieldy.

Pertinent to this investigation are three studies that use LEN as a model protein. Firstly, Rosemarie Raffen and colleagues demonstrated through chromatographic techniques in a proof of principle assay, that non-naturally occurring single mutations of LEN at positions Q38R, Q38D, and Q38K could reduce the dimerisation constant of the protein, most likely through disruption of conserved hydrogen bonds situated in the dimer interface (Raffen, Stevens et al. 1998).

In a follow up study, it was found that substitution of Gln for Glu at this same position in LEN surprisingly lead to an increase in dimer association. Structural data of the Q38E mutation revealed that this single mutation induced a flipping of the LEN dimer by 180 ° (Pokkuluri, Huang et al. 1998). The authors suggested that this observed increase in dimer affinity could be attributable to an increase in the number of salt bridges, hydrogen bond and total buried surface area found in this particular arrangement. Despite being more energetically favourable than the canonical arrangement, the authors

concluded that LEN cannot adopt this interface naturally as the excess positive electrostatic potential of the dimer interface is not compensated by any other forces (Pokkuluri, Huang et al. 1998).

In an independent study, the single mutant K30T, a mutation that occurs in REC also achieved the same result, causing the LEN dimer to adopt an altered interface (Pokkuluri, Huang et al. 1998). An additional factor governing V_L dimerisation is also noted by PR Pokkuluri and colleagues who demonstrated that the substitution of Glu at position 89, (a residue that forms hydrogen bonds both inter and intramolecularly to Tyr36) for an alanine or leucine (non-naturally occurring in these proteins), also resulted in LEN adopting an altered 180 ° dimer arrangement due to the unsatisfied hydrogen bonding potential that was satisfied in an alternate arrangement (Pokkuluri, Cai et al. 2000).

We hypothesised that destabilisation of one dimer orientation would lead to an altered 180 ^{o} to be preferential, and such an arrangement would be reflected in the energetic features of that dimer interface, which could be calculated using computational methods, specifically using PISA, software designed for the analysis of macromolecular interfaces with a reported 80-90% accuracy (Krissinel and Henrick 2007).

As a control for this hypothesis, we included a structure of LEN that is posed into a flipped 180 ^{o} dimer, an arrangement that as highlighted above, is reputed to be non-physiological where it has consistently crystallised as a canonical dimer (Stevens, Raffen et al. 1995, Huang, Chang et al. 1997, Pokkuluri, Huang et al. 1998). In order to identify the most favourable quaternary arrangement of SMA out of the two possible alternatives, the

geometric and energetic features of each finalised SMA model (*SMA.1lveCLUS1, SMA1ek3CLUS2*, *SMA.1lveCLUS3,* and *SMA.1ek3CLUS4*) were compared to the experimentally resolved template structures LEN and REC. For greater comparison, all published X-ray crystal structures of LEN from the aforementioned studies which show an altered, 180 ^{o} interfaces as a result of a single amino acid mutation (Q38E, K30T, Q89A, and Q89L which corresponds to PDB ID's 3LVE, 4LVE, 5LVE and 1QAC respectively) were also included. All structures were energy minimised so they were directly comparable, and the interfaces analysed by PISA. The energetic and geometric features of each interface are shown below Table 3.

Table 3 Statistical analysis of protein-protein dimer interface of V_L domains using PISA. The PISA program was used to calculate the energetic features of the homodimer dataset presented below.

INTERFACIAL ANALYSIS OF V_L HOMODIMERS											
	LEN	**LEN MUTANTS**				**REC**	**COMPUTATIONAL MODELS**				**Control**
							Based on LEN		Based on REC		
Struc	1LVE	3LVE	4LVE	5LVE	1QAC	1EK3	SMA.1lveCLUS1	SMA.1lveClus3	SMA.1ek3Clus2	SMA.1ek3Clus4	LEN FALSE POSE
Ori[a]	A⇈B	X-ray crystal determined A⇅B					PUTATIVE A⇅B		PUTATIVE A⇅B		A⇅B
IA[b]	747.2	800.5	833.6	852.5	811.3	843.4	796.8	811.3	900.4	898.8	635.6
iNres[c]	24	24	26	24	26	26	24	26	24	26	20
BE[d]	-13.6	-2.8	-15.0	-14.6	-13.6	-14.5	-10.5	-12.1	-12.9	-14.1	-11.2
ΔG^{diss}[e]	2.1	1.4	3.5	3.1	2.1	3.0	-1.0	0.6	1.4	2.6	-0.2
$\Delta^{i}G$[f]	-10.4	-8.8	-8.4	-11.2	-10.2	-10.9	-9.6	-11.3	-9.7	-10.9	-10.0
HB[g]	6	6	10	8	6	4	2	2	4	4	2
SB[h]	0	2	4	4	4	4	0	0	4	4	0
Amyloid[i]	✘	?	?	?	?	✓	?	?	?	?	-

Struc: The name of the structure. This is either the PDB ID of the dimer, names of models generated earlier in this chapter (SMA.1lveCLUS1 and 4 and SMA.1ek3Clus2 and 4) or LEN FALSE POSE for the control structure.

a: The orientation (**Ori**) of the homodimer. Each monomer of the dimer is labelled A and B. In a canonical dimer both monomers (A and B) are orientated in the same direction (A⇈B). In non-canonical structures, B is rotated 180^{o} relative to A (A⇅B)

b: Interfacial area (**IA**) = This is the calculated buried surface area between monomers A and B and is measured in square angstrom (Å^{2}).

c: iNres = The number of amino acids that lay within the interface of A and B. In the case of non-canonical dimers, where symmetry is lost, one monomer may contribute a different amount of residues (higher or lower number) to the dimer interface.

d: Binding energy (shown as **BE** in the table) = is the binding, or interaction energy (ΔG^{int}) between individual V_L monomers. The ΔG^{int} is calculated as the difference in total solvation energies ΔG^{s} , the contact $\Delta G^{contact}$ and electrostatic ΔG^{es} interactions between individual V_L monomers in isolation and upon their complexation. The equation (3,4) used to calculate this value can be found in the following publication (Krissinel and Henrick 2007). This value also takes into account the effect of hydrogen bonds, salt bridges and disulphide bonds, although these homodimers do not contain any intermolecular disulphide bonds. Values are presented in kcal/mol^{-1}.

e: Δ^{i}**G** = Represents the solvation free energy gain upon formation of the homodimer. (Sowmya, Breen et al. 2015). Negative $\Delta^{i}G$ values indicate a stronger more hydrophobic interface and contribute to positive protein affinity (Aeschimann, Staats et al. 2017). Unlike the calculation for the BE, $\Delta^{i}G$ is calculated by the difference in total solvation energies between the homodimer and isolated monomers (Sowmya, Breen et al. 2015). However, this value does not consider hydrogen bond, or salt bridges that occur across the interface which the BE value does. Values are presented in kcal/mol^{-1}.

f: Δ**G^{diss}** = The calculated Gibbs free energy of dissociation. The dissociation pattern in this study if dimer to monomer. The ΔG^{diss} (given in kcal mol^{-1}) can be used to infer the stability of the V_L homodimer. Dimers that possess a low ΔG^{diss} in PISA are likely to have a low dissociation constants (K_d) which indicates a weak assembly (Hashimoto and Panchenko 2010, Krissinel 2011). Substitution of an amino acid that results in a dimer with a lower ΔG^{diss} than the control (LEN) are likely to be less stable and more prone to dissociation. Homodimers that possess dissociation free energies below zero ΔG^{diss} however are considered thermodynamically unstable and are coloured red. (Ravn, Madhurantakam et al. 2013). Values are expressed in kcal/mol.

The equations (1,9 and 17) used to calculate this value can be found in the following publication (Krissinel and Henrick 2007).

g: Hydrogen bond (**HB**) = The total number of hydrogen bonds that lay across the interface of the V_L homodimer

h: Salt bridge (**SB**) = The total number of salt bridges that lay across the interface of the V_L homodimer

i: **Amyloid** = If the structure is deemed amyloidogenic (✓) or non-amyloidogenic (✘) based on previous work (Stevens, Raffen et al. 1995). Structures where this data is unavailable are indicated (?)

Firstly, we used PISA to calculate the energetic properties of LEN (1LVE) in its native orientation (determined by crystallography), and compared these values to LEN in a non-native orientation, where we had twisted one monomer 180 ^{o} relative to one another, forming a non-canonical dimer. PISA calculated that native LEN is a stable assembly (Table 3) with a solvation free energy gain upon solvation ($\Delta^{i}G$) of -10.4 kcal/mol^{-1} and a free energy of assembly dissociation (ΔG^{diss}) of +2.1 kcal/mol^{-1} (Capitani, Duarte et al. 2016, Gandini, Reichenbach et al. 2017). These values have been previously used as an indication of complex stability (Sowmya and Ranganathan 2015), where the strength of the hydrophobic interaction can be estimated from the $\Delta^{i}G$ (Boyko, Rakitina et al. 2016). The more negative the $\Delta^{i}G$, the more hydrophobic the dimer interface is estimated to be, which leads to a greater affinity of the dimer (Aeschimann, Staats et al. 2017). The free energy of dissociation, here describes the transition from a dimer to a monomer (Baden, Owen et al. 2008) where a positive ΔG^{diss} (greater than 0) is indicative of a stable dimer (Krissinel 2011).

Comparing native LEN to our non-native control model reveals that there is little difference in the strength of hydrophobic interactions between the native LEN dimer and non-native LEN dimer, where each show high hydrophobic interactions ($\Delta^{i}G$ of -10.4 and -10 kcal/mol^{-1}respectively) (Krissinel 2011, Boyko, Rakitina et al. 2016). This suggests that hydrophobic interactions in this scenario do not play a large part in defining how the LEN dimer forms. Non-native LEN does however, possess fewer intermolecular interactions (only 2 hydrogen bond in comparison to the 6 present in native LEN) and less extensive surface area in comparison to the native LEN structure. This manifests into an overall weaker binding energy than native LEN (ΔG^{int} = -13.6 kcal/mol^{-1} for native LEN and -11.2 kcal/mol^{-1} for the non-native structure)

If there were no factors preventing the formation of a particular dimerization state (canonical or non-canonical) such as steric hindrance, we would expect the dimer to adopt the most energetically favourable conformation, which, based on binding energies alone, is the canonical state. Interestingly, the estimated ΔG^{diss} of the non-native LEN state model (-0.2 kcal/mol^{-1}) indicates that this complex is unstable (Krissinel and Henrick 2007, Tomovic and Oakeley 2008, Gopavajhula, Chaitanya et al. 2013, Ravn, Madhurantakam et al. 2013). Together these values (ΔG^{int} and ΔG^{diss}) indicate that LEN is most favourable in the native state, which based on crystallographic analysis that shows this to be the case, provides a reference point for us to now assess the likely dimerization mode of SMA.

Having established the ability of PISA to discriminate stable and unstable associations on our control models, we next turned our attention to the computational models of SMA. While there are slight differences between

SMA models that were generated using different methods (Table 3) several conclusions can be drawn. The estimated $\Delta^{i}G$ alone, again provides little indication which orientation is likely to be more favourable however where the strength of hydrophobic interactions are quite similar. However, the greater number of intermolecular interactions (hydrogen bonding and salt bridges) observed in both models of non-canonical SMA (*SMA.1ek3Clus2* and *SMA.1ek3Clus4* contribute to a marginally overall stronger binding energy of ΔG^{int} = -12.9 and -14.1 kcal/mol^{-1} in comparison to ΔG^{int} = -10.5 and -12.1 kcal/mol^{-1} possessed by canonical models of SMA (*SMA.1ek3Clus1* and *SMA.1ek3Clus3*) indicating that SMA is more favourable in a non-canonical state.

The most substantial difference between SMA models again is reflected in the free energy of assembly dissociation (Table 3). Both canonical SMA models (*SMA.1ek3Clus1* and *SMA.1ek3Clus3)* are predicted to be less stable than their non-canonical counterparts (*SMA.1ek3Clus2* and *SMA.1ek3Clus4)* with ΔG^{diss} values of -1.0 kcal/mol^{-1} and 0.6 kcal/mol^{-1}, and ΔG^{diss}= 1.4 kcal/mol^{-1} and 2.6 kcal/mol^{-1} respectively. The PISA calculated values of ΔG^{diss} for these canonical models suggest that *SMA.1lveClus3* is only marginally stable (with a ΔG^{diss}= 0.6 kcal/mol^{-1}) where *SMA.1lveClus1* is predicted to be unstable (ΔG^{diss}= -1.0 kcal/mol^{-1}).

To further derive more details from PISA which could indicate which interface SMA is more likely to adopt, we included a number of LEN structures that had crystallised with an altered interface due to amino acid substitutions (4LVE, 5LVE, 1QAC which possess single amino acid substitutions from LEN and REC which contains 14). At first glance we can similarities between these

dimers and the non-canonical SMA models of our study. Enhanced binding energies, positive ΔG^{diss} values and a greater number of salt bridges are reflected in both experimentally derived and our computationally predicted models, where in one case (*SMA.1ek3Clus2* and 3LVE) identical values are demonstrated.

As a final experiment, we used *FoldX* to calculate the overall stability (ΔG) of a canonical SMA model (*SMA.1ek3Clus3*) and compared it to a non-canonical SMA (model *SMA.1ek3Clus4*). *FoldX* predicted the stability of the non-canonical SMA dimer to be similar to native LEN *with* values of ΔG = -3.34 kcal/mol^{-1} and ΔG -3.06 kca/mol^{-1} respectively. The stability of the SMA dimer was predicted to be much higher (ΔG + 17.00 kcal/mol^{-1}) suggesting this assembly is not stable. As a further method for validation, we also used *FoldX* to calculate the interaction energies (ΔG^{int}) of these final models, where we again found the non-canonical model (*SMA.1ek3Clus4*) to be more favourable (ΔG^{int} = -20.21 kcal/mol and ΔG^{int} = -18.19 kcal/mol respectively). While the calculated free energy binding of the homodimer is different in *FoldX* than in PISA the differences between the two models for each program is remarkably similar (2.00 kcal/mol for PISA between models *SMA.1ek3Clus3 and SMA.1ek3Clus4*) and 2.02 kcal/mol for *FoldX* between models *SMA.1ek3Clus3 and SMA.1ek3Clus4*). Combined together, the data presented in this section suggests that SMA is likely to favour a 180 ° rotated dimer conformation, similar to that adopted by REC rather than the arrangement adopted by LEN.

2.3.3. Molecular dynamics: probing the stability of putative SMA dimers

An additional measure was taken to probe the two putative orientations of the SMA dimer by using the molecular dynamics package Gromacs (Abraham, Murtola et al. 2015). Here, the tip3p water solvation model and the force field AMBER99 SB (Cornell, Cieplak et al. 1995) were used to observe each structure over a 150-ns timeframe. As the canonical models of SMA were predicted to be unstable in comparison to non-canonical SMA models we hypothesised here that this made lead to dissociation of the dimer which could be detectable in molecular dynamics simulations. Unfortunately, very little differences between models of SMA in both orientations were observed. Although we believe the timescale to be sufficient, external pulling motions may be required to initiate dissociation of the dimer.

2.3.4. Experimental analysis of dimer by X-ray crystallography

In the attempt to acquire experimental results to support our computational investigations, recombinant SMA (method is detailed in Chapter 5 of this thesis) was purified to homogeneity and subject to a number of commercially available crystallisation screens (Figure 16).

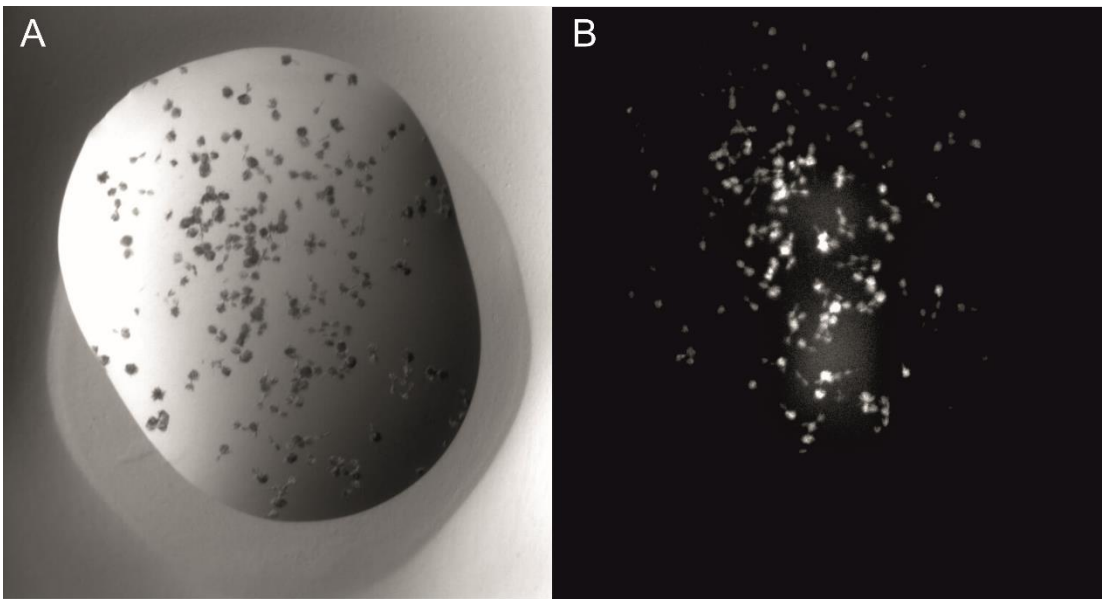


Figure 16 Example of crystal formation in condition screen. A) Sample well containing crystals under visible light, B) sample well under UV light.

A total of 962 screens were trialled at a protein concentration of 10 mg/mL and 30 mg/mL (see materials and methods for full screen listings). Despite both room temperature (25 °C) and 4 °C incubation, only 2 conditions revealed UV positive crystals. Unfortunately, crystals did not diffract at the SOLEIL beamline. Thank you to Dr. Michael Capper for attempting data collection.

2.3.5. Experimental analysis of dimer by mass spectrometry

In the attempt to experimentally validate the orientation of the SMA dimer using a different technique, drift tube ion mobility-mass spectrometry was used to derive information on the collisional cross section (CCS) of the SMA dimer. This procedure has the ability to separate the molecule both on mass (by analysis of the mass to charge ratio), but also on the size of the molecule. Visual analysis of the published crystal structures of the light chain variable domains LEN and REC), led us to hypothesise that the non-canonical dimers

may be more tightly packed in comparison to their canonical counterpart, where a more tightly packed molecule would typically display a lower CCS. Using this procedure, the CCS for recombinant SMA (the expression and purification of this V_L is detailed in chapter 5 of thesis) was acquired using a Waters Synapt G2-S*i* instrument Waters Synapt, Manchester U.K.). CCS values for both experimentally determined values and those derived from computational analysis were calculated using MOBCAL using the exact hard sphere scattering model (Shvartsburg and Jarrold 1996, Mesleh, Hunter et al. 1997). Crystal structures were subject to Amber 16 geometry optimization using the ff14SB force field prior to any analysis (Maier, Martinez et al. 2015).

The results of this experiment were compared to calculated CCS values of published crystal structures (REC that is a non-canonical dimer and LEN which is a canonical dimer).

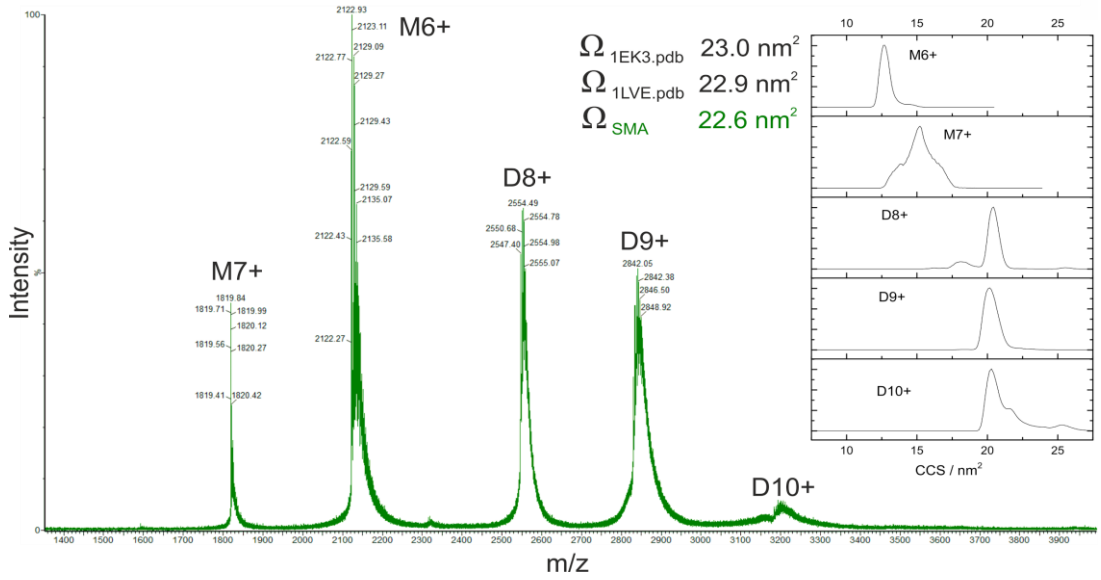


Figure 17 Intact mass spectra of SMA conformers. The molecular ion peaks M7+ and M6+ correspond to monomeric (M= monomer) SMA. D9+ indicates the presence of an SMA dimer (D= Dimer). The inset (right panel) shows analysis of the Ω (collision cross-sectional area, units are presented in nm^2). Analysis of SMA indicates the Ω is 22.6 nm^2 where REC (1EK3) and LEN (1LVE) have theoretical Ω values of 23.0 nm^2 and 22.9 nm^2 respectively which were acquired following structure geometry optimisation using Amber (Hornak, Abel et al. 2006) Mass spectrometry was performed by Dr Matthias Vonderach in the Centre for proteome research at the University of Liverpool . The spectrum was acquired using 40 µM SMA in 100 mM ammonium acetate (pH 7.2).

Despite our best efforts, the attempts to derive the orientation of the dimer were unsuccessful using this technique (Figure 17). CCS derived values for SMA were calculated to be more compact (Ω of 22.6 nm^2) than all structures, where such minor changes in Ω between the known crystal structures of LEN and REC make it difficult to be confident in the orientation of the SMA dimer.

2.3.6. Predicting changes in binding as a result of an SMA-like mutation

Comparisons made between amyloidogenic and non-amyloidogenic V_Ls have identified the ability of a somatic mutation to alter the stability and enhance aggregation potential by three independent mechanisms; 1) by weakening the homodimer interface leading to an increased number of aggregation prone monomers (Bhavaraju and Hansmann 2015), 2) or causing dimers to adopt altered conformations that are less stable 3) or by altering the thermodynamic stability of the monomer, where partially unfolded species favour aggregation. Deciphering the outcome of each mutation and highlighting trends between V_Ls is important in a disease where each patient harbours a V_L with a unique amino acid composition, and a unique set of mutations. However, routine methods of acquiring such information are not trivial. In this section we seek to address point 1 and evaluate the impact that each SMA-like (S29N, K30R, Q89H, T94H, Y96Q, S97T P40L and I106L) mutation has on the stability of the LEN dimer.

We use *the* empirical force field implemented in *FoldX* (Schymkowitz, Borg et al. 2005) to calculate changes to the binding affinity of the V_L homodimer ($\Delta\Delta G_{bind}$) upon mutation (Dourado and Flores 2016). *FoldX* was chosen due to a number of studies that have described the programs high accuracy in test cases (Thiltgen and Goldstein 2012, Studer 2015, Kumar, Rahman et al. 2017) and reported high levels of accuracy when compared to experimental data (Guerois, Nielsen et al. 2002). We also chose *FoldX* as a benchmark for several other well cited software including; *mCSM-PP, ELASPIC, Mutabind*, and *BeAtMuSic as FoldX* has been directly compared to

these programs previously (Berliner, Teyra et al. 2014, Li, Simonetti et al. 2016, Kumar, Rahman et al. 2017). Here, we ask if these digital platforms are able to highlight destabilising and stabilising residues, where agreement between calculations made between different methods would add confidence to the findings.

FoldX calculations were made with a total of five runs for each mutation. This is to allow for residues with larger side chains that may sample different rotamers on different runs to adopt the lowest energy conformation (detailed in the supplemental note of the following papers (Studer, Christin et al. 2014, Studer 2015). The reported value (Table 4) is the average of these runs. Calculations were conducted using parameters that best match the experimental conditions in the study of single point mutations of LEN performed by Raffen *et al* temperature (37 °C (performed at 310K) and ionic strength (performed at 50 mM) (Raffen, Dieckman et al. 1999). For clarity, the results from each platform are categorised into destabilising, neutral and destabilising based on the programs cut-off value.

Table 4 The effects of 8 SMA-like mutations on the protein-protein interaction energies of the LEN homodimer. Residues that *FoldX* indicated were part of interface are highlighted (✓). Units for energy calculations ($\Delta\Delta G_{bind}$) are expressed in kcal/mol. Software and their associated cut-off for a neutral mutation (indicated by +-) are as follows: *FoldX* +-0.46 kcal/mol; *ELASPIC* +-0.46 kcal/mol; *MutaBind* +-0.86 kcal/mol; *mCSM-PPI* +-0.5 kcal/mol; BeAtMuSiC (+-0.47 kcal/mol). Colours are used to indicate no change (blue) increase to dimer stability (green) and decrease to dimer stability (red). Residues previously shown to dictate the orientation of the LEN dimer (Q38E, K30T, Q89A, Q89L were also calculated for comparison).

	FoldX	BeAtMuSiC	MutaBind	mCSM-PPI	ELASPIC	**Location**	**Interface**
Q38E	2.47	0.93	1.98	0.22	0.18	FR2	✓
K30T	0.00	0.07	0.91	1.24	1.33	CDR1	✗
Q89A	2.35	-0.11	2.40	0.43	0.88	CDR3	✓
Q89L	0.21	-2.45	2.12	0.55	0.60	CDR3	✓
S29N	0.00	-0.34	-0.10	0.36	n/a	CDR1	✗
K30R	0.01	-0.29	0.48	0.70	0.43	CDR1	✗
P40L	0.01	0.03	0.11	0.09	n/a	FR2	✗
Q89H	3.01	-2.04	2.78	0.29	1.44	FR3	✓
T94H	0.42	0.43	0.84	0.74	-0.10	CDR3	✓
Y96Q	1.27	4.22	2.73	2.49	0.55	CDR3	✓
S97T	0.07	-0.56	0.62	0.67	0.30	CDR3	✓
I106L	0.00	-0.16	1.01	0.57	n/a	FR4	✗

To first test the ability of *FoldX* to recognise mutations that have previously been able to alter the dimerization state of LEN (causing it to adopt flipped dimers) we introduced the following mutations; Q38E, K30T, Q89A and Q89L into the native structure of LEN in four separate sets of analysis. Surprisingly, only Q38E and Q89A were predicted to reduce dimer affinity ($\Delta\Delta G_{bind}$ = 2.47 kcal/mol and 2.35 kcal/mol respectively) where K30T and Q89L were predicted to be neutral ($\Delta\Delta G_{bind}$ = 0.00 kcal/mol and 0.21 kcal/mol respectively. *Mutabind* and BeAtMuSic also predicted this mutation to be neutral ($\Delta\Delta G_{bind}$ of 0.91 kcal/mol and ($\Delta\Delta G_{bind}$ 0.07 kcal/mol respectively). On the other hand, *ELASPIC*, *and mCSM-PPI* both estimated this mutation to be deleterious to dimer binding affinity with binding free energies of +1.33 and +1.24 kcal/mol (respectively).

A likely explanation for the discrepancies in these values lays in how the programs work. *ELASPIC* is the most modern software from this selection, and while the program still uses *FoldX* at its core, it is trained on a modified. SKEMPI dataset (database that details how missense mutations affect protein-protein complexes) (Moal and Fernandez-Recio 2012, Pires, Ascher et al. 2014) which *FoldX* is not (instead using FOLDEF (*FoldX* energy function) and a selection of 10 different databases consisting of over 1088 mutations to base its energy predictions (Guerois, Nielsen et al. 2002). This reportedly gives ELASPIC a greater degree of accuracy than *FoldX*, with a correlation coefficient of 0.77, and standard deviation of σ = 0.002 in comparisons to *FoldX'*s 0.83 and σ = 0.46 (in the software's latest iteration which was 0.81 previously) between the experimental values (provided by the database which they were trained upon) and predictions made by the software (Guerois,

Nielsen et al. 2002, Berliner, Teyra et al. 2014). ELASPIC also benefits from an enhanced method of being able to discriminate between interface and non-interfacial residues with improvements over *BeAtMuSiC* (Berliner, Teyra et al. 2014). For the K30T mutation the latter point is certainly explains the findings, where *FoldX*, *BeAtMuSic* and *Mutabind* do not recognise this mutation to be part of the dimer interface, and so, its substitution has no impact on the stability $\Delta\Delta G_{bind}$.

For Q38E, there is also a mismatch in the predictions. While *FoldX*, *Mutabind* and BeAtMuSic predicted this mutation to be destabilising ($\Delta\Delta G_{bind}$ of 2.47, 0.93 and 1.98 kcal/mol respectively) *mCSM-PPI* and *ELASPIC* were unable to predict the deleterious nature of this mutation. This outcome is slightly harder to explain as each program recognised glutamine to be part of the dimer interface. Substitution of a glutamine to glutamic acid results in no perceivable structural changes (as the amino acids structures are almost identical) when closely examined in *the* software's visualisation panel. We would expect that the introduction of a charged residue at this position however would be disruptive, given the ability of this mutation to destabilise the dimer interface of LEN leading to altered conformations in this study, and a recent publication that documents a large lab scale analysis of 34,373 mutations where the authors also found introducing Glutamic acid at protein interfaces be highly destabilizing (Gray, Hause et al. 2017). The atom classification systems incorporated into *mCSM-PPI* (one, binary and Pmapper pharmacophoric) and the ability to consider the different properties of amino acids and their local 3d- structural environment should be able to detect this change based on the information provided within the software's documentation

(Pires, Ascher et al. 2014) yet this is not the case. For *ELASPIC*, which we mention here again, is trained on the SKEMPI database (Moal and Fernandez-Recio 2012) it is conceivable that thermodynamic information for such a mutation, located within a protein-protein complex and deposited within the PDB (which SKEMPI is based upon) is limited enough to prevent accurate predictions (2317 mutations for 150 proteins that contain experimental data) (Pires, Ascher et al. 2014).

Q89A, Q89L and the SMA mutation Q89H also provide a set of mixed results. *BeAtMuSic* was consistently unable to predict the destabilising nature of any mutation at position 89 and in two cases predicted the mutation to be stabilising to the dimer interface (Q89L and Q89H with a $\Delta\Delta G_{bind}$ of -2.47, and -2.04 respectively. A clear limitation of this software which is arguably a reason for these observations, is that *BeAtMuSic* is trained on a subsection of the complete SKEMPI dataset, achieving a correlation coefficient of 0.47, and a standard deviation of σ = 1.80 (when comparing to experimentally determined $\Delta\Delta G$ values) (Dehouck, Kwasigroch et al. 2013, Pires, Ascher et al. 2014). The results for the Y96Q mutation are quite interesting. There was much more consistency between platforms with a calculated $\Delta\Delta G_{bind}$ of +1.27 4.22, 2.73 2.49, and 0.55 kcal/mol for *FoldX, BeAtMuSic, Mutabind, mCSM-PPI and ELASPIC* respectively suggesting that this mutation is able to significantly destabilise the dimer and is a key residue we wish to investigate in further analysis.

We recognise these discrepancies to be a weakness of the programs presented here. Inconsistencies between these predictors have also been noted before (Kumar, Rahman et al. 2017). It is also important to document

that in the original *FoldX* publication the authors provide a list of possible reasons to why discrepancies between experimental values and calculated values were observed in their test data, but are unable to answer this for all mutations (Guerois, Nielsen et al. 2002). For this reason, we go onto assess using a different procedure, conducting further structural analysis and combine these results with experimental data in subsequent chapters to reach a more educated conclusion into the destabilising, neutral or stabilising nature of each mutation.

We conclude here that while the predictions performed using different platforms are not in agreement, the consistency between programs (that are trained on different datasets) demonstrated for mutation Y96Q (of the CDR3 region) suggests that out of this set of mutations, this residue may be key for altering dimer stability.

2.3.7. A single mutation can dictate the dimer orientation; can this be pinpointed using bioinformatics?

Having demonstrated the inconsistencies in the programs *FoldX*, *BeAtMuSiC*, *MutaBind*, *mCSM-PPI* and *ELASPIC* to predict the destabilising/ stabilising effect of a mutation on the LEN dimer, we opted for a different approach. Instead of using the programs listed above to assess changes to $\Delta\Delta G$ based on the datasets which they were trained, we wished to assess if introducing a mutation directly into the crystal structure of LEN and performing a series of energetic minimisation steps could provide these answers. We also asked that if one mutation was shown to destabilise the native LEN dimer (canonical) could this same residue stabilise the non-canonical arrangement? This could indicate which mutations lead to “flipped dimers” (those with altered quaternary arrangements) and which mutations may completely abolish dimerization capability.

To assess this possibility we used in silico mutagenesis to introduce double SMA-like mutations (a single mutation introduced into each monomer of the dimer) into the native LEN dimer (canonical) and into the structure of LEN that is posed into a flipped 180 ^{o} arrangement (this quaternary arrangement is illustrated in Figure 10B and was previously used as a control in Table 3) which does not occur natively (Stevens, Raffen et al. 1995, Huang, Chang et al. 1997, Pokkuluri, Huang et al. 1998). The energetic features of the best-ranking structure were calculated by PISA.

.

Table 5 Interfacial analysis of LEN harbouring a single SMA-like mutation. A single SMA-like mutation was incorporated into the dimer of LEN (PDB ID: 1LVE). Structures were posed into two orientations (canonical A⇈B and non-canonical A⇅B with arrows indicating the direction of the monomer) and the energetic features of each interface calculated by PISA. Descriptions and their abbreviations are shown previously in Table 3. Colours are used to indicate no change (blue) values that stabilise the homodimer (green) and values that destabilise the homodimer (red). In the case of non-canonical dimers, where symmetry is lost, one monomer may contribute a different amount of residues (higher or lower number) to the dimer interface.

A: INDIVIDUAL RESIDUE CONTRIBUTIONS									
	LEN	S29N	K30R	P40L	Q89H	T94H	Y96Q	S97T	I106L
Ori[a]	X-ray crystal determined canonical dimer A⇈B								
IA[b]	747.2	742.3	745.2	776.7	728.1	744.6	693.1	723.7	753.3
iNres[c]	24	24	24	26	22	22	22	20	24
BE[d]	-13.6	-13.8	-13.6	-14.5	-12.2	-14.4	-10.4	-13.6	-13.6
ΔiG[e]	-10.9	-10.2	-11.0	-11.8	-10.5	-12.6	-8.7	-10.0	-10.9
ΔG^{diss}[f]	2.1	2.3	2.2	3.0	0.8	2.9	-1.0	2.1	2.1
HB[g]	6	8	6	6	4	6	4	8	6
SB[h]	0	0	0	0	0	0	0	0	0

B: INDIVIDUAL RESIDUE CONTRIBUTIONS									
	LEN	S29N	K30R	P40L	Q89H	T94H	Y96Q	S97T	I106L
Ori[a]	Non-canonical “flipped” dimer A⇅B								
IA[b]	635.6	699.6	809.4	720.6	717.5	878.0	776.2	802.5	698.8
iNres[c]	20	19	23	21	20	25	22	23	19
BE[d]	-11.2	-10.4	-14.0	-11.7	-13.0	-14.5	-11.3	-12.0	-8.1
ΔiG[e]	-10.3	-8.0	-10.7	-8.4	-10.6	-11.2	-7.1	-9.7	-10.5
ΔG^{diss}[f]	-0.2	-0.5	2.5	0.3	1.5	3.0	-0.2	-0.6	-1.0
HB[g]	2	2	4	4	2	4	6	2	2
SB[h]	0	4	4	4	4	4	4	4	4

Firstly, we assessed which mutations were unable to destabilise the native LEN dimer. As illustrated, (Table 5 part A) structures harbouring mutants S29N, K30R, S97T, and I106L show either a minor increase, or no substantial difference in the energetic features of the dimer (judged by changes to ΔiG, the binding energy and the $\Delta G^{diss)}$ and so are unlikely to have any effect on the stability of the LEN (A⇈B) dimer.

The mutations P40L and T94H demonstrate an increase in some of the energetic features of the canonical dimer interface (increase in BE of 14.5 and -14.4 kcal/mol over the native LEN structure which has a calculated BE of -13.6 kcal/mol and increase in free dissociation energies ΔG^{diss} 3.0 and 2.9 kcal/mol). Substitution of a threonine for a histidine at position 94 results in a loss of interfacial residues, and as a consequence there is diminished interfacial area (744.6 kcal/mol) however, there is an increase in the strength of hydrophobic interaction (ΔiG= -12.6 kcal/mol) and binding energy (14.4 kcal/mol) over the native LEN structure (ΔiG = -10.9 kcal/mol and -13.6 kcal/mol respectively) suggesting that this mutation is likely to be neutral/stabilising to the dimer. Edward G. Randles *et al.* and colleagues suggested that the P40L mutation may be able to dictate the orientation of the dimer (Randles, Thompson et al. 2009). However, we predict overall that neither of these mutations is compromise the stability of the native LEN dimer.

In previous work P40L has demonstrated a profound impact on the stability and aggregation of the protein, and was able to promote fibril formation under physiological conditions, without the need for addition of

preformed misfolded protein templates or ‘seeds’ (Raffen, Dieckman et al. 1999, Davis, Raffen et al. 2000). These original experiments however were performed in conditions where the protein is in its monomeric state. Dissociation of the dimer into monomers is believed to be a necessary pre-requisite for aggregation to occur ((Brumshtein and Esswein 2015)). It is important to clarify there then, that the results presented in Table 5 do not assess the ability of this mutation to destabilise the monomer, only the dimer interface and all energetic and geometric values calculated reflect this. This ability of the P40L mutation to destabilise the monomer – independent of the dimer is later investigated in this thesis where we document an alternative explanation to the destabilising nature of this residue to that currently published. The observations here (PISA and those presented in Table 4 despite the overall inconsistencies in other mutations the programs unanimously agreed P40L to be neutral) indicate that P40L is unlikely to destabilise the native dimer.

The mutation Q89H was calculated to destabilise the native LEN dimer, leading to a reduction in the BE (-12.2 kcal/mol) ΔiG (-10.5 kcal/mol) and more substantially, the ΔG^{diss} (-0.8 kcal/mol) which was reduced 1.3 kcal/mol over the native LEN structure. We note here again that a dimer with a low ΔG^{diss} would likely have a low *Kd,* suggesting this assembly is only marginally stable (Hashimoto and Panchenko 2010; Krissinel 2011). These values may indicate that Q89H has the ability to destabilise the native LEN dimer.

By and large, Y96Q is the most destabilising of all SMA-like mutations to the native LEN dimer. Introduction of this mutation substantially diminishes the BE, the ΔiG and has a ΔG^{diss} of -1.0 kcal/mol suggesting this assembly is

unstable (Krissinel and Henrick 2005). Out of all mutations this corresponds to the values previously presented in Table 4 and may be a true indication to the destabilising nature of this mutation.

Having now assessed the ability of each SMA-like mutation to destabilise the native canonical LEN dimer (Table 5 part A - A⇈B) and narrowed down our search to just two residues (Q89H and Y96Q that can destabilise the native dimer), we asked if these same residues could stabilise the non-native LEN dimer. Based on the calculations of PISA (Table 5 B), it appears that Q89H is able to stabilise the altered dimer arrangement leading to an increase in BE, ΔiG and ΔG^{diss} over the non-native LEN structure, and having more favourable values than the LEN structure harbouring this mutation. The outcome of this mutation on dimer stability has not been tested experimentally, however it is interesting to note that a single mutation located just 2 residues apart, Y87H was able to dictate the orientation of κI light chain which crystallised as a dimer with a 180 ° rotated interface (Peterson, Baden et al. 2010). It is therefore arguable that Y89H may be responsible for forming dimers with an altered interface. Y96Q on the other hand is poorly tolerated in this alternative dimer orientation. The mutation is slightly better tolerated in this interface than the native LEN dimer (based on a higher BE) however the structure is still classed as unstable with a calculated ΔG^{diss} of -0.2 kcal/mol.

In a similar mechanism to the aggregation of TTR, mutations that destabilise the quaternary state leading to dissociation into monomeric species is believed to be a necessary pre-requisite in the aggregation of light chain variable domains (Baden, Owen et al. 2008, Brumshtein, Esswein et al. 2014). Our data suggests that Q89H may induce flipped dimers that possess reduced stability over the native LEN dimer. Y96Q on the other hand could abolish dimerization ability. This has not been shown previously, and could indicate that these mutations are able to increase the likelihood of dimer dissociation into the aggregation prone monomer.

We acknowledge that these computational methods are energy model approximations and such values would need experimental validation using a technique such as analytical ultra-centrifugation however the data provides an insight to which residues can alter destabilise the LEN dimer which has previously been unknown.

2.3.8. Small molecule stabilisation of the SMA homodimer

In the introduction to this thesis, the ability of the small molecule, sulfasalazine to bind within the dimer interface of the light chain variable domain of Mcg was highlighted. This stabilised the native dimer and inhibited its dissociation into the aggregation prone monomer (Brumshtein, Esswein et al. 2015). Co-crystallisation of the Mcg dimer in complex with sulfasalazine (PDB ID: 5ACL) and methylene blue (PDB ID: 5MCG) in this study revealed that both agents were able to occupy the concave shaped intermolecular cavity at three distinct locations (illustrated in Figure 18 (Brumshtein and Esswein 2015).

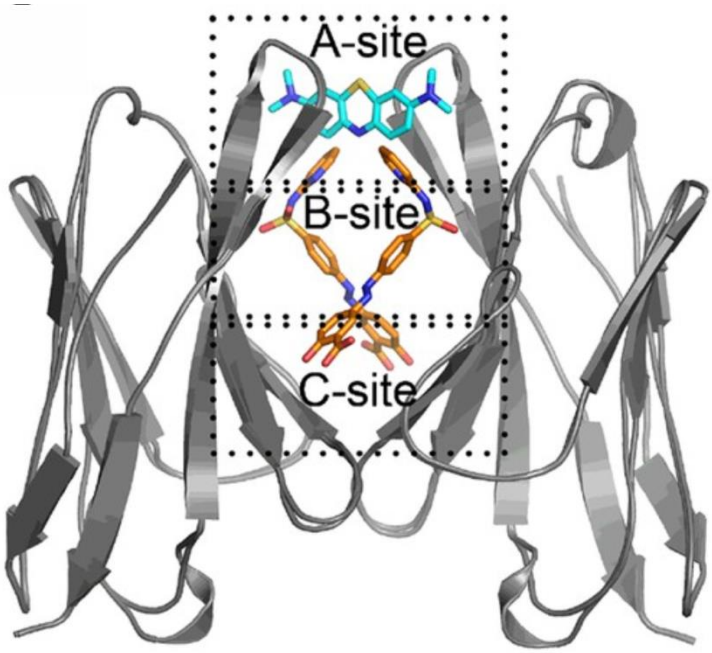


Figure 18 Overlaid ligand bound Crystal structure of V_L-Mcg. Structure and Image was adopted from *Brumshtein et al.,* and used here to highlight the binding of a single methylene (blue) and two sulfasalazine molecules (orange) at three sites located within the Mcg interfacial cavity labelled A-B-C (Brumshtein and Esswein 2015).

This cavity is maintained amongst many V_Ls, suggesting the findings could have a wider breadth for therapeutic application. However, as Mcg has a canonical dimer interface (Figure 10) it is yet unclear if such a strategy is applicable to other V_L-homodimers which possess an alternate, flipped dimer arrangement, where the interfacial residues and intermolecular landscape are vastly different. This gap in the literature motivated us to assess the druggability of SMA, which possesses minimal sequence identity to Mcg (52.75% assessed by Mustang sequence alignment algorithm (Konagurthu, Whisstock et al. 2006) and as our computationally generated models suggest a 180 ° flipped dimer arrangement similar to that of REC, AL-T05, and kI Y87H variant, rather than a canonical Mcg dimer. To this end, we employed the grid-based pocket detection system of DoGSiteScorer, a digital platform that identifies potential binding sides, and assess the druggability of structures based on geometric properties (size and shape) as well as their physiochemical composition such as amino acid composition, hydrophobicity and the solvent accessible lipophilic surface (for full description see (Volkamer, Kuhn et al. 2012, Zhang and Li 2017)). To first assess the accuracy of this software for this case, we used Mcg (PDB ID: 5ACL) as the query structure, where any similarities between pockets detected by DoGSiteScorer and the experimentally (X-ray crystallography) determined binding site of methylene and sulfasalazine would add confidence to our findings. This is a similar tactic employed by Zhang *et al*., (Zhang and Li 2017) (Table 6 Figure 19).

Table 6 Predicted druggable sites for the canonical Mcg dimer (5ACL) calculated by DogSiteServer. Identified pockets of the Mcg structure, pocket properties and scores are presented in best ranking order. Drug scores range from 0-1 where 1 indicates a more druggable pocket. Table colours correspond to mesh in Figure 19.

Pocket	Volume Å^3	Surface Å^2	Drug score
P0	1095.94	1205.44	0.78
P1	552.13	647.39	0.71
P2	306.11	663.21	0.36
P3	110.08	296.16	0.16

Table 6 shows the calculated pockets, their descriptors and associated druggability scores for the Mcg structure which were identified by DogSiteScorer. Druggability scores range from 0–1, where higher values indicating a more druggable pocket (Volkamer, Kuhn et al. 2012, Sivakumar and Niranjali Devaraj 2014). In total, DogSiteScorer server was able to identify 4 pockets (orange P0, green P1, purple P2, and red P3). As illustrated in Figure 19, pockets P0, P1 and P2 are located within the dimer interface (orange, green, and purple respectively). Site P3 however (red) is located too far from the dimer interface for it to have any meaningful use, any small, molecules that bound to this pocket would be unable to stabilise the dimer. Remarkably, the two pockets with the highest druggability scores (P0 orange 0.78 and P1 green mesh 0.71 Figure 19) correspond well to the ligand binding sites of the Mcg structures (Figure 18) (Brumshtein and Esswein 2015).

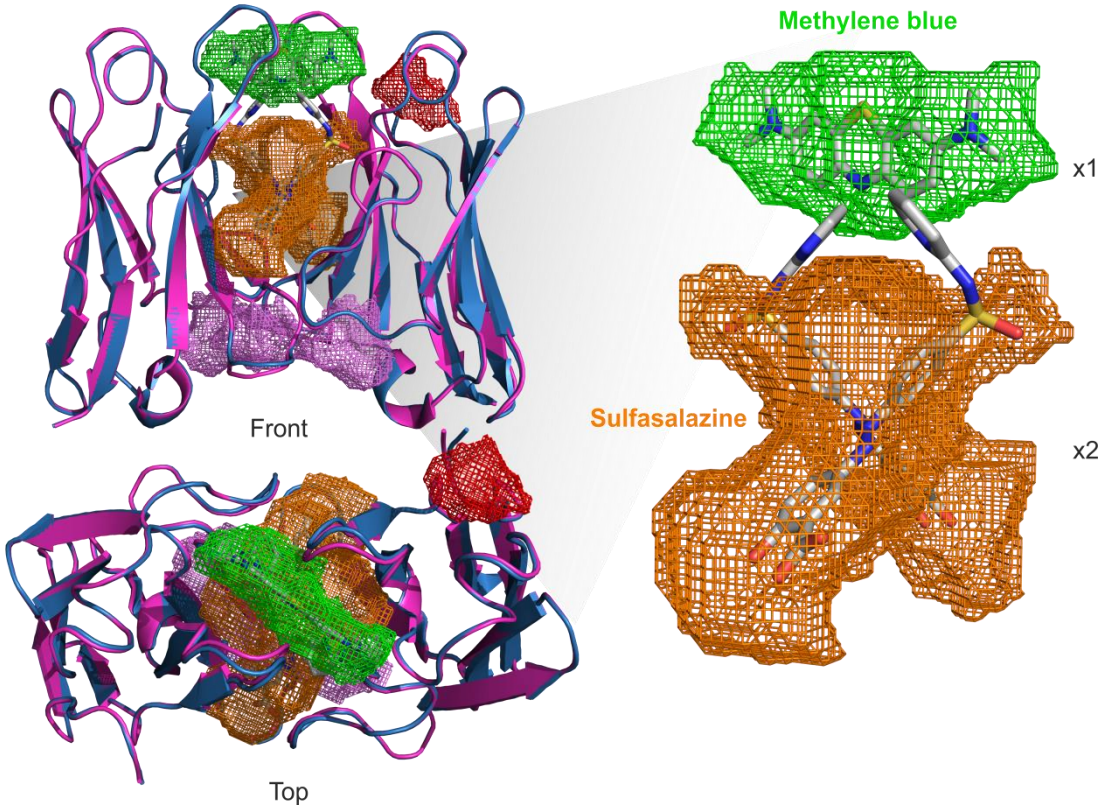


Figure 19 Predicted druggable sites for the canonical Mcg dimer (3mcg). DogSiteScorer was used to predict pockets of the Mcg dimer. The server was able to predict a total of 4 sites (all shown). The top scoring sites (P0 orange, and P1 green) both occur within the hydrophobic cavity of the Mcg dimer. Pocket P01 possesses a drug-score of 0.78, a simple score of 0.71 and occupies the largest volume of 1095.94 $Å^3$. Pocket P1 (green) possesses a drug-score of 0.71, a simple score of 0.42 and occupies the second largest volume of 552.13 $Å^3$. As indicated by the superimposition of Mcg structures in complex with sulfasalazine (5ACL purple), and methylene (3ACL blue) the best ranking pockets (P0 and P1) match well with the actual ligand binding site found within the crystal structures.

Overlaying the DogSiteServer predicted cavities (orange and green mesh) with ligand bound crystal structures of Mcg (Figure 19) reveal the extent of this accuracy. As illustrated, the co-crystalised ligands (a single methylene molecule bound to the top of the cavity and two sulfalazine molecules bound within the central cavity of the dimer interface) fit within the calculated binding

sites (green and orange mesh). Confident in the performance of DogSiteServer, we next used the software to probe the druggability of SMA.

Table 7 Predicted druggable sites for the non-canonical SMA model calculated by Dogsiteserver. Identified pockets of the SMA structure, pocket properties and scores are presented in best ranking order. Drug scores range from 0-1 where 1 indicates more a druggable structure. Table colours correspond to mesh in Figure 20.

Pockets	Volume Å^3	Surface Å^2	Drug score
P0	776.32	938.69	0.84
P1	425.09	747.88	0.67
P2	424.19	722.51	0.68
P3	360.32	513.24	0.60

Table 7 shows the calculated pockets and their descriptions for the SMA structure. Manual inspections of these pockets on the queried SMA structure show that none occur within the dimer interface (Figure 20). In comparison to the Mcg structure, SMA does not possess such a deep concave interface (Figure 20 top). Instead, the highest scoring pocket (P0 blue mesh Figure 20 with a drug score of 0.84) is located on the surface of the dimer, and involves a total of 20 residues, specifically $_{35}$YQQK$_{40}$ $_{41}$QPPKLL$_{48}$ $_{54}$EVP$_{58}$ $_{60}$RFED$_{65}$ Y86 and H89 and F98 (identical residues involved in each subunit). Interestingly, pocket P0 in the SMA structure scores higher than in Mcg (P0 0.84 Table 7) in comparison to P0, drug score of 0.78 (Table 6). However, the surface coverage of this pocket is substantially less, in addition, covering only 938.69 cubic angstroms (Å^3) in comparison to 1205.44 Å^3. Each of the other sites, also span the two protein chains, and score reasonably well (0.67, 0.68,

0.60 drug score) suggesting there may be multiple sites to accommodate small-molecule stabilisers (Figure 20).

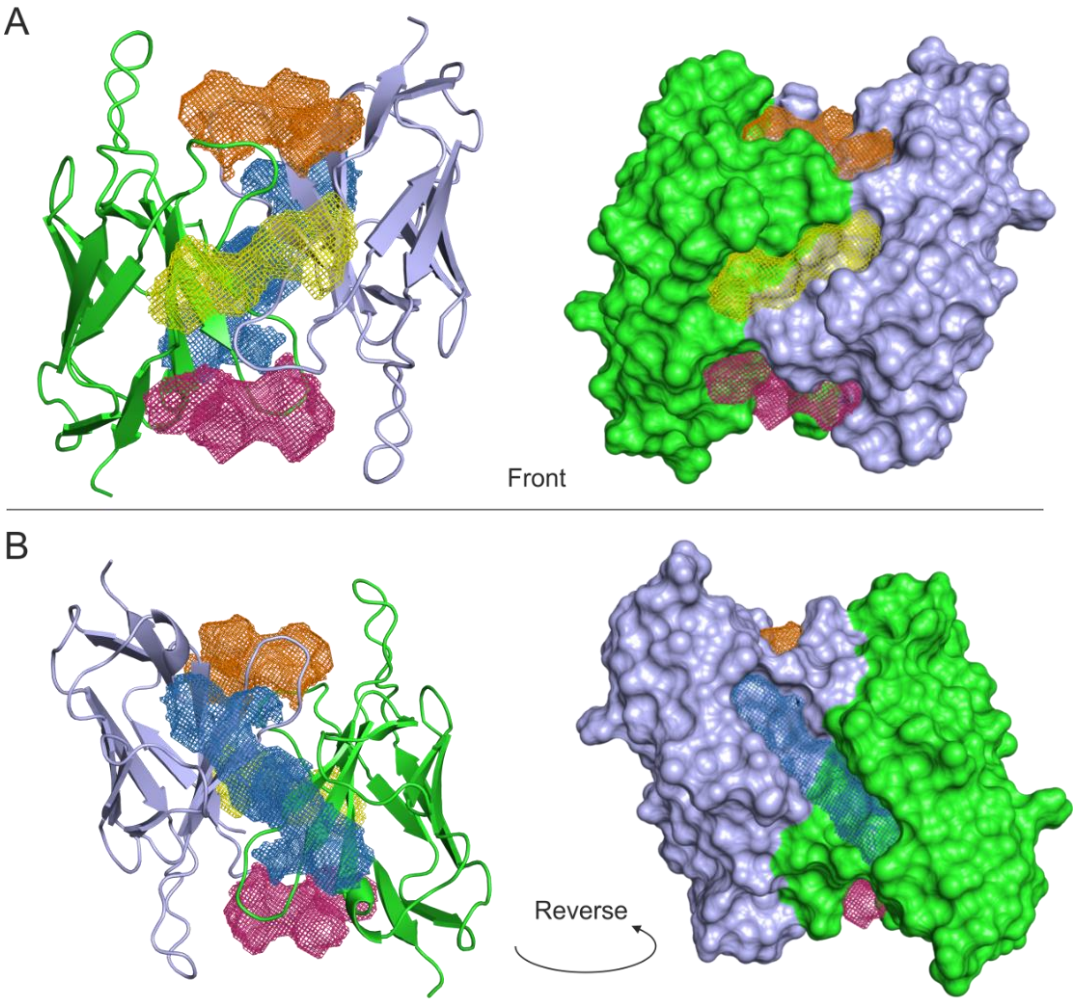


Figure 20 Calculated druggable sites for the non-canonical SMA dimer. Models of SMA generated using computational methods were used as query structure for the DogSiteScorer. Here, the server was able to predict a total of 4 sites. A) shows cartoon and space fil model of SMA (from front) with calculated pockets P0 (blue) P1 (yellow) P2 (orange) P3 (purple). B) shows structure rotated 180 ° on x-axis.

Overall, the data presented here suggests that the small molecule stabilisers (methylene and sulfasalazine) shown to bind to Mcg and prevent its aggregation would not be applicable for V_L-homodimers that possess a dimer interface rotated 180 ° with respect to the canonical dimer. Instead, several

new pockets have been identified. However, the large cavity suitable to accommodate such larger molecules like sulfasalazine is not present in SMA; this is likely to be the case for other V_L-homodimers that possess a non-canonical dimer interface. Inspection of the native crystal structure of REC reveals a cavity of 832.96 $Å^3$ (volume), 1135.83 $Å^3$ (surface), and a drug score of 0.79. Here, the software was able to successfully predict binding pockets that match well to the ligand bound crystal structures of Mcg. Thus, the results demonstrate the validity of performing druggablity assessment using structure-based computational tools in order to highlight protein-protein interaction targets that are suitable for molecular intervention.

2.3.9. Main findings and summary

Summary

Within this chapter, a three-dimensional structure of SMA was generated using the deposited crystal structures of LEN and REC. By assessing the energetic features of two putative dimer orientations we conclude that SMA is most favourable in the non-canonical dimer orientation. As there is currently no high resolution three-dimensional structure of SMA, its orientation has gone unknown until now. Structure-based pocket detection of this model has revealed a cavity that is suitable to accommodate a small-molecule.

Highlights

In this chapter, we have used a computational approach to obtain a model of SMA, and have established its orientation by assessing the energetic features of two putative dimer interfaces. This model of SMA then served as a query structure for DoGSite Scorer, an online server that assesses the druggability of

a given structure. The software was able to identify a cavity that spans both domains, suggesting the structure could possible accommodate a small-molecule that is capable of stabilising the homodimer and prevent its dissociation into aggregation prone monomers. A cavity with similar features was also identified in REC. Although each of the identified pockets (in SMA and REC) are somewhat shallower than the large hydrophobic cavity usually found within the canonical dimer, it scored well by DoGSiteScorer (drug score of 0.84 for SMA and 0.79 for REC), and bettered the drug score of Mcg (0.78), which has been previously shown to accommodate the small molecule sulfasalazine and methylene blue by crystallographic methods.

In addition, by using a number of digital platforms to assess, specifically, the changes in binding free energies between 2 monomers as a consequence of mutation, it has shed light on which of the SMA-like mutation(s) that may be responsible for disrupting key interactions. We document that 5 predictive software *mCSM-PP, ELASPIC, Mutabind*, and *BeAtMuSic as FoldX* were unable to consistently indicate residues to be destabilising/ neutral or stabilising however Q89H and Y96Q of the CDR3 region (also known as the Proline 95 loop) (Randles, Thompson et al. 2009) were able to significantly reduce the binding affinity and stability of the dimer based on the estimated free energy of dimer dissociation, the strength of hydrophobic interactions, and the binding energy which were each calculated by the PISA algorithm. This has not been shown previously, and could indicate that these mutations are able to increase the likelihood of dimer dissociation into the aggregation prone monomer.

We hope that the methods presented here may be applicable in defining both the orientation of V_L dimers, and elucidate the consequence of a particular mutation on the dimer interface in new and retrospective studies when a protein source is no longer available from a patient or is difficult to acquire by recombinant methods.

3. Assessing the effects of SMA-like mutations on V_L stability

3.1. Introduction

Destabilisation of a proteins native state is proposed to have a major impact on its ability to adopt pathological structural conformations that underlie a number of human disorders (Tokuriki, Stricher et al. 2008, Redler, Das et al. 2016, Kumar, Rahman et al. 2017). In AL amyloidosis, the accumulation of somatic mutations at key structural positions of the V_L renders the protein more likely to form amyloid. This can happen by two mechanisms. Firstly, mutations that occur within the dimer interface weaken the dimer and increase the chance of dissociation into aggregation prone monomers. In the monomeric state, mutations can decrease the folding stability allowing the protein to sample partially unfolded states that favour aggregation (Poshusta, Sikkink et al. 2009, Brumshtein, Esswein et al. 2014).

In chapter 2 we evaluated the ability of each SMA-like somatic mutation to alter the stability of the native LEN dimer using a computational-based strategy in the aim of identifying which residues(s) could potentially destabilise the dimer, an experiment that has not been conducted previously.

Unfortunately, there were inconsistencies between the 5 predictors used which ultimately prevent us from unequivocally stating that a particular amino acid substitution had a particular effect, and so, we built on these results by adopting a different *in silico*, structural based strategy which pinpointed two residues in particular (Q89H and Y96Q) that could potentially destabilise the dimer interface (2.3.7.).

In this chapter we turn our attention to the effect of these same 8 mutations on the folding stability (Gibbs free energy of folding ΔG_{fold}) of the LEN monomer (Li, Simonetti et al. 2016). In comparison to prediction of binding energies (which proved to be quite variable) predicting changes to the Gibbs free energy upon folding ($\Delta\Delta G$) is reportedly more accurately due to the higher number of available datasets from which programs can be developed and trained Acquiring such values using *in vitro* experimental methods is a timely process, requiring site-directed mutagenesis, followed by the expression and purification of the mutant structure before such values can be determined using thermal denaturation methods (Poshusta, Katoh et al. 2013, Miller, Lee et al. 2014, Blancas-Mejia, Hammernik et al. 2015).

Given the sheer number of possible mutations that can occur in variable domains, the time it takes to acquire experimental values – and the possibility that substitution of a single amino acid in a protein sequence can have devastating consequences on stability, preventing its expression by recombinant methods the ability to predict these values, or at the very least narrow down a large selection of substitutions to a few which can be assessed using experimental methods carries many benefits. In this chapter we use *FoldX*, *mCSM*, *SDM*, *DUET*, *ERIS* and *I-mutant 3.0, six* protein stability

predictors that reportedly high levels of accuracy (Guerois, Nielsen et al. 2002, Kumar, Rahman et al. 2017) to calculate the change in Gibbs free energy upon folding of the LEN monomer. These values were compared to experimental values. In addition, a structural link to the outcome of each mutation is demonstrated where we manually annotate, at atomic-level detail the conformational changes and including loss or gain of crucial intra and/or intermolecular interactions that occur upon mutation.

A portion of this chapter has been submitted for publication in collaboration with Dr. Sujoy Mukherjee (Indian Institute of Chemical Biology), Dr. Simon Pondaven, Dr. Christopher Janoreic (The Ohio State University) and Dr. Jillian Madine (University Of Liverpool) (Mukherjee, Pondaven et al. 2017)

3.2. Results

3.2.1. Assessing the effect of SMA-like somatic mutations on protein stability

Firstly, we used *FoldX*, *mCSM*, *SDM*, *DUET*, *ERIS* and *I-mutant 3.0* to calculate changes to the folding free energy (ΔΔG) of the LEN monomer upon a single point mutation. Mutations that are destabilising to the protein fold will contribute positively to the ΔG (ΔG >0 kcal/mol) where mutations that remove energy (ΔG <0 kcal/mol) will be stabilising SMA-like mutations that severely destabilise the LEN monomer may indicate which are more likely to form amyloid. The results of our analysis are shown in Table 8. For comparison, stability changes acquired by experimental methods (chemical induced equilibrium unfolding experiments and monitoring changes to spectra using

circular dichroism and tryptophan fluorescence) and the ability of a mutation to lead to fibril formation (ThT) previously performed by *Raffen et al* (Raffen, Dieckman et al. 1999) is also included.

Table 8 Structure-based energy calculations of SMA like mutations. *FoldX*, *mCSM*, *SDM*, *DUET*, *ERIS* and *I-mutant 3.0 was used to calculate* changes to the ΔG upon mutation. Software and their associated cutoff for a neutral mutation (indicated by +-) are as follows: *FoldX* +-0.46 kcal/mol; *mCSM*, SDM and DUET +-0.5 kcal/mol; *ERIS* (not reported) I-mutant 3.0 +- 0.5 kcal/mol. experimentally determined calculations of changes to folding stability are shown (displayed as $\Delta\Delta G_{unf}$). The ability of a mutation protein structure to forma amyloid for each mutation is also indicated (✘= does not form amyloid ✓ = forms amyloid). These experimental values are taken from the following publication (Raffen, Dieckman et al. 1999). Values are colour coded according to their ability to stabilise (green) or destabilise (red) the structure where blue is neutral.

	Computational Methods						Experimental Methods			
Mutation	*FoldX*	*mCSM*	*SDM*	*DUET*	*Eris* *Ερισ*	*I-mutant 3.0*	$\Delta\Delta G_{unf}$	Ability to form amyloid	Location	Solvent accessibility %
LEN	-	-	-	-	-	-	-	-	-	-
S29N	-0.55	-0.20	2.43	0.01	3.54	-0.56	-1.0	✘	CDR1	81.2%
K30R	0.30	-0.56	1.22	-0.23	-1.19	-0.69	0.1	✘	CDR1	46.0%
P40L	1.02	-0.27	4.89	0.08	-0.77	-0.44	0.7	✓	FR2	74.7%
Q89H	8.19	-1.39	-0.21	-1.48	8.34	-0.82	1.0	✘	CDR3	14.8%
T94H	0.27	-0.88	1.31	-0.74	1.69	-1.21	-0.7	✘	CDR3	72.8%
Y96Q	0.61	-0.57	-0.92	-0.41	0.96	-1.03	3.2	✓	CDR3	53.7%
S97T	0.77	-0.82	0.34	-0.78	-1.04	-0.66	-0.6	✘	CDR3	32.5%
I106L	0.83	-0.79	-0.78	-0.68	5.10	-0.51	-0.2	✘	FR3	14.6%
SMA	-	-	-	-	-	-	2.6	✓	-	-

We immediately focus on the accuracy of each program, where unfortunately we notice inconsistencies. One example can be found in the two most destabilising residues of this set; Q89H and Y96Q (by experimental methods of $\Delta\Delta G_{unf}$ = of +1.00 and +3.20 kcal/mol respectively) were consistently flagged as destabilizing mutations by the majority of the predictive software presented here (Table 8 colour coded). However *FoldX, mCSM, DUET, and ERIS,* calculated Y96Q to be only be mildly destabilising ($\Delta\Delta G_{fold}$ + 0.61, 0.57, 0.41, 0.96 kcal/mol respectively) yet this residue was the most destabilising of the set by *in vitro* measurements set ($\Delta\Delta G_{unf}$ +3.20 kcal/mol) (Raffen, Dieckman et al. 1999). In comparison, Q89H was shown to be much more destabilising using these platforms ($\Delta\Delta G_{fold}$ + 8.19, 1.391, 1.48, 8.34 kcal/mol respectively) yet, this mutation was only mildly destabilising *in vitro* (Raffen, Dieckman et al. 1999). The solvent accessibility between these two mutations (Q89H 14.8% solvent exposure and (Y96Q solvent exposed 53.7%) may be an explanation for this difference, where a mutation is usually more destabilising if it is located within the protein core (Tokuriki, Stricher et al. 2008).

For other residues that show inconsistencies either between themselves or between experimental values, the answer to why this occurs arguably lies within how the programs calculate the prediction where each software will have a particular bias to the database/ dataset it is trained on (Pandurangan, Ochoa-Montano et al. 2017).

Next, we place focus on the following three substitutions; P40L, Q89H and Y96Q due to their highly destabilising nature (both *in vitro* and *in silico*) and address a larger limitation of this study, which we overcome in subsequent analysis- specifically by combining our analysis with NMR . Interestingly, P40L was shown to be neutral /stabilising in all programs apart from *FoldX* (+1.02 kcal/mol) which corresponds somewhat to the stability values that were calculated ($\Delta\Delta G_{unf}$ -1.0 kcal/mol) experimentally. Q89H was shown to be destabilising by the majority of programs which corresponds to the experimental value of ($\Delta\Delta G_{unf}$ = 1.0 kcal/mol) and Y96Q was also shown to be destabilising ($\Delta\Delta G_{unf}$ = 3.2 kcal/mol).

Both Y96Q and P40L were able to enhance aggregation potential leading to fibril formation (Davis, Raffen et al. 2000, Joob and Wiwanitkit 2014) Q89H on the other hand, was unable to promote aggregation (negative for ThT fluorescence) despite the ability of this mutation to significantly reduce the folding stability of LEN (more than P40L) with experimental calculations of $\Delta\Delta G_{unf}$ of +1.00 kcal. The answer to this puzzling outcome circulates around an emerging concept that thermodynamic stability is not always an indication of aggregation propensity where it has been typically associated ((Marin-Argany, Guell-Bosch et al. 2015)).

It appears that P40L promotes fibril formation by a different mechanism, potentially altering structural integrity of a local region such as loop it is positioned, rather than having any global destabilising effects which could lead to unfolding of the monomer. Indeed, Davis *et al.,* has previously suggested that the substitution of the Proline at this position (a residue that is

conserved in 98% of kappa and lambda germlines) may disrupt the intramolecular interactions between residue Gln38 and intramolecular interactions between Lys 39 and Glu91, Arg61 and Asp 82 of neighbouring strands (Ramirez-Alvarado 2012, Joob and Wiwanitkit 2014). The precise structural changes for this mutation however, have not been experimentally tested to date and are later investigated in this thesis where we consider the residues position and the unique properties of proline (fixed torsion angles) and leucine (hydrophobicity) (Raffen, Dieckman et al. 1999). It is therefore important to take into account specific structural changes what assessing the consequence of a mutation and attempting to assess the ability of a mutation to increase aggregation cannot be taken on face value, rather a more sophisticated approach, (presented in the next section) is required to acquire a more firmer understanding of an outcome of a particular mutation.

3.2.2. SMA-like mutations of LEN induces structural changes

To provide a structural link to how certain SMA-like mutations confer a profound or negligible effect on the folding free energy ΔG_{fold} (presented in Table 8 or binding free energy ΔG_{bind} (presented last chapter Table 4), structural changes were elucidated at atomic-level. To provide a more in-depth analysis, and elucidate the dynamic behaviour of single mutant proteins in solution, which may overcome our limitation described previously, NMR relaxation exchange (R_{ex}) data, values which are indicative of motional fluctuations and intrinsic conformational flexibility over a milli-second (ms) timescale (Chapman, Davulcu et al. 2015) were matched with our own computational analysis of each mutant in the attempt to understand how intramolecular changes could enhance aggregation propensity. The work presented here, along with NMR R_{ex} data collected by Simon P. Pondaven at Ohio State University has been accepted for publication (Mukherjee, Pondaven et al. 2017). Where indicated, $\Delta\Delta G_{unf}$ values are taken from the original analysis of structures harbouring individual mutants (Raffen, Dieckman et al. 1999)

Mutants were generated *in silico*, using the LEN crystal structure as template with Rosetta based energy-minimisation. The lowest energy structure for each mutation, assessed by the Rosetta energy function was taken for further analysis. In this first section, SMA-like mutants K30R, P40L, Q89H, T94H or Y96Q, were chosen for molecular-level analysis to further investigate changes observed in ^{15}N-^{1}H heteronuclear single quantum coherence (HSQC) spectra or R_{ex} profiles (Mukherjee, Pondaven et al. 2017). For clarity, R_{ex}

profiles have been mapped onto the structure of LEN, SMA, and individual SMA-like mutants of LEN and is illustrated in Figure 21.

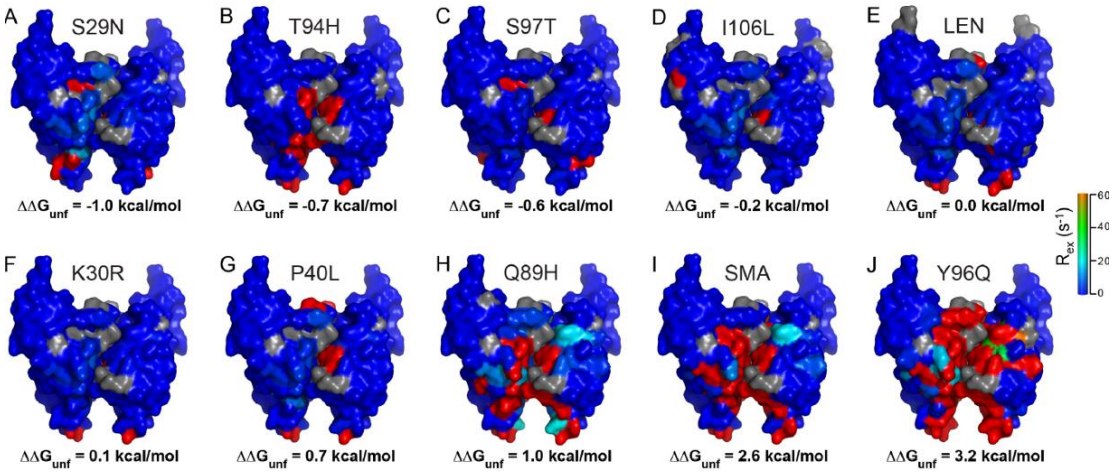


Figure 21 Surface representations of LEN, SMA and the eight SMA-like mutants of LEN in dimer form, depicting the locations of the most dynamic residues. The proteins are arranged in the order of decreasing thermodynamic stability with respect to LEN as reported by Raffen et al.,(Raffen, Dieckman et al. 1999) with LEN S29N (A) being the most stable and Y96Q (J) the most unstable. R_{ex} values were obtained at 600 MHz ^{1}H frequency is mapped onto the X-ray structure of LEN with colour coding. Text adopted from manuscript (Mukherjee, Pondaven et al. 2017).

As previously highlighted, P40L displays enhanced aggregation propensity relative to LEN (Raffen, Dieckman et al. 1999). It also displays decreased thermodynamic stability, with an experimentally derived $\Delta\Delta G_{unf}$ value of 0.7 kcal/mol (Table 8). Yet, it is not as destabilising as the Q89H mutation, ($\Delta\Delta G_{unf}$ value of 1.0 kcal/mol) which did not show the propensity to aggregate. To determine the molecular-level changes following this mutation, *in silico* analysis was used.

In silico analysis of mutant structure K30R indicate that the conformational changes induced by the conservative substitution of Lysine for Arginine at position 30 are quite subtle with overall root mean square (RMS)

value of 0.057 (Figure 22A). This is most likely due to swapping like for like of positively charged basic residues at the surface-exposed CDR1. This position is documented to tolerate substitution well without much impact on structure, as observed for a Serine to Asparagine mutation at the same position (S30N) in AL-09, an amyloidogenic member of the κI O18:O8 germline (Martin and Ramirez-Alvarado 2010). This particular mutation also shows no ability to cause any changes in binding energies of the dimer (Table 4), the thermodynamic stability (Table 8) and was unable to perturb the geometric or energetic features of the interface of models harbouring this mutation (Table 5). As seen in Figure 22A, the only detectable changes include minor alterations in hydrogen bond length between residues Tyrosine 25 and Serine 29 and the addition of a hydrogen bond of 2 Å between the long flexible chain of the Arginine where the residue forms a hydrogen bond with the adjacent backbone carbonyl of Serine 29. This corresponds to minimal alterations in dynamics of the ms timescale observed in NMR experiments (Figure 21) (Jorgensen, Chandrasekhar et al. 1983, Pondaven 2012).

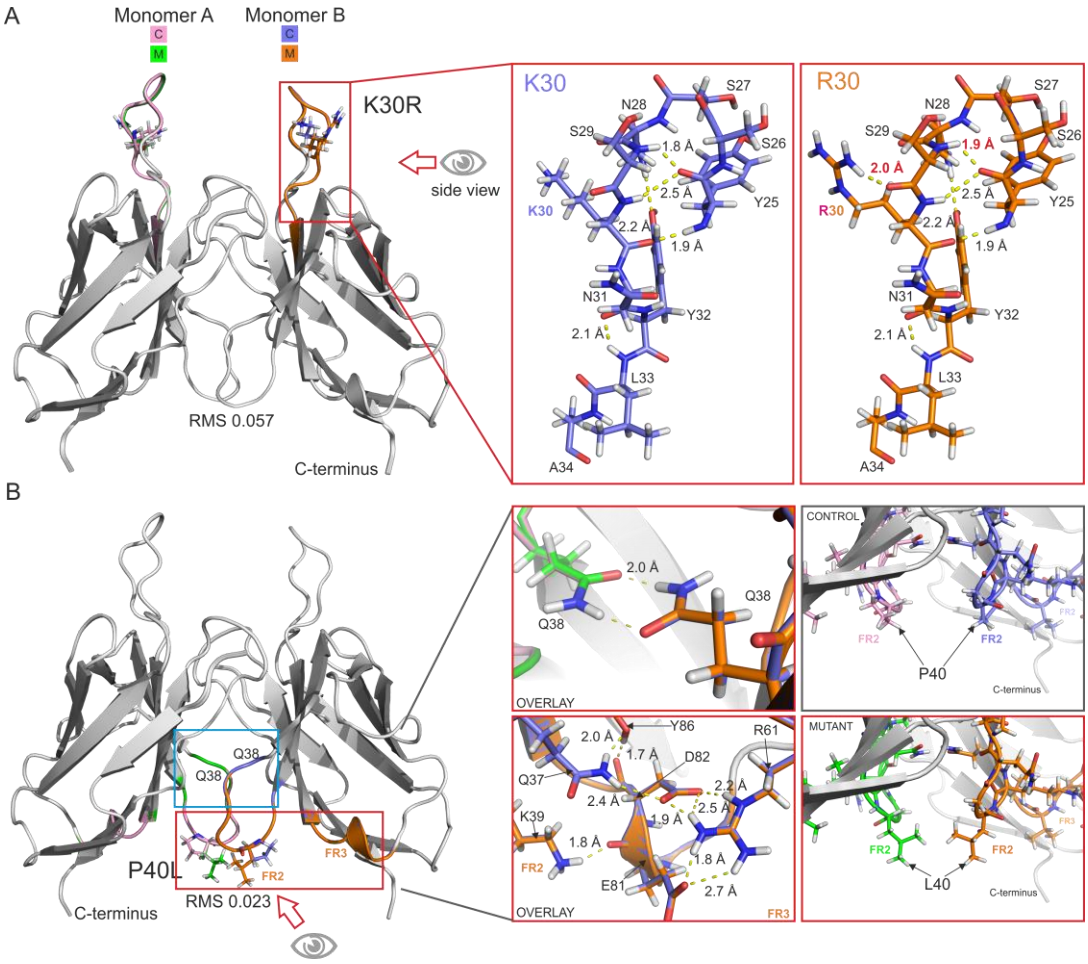


Figure 22 Computational analysis of mutation-induced structural changes of K30R and P40L. Dimer structure of native LEN is shown as cartoon overlaid with mutant K30R (A) and P40L (B) with mutated residue side-chains shown as sticks. RMS values between control LEN and mutant are indicated. Colour coded expanded regions highlight key areas of interest in greater detail. Control (C) monomers are shown in pink (monomer A) and blue (monomer B), with mutant (M) structures in green (monomer A) and orange (monomer B). Hydrogen bonds are shown as yellow dashed lines with lengths given, changes in lengths are shown in red text on mutant images. Framework regions (FR) 2 and 3 are shown to help to locate positions within the structure. P40 is predicted to be outside of the dimer interface before mutation.

In previous work, P40L has demonstrated a profound impact on the stability and aggregation of the protein, and was able to promote fibril formation under physiological conditions, without the need for addition of preformed misfolded protein templates or 'seeds' (Raffen, Dieckman et al. 1999, Davis, Raffen et al. 2000). The importance of this residue is also emphasised by noting that Proline 40 is a highly conserved residue present in 98% of all κ and λ germline sequences (Ramirez-Alvarado 2012). The R_{ex} profile (Figure 21) (Jorgensen, Chandrasekhar et al. 1983) is not consistent with the previously proposed suggestion that the P40L mutation increases backbone mobility over LEN in the FR2 loop. We use computational analysis to confirm a lack of increased dynamics in the FR2 loop, with a low RMS value of 0.023 Figure 22B consistent with the observed similar R_{ex} profiles between P40L and LEN (Pondaven 2012). In the previous study, it was also proposed that the mutation of Proline 40, a residue residing in a β turn within the FR2 loop between β-strands C and C' (CDR2 region) facilitates interactions between residues of neighbouring FR3 region, specifically by displacing interchain interactions between Glutamine 38 across the dimer interface, and interrupting the hydrogen bonding between Lys39-Glu81 and Arg61-Asp82 (Davis, Raffen et al. 2000, Baden, Randles et al. 2008, Poshusta, Sikkink et al. 2009). To investigate this possibility, we used computational analysis to predict that these inter and intrachain hydrogen bonds are not perturbed by this mutation (Figure 22B). To gain an insight into the dynamics of V_L LEN harboring the P40L mutation that extend further than that can be probed by R_{ex} (which was conducted at concentrations where the protein is a dimer and showed essentially no difference to the R_{ex} profile of wildtype LEN), CABS-FLEX

(Jamroz, Kolinski et al. 2013) was used to perform a 10 nanosecond MD simulation. A comparison between Cα-Cα fluctuations between LEN and the mutant structure (Figure 23) reveals that for the majority of the protein, motions of the backbone are near identical. There are however, 2 regions (CDR1 and CDR3) that exhibit enhanced fluctuations in the mutant.

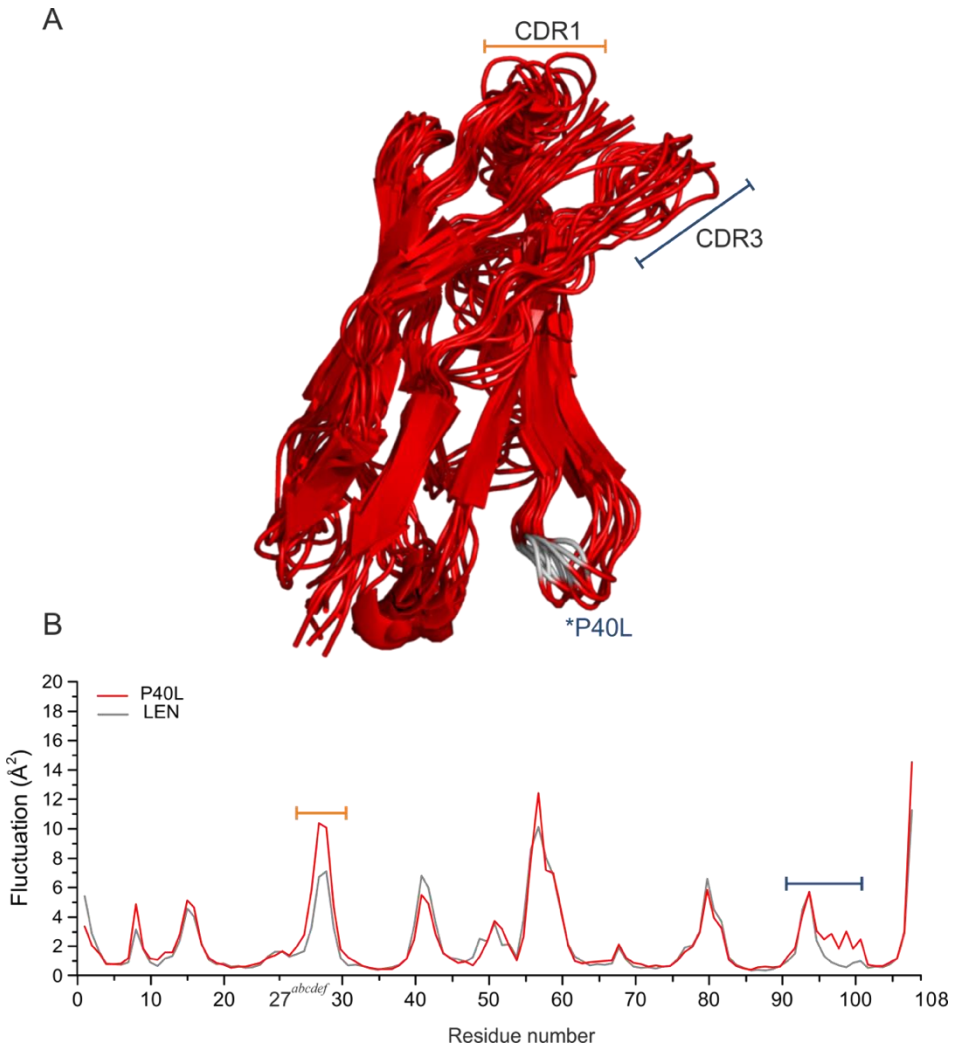


Figure 23 Residue fluctuation profile and structural ensemble of LEN containing REC-type mutant P40L calculated using CABS-FLEX. (A) A single leucine to proline mutation in the β-turn of the FR2 region affects loop dynamics at three distinct regions; CDR1 (orange) and CDR 2 (blue). B) The root mean square fluctuations (RMSF) of α-carbon atoms per residue plotted as a function of residue against fluctuation in Å acquired from a 10 nanosecond timescale.

The removal of a proline is likely to be accountable for the decreased thermodynamic stability, give than a proline is entropically stabilising, and limits the transition between one particular conformation to another. It is surprising to observe enhanced motions of the CDR1 and CDR2 region given the location of the mutation. Aside from the calculated decrease in stability (ΔG_{fold} 1.024 kcal/mol) calculated by *FoldX*, and enhanced motions in CDR regions another interesting feature of this mutation is in illustrated in where, it is the only mutation to be predicted to enhance aggregation potential by Aggrescan3D (Figure 24). This is consistent with the observation that this SMA-like mutation is the only one which can form fibrils without seeding (Davis, Raffen et al. 2000).

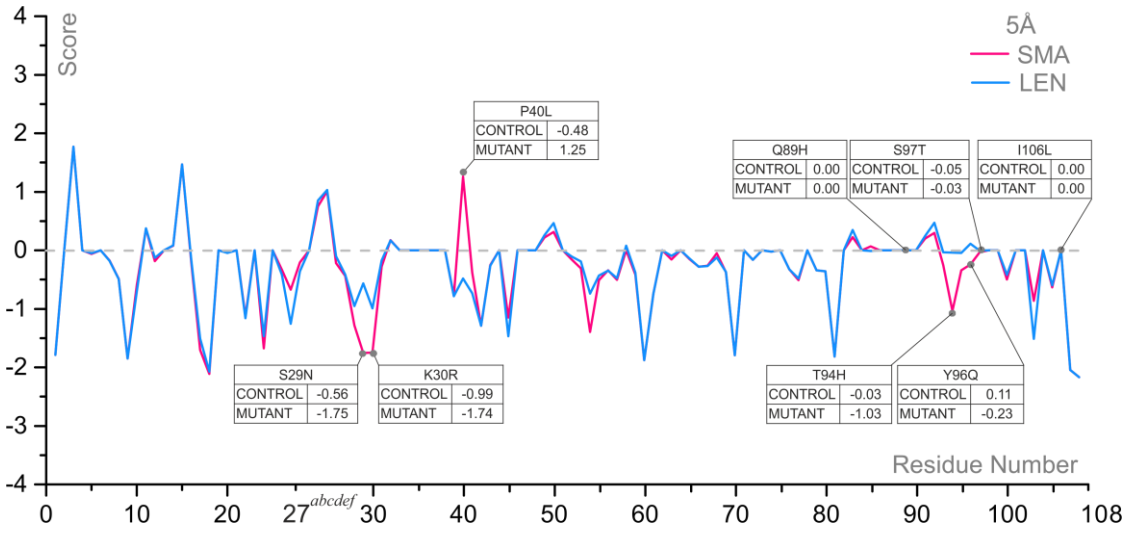


Figure 24 Aggrescan3D prediction profile. Aggrescan3D was used to assess the aggregation propensity of LEN (blue) and SMA (pink). The Y axis of the profile corresponds to the prediction scores. Positive scores indicate regions that calculated to be aggregation-prone. Only a single mutation (P40L) was calculated to increase aggregation potential. Analysis was performed using a 5 Å radius to pinpoint the individual residue contributions to overall aggregation potential.

In addition, analysis of dimer contacts using PISA revealed that residue 40 is the only residue out of all of the mutations found in SMA that is outside of the dimer interface prior to mutation, but becomes incorporated into the dimer interface following mutation (Figure 25). This observed change is presumably due to the inherent properties of Leucine, a highly hydrophobic residue (3.8 based on the Kyte and Doolittle hydrophilicity scale (Kyte and Doolittle 1982) which prefers to be buried within the interface rather than surface exposed.

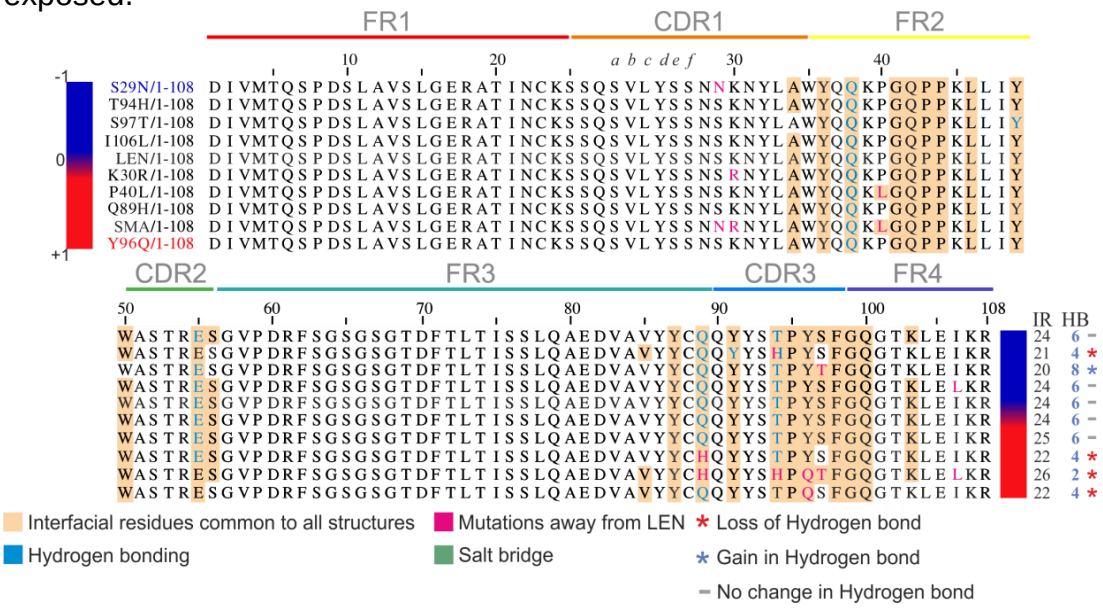


Figure 25 Protein sequence alignment of LEN, SMA and the eight SMA-like mutants of LEN. PISA was used to identify residues involved within the dimer interface of each structure within its 1 Å^2 cutoff. The alignment highlights residues common to both interfaces (sand). Residues involved in hydrogen bonding are highlighted in blue. White space indicates residues that were not involved in the dimer interface. Mutants highlighted in blue (S29N) and red (Y96Q) indicate the most and least stable (respectively) based on experimentally derived values from thermal unfolding experiments (Raffen, Dieckman et al. 1999). The number of interfacial residues (IR) and hydrogen interfacial hydrogen bonds (HB are indicated. In the case where N-terminal residue D1 was detected to be part of the interface, it was omitted from analysis, as performed previously by *Baden et al (Baden, Owen et al. 2008)*. Residue and region annotation were taken from *Stevens et al* (Stevens, Raffen et al. 1995) and are presented according to Kabat nomenclature (Kabat, Te Wu et al. 1992).

Overall, the removal of a proline at this position and its substitution to a leucine may be sufficient in allowing the protein to access the amyloidogenic landscape by a number of mechanisms described here (Padlan 1994, Graña-Montes, Pujols-Pujol et al. 2017).

Next, we analysed mutations that displayed larger differences in ^{15}N-^{1}H HSQC and CD spectra (Pondaven 2012), T94H, Y96Q and Q89H, suggesting conformational rearrangement within these mutants. Out of these mutants, Q89H and Y96Q display $\Delta\Delta G_{unf}$ stabilities of +1.0 kcal/mol and +3.2 kcal/mol relative to LEN, and both have been shown to be capable of forming fibrils upon the addition of pre-formed 'seeds' (Davis, Raffen et al. 2000). In contrast, T94H is stabilising (-0.7 kcal/mol $\Delta\Delta G_{unf}$), and shows no propensity to aggregate *in vitro* under non-destabilising conditions (Raffen, Dieckman et al. 1999, Davis, Raffen et al. 2000). This residue is also calculated to reduce the aggregation propensity of LEN (Figure 24). To investigate the differences between these mutants we performed an exhaustive analysis of inter and intrachain interactions.

For T94H, a number of changes in hydrogen bond length and a loss of hydrogen bonding between main and side chain atoms are observed, consistent with the largest RMS value of 0.5 out of all of the mutations studied (Figure 26A). The most profound alteration is the loss of a hydrogen bond between interfacial residues Threonine 94 and Glutamate 55 on opposite sides of the dimer (Figure 26, expanded region, top panel). However, despite this loss, the increase in stability seen in T94H mutant and in binding properties of a structure harbouring this mutation (Table 9) could be attributable to the creation of new intersubunit contacts between the resulting Histidine at

position 94, and Tyrosine 90 on the adjacent monomer (Figure 26A, expanded region, top panel and Figure 26B). T94H displayed one of the least changes in ms timescale dynamics (in comparison to LEN), with a quenched R_{ex} profile (Mukherjee, Pondaven et al. 2017), suggesting a less dynamic structure possibly due to restraints from the hydrogen bond formed between His94-Tyr90 across the dimer interface (Figure 26B).

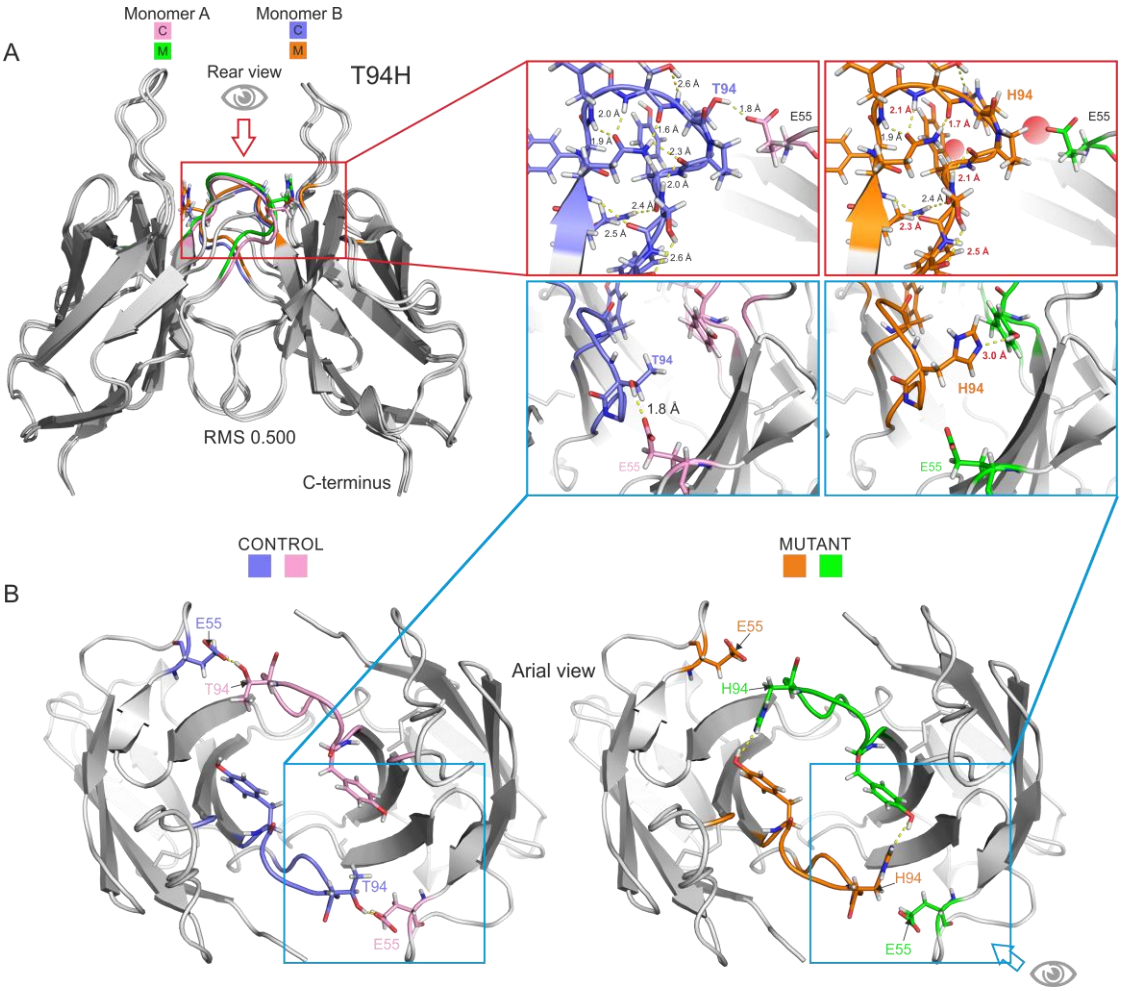


Figure 26 Computational analysis of mutation-induced structural changes of T94H. (A) Dimer structure of native LEN is shown as cartoon overlaid with mutant T94H with mutated residue side-chains shown as sticks. (B) Top view reveals hydrogen bond formed between His94 and Tyr 90 across the interface. RMS values between control LEN and mutant are indicated. Colour coded expanded regions highlight key areas of interest in greater detail. Direction of view for these regions is indicated. Control (C) monomers are shown in pink (monomer A) and blue (monomer B), with mutant (M) structures in green (monomer A) and orange (monomer B). Hydrogen bonds are shown as yellow dashed lines with lengths given; changes in lengths are shown in red text on mutant images. Red spheres

For Q89H, losses of multiple intra and interchain hydrogen bonds are observed (Figure 27). There is substantial alteration of most of the hydrogen bonds within the CDR3 loop (Figure 27, expanded region, top). Glutamine 89 in LEN forms a network of hydrogen bonds with neighbouring residues on the adjacent monomer that is lost upon mutation to Histidine (Figure 27, expanded region, top).

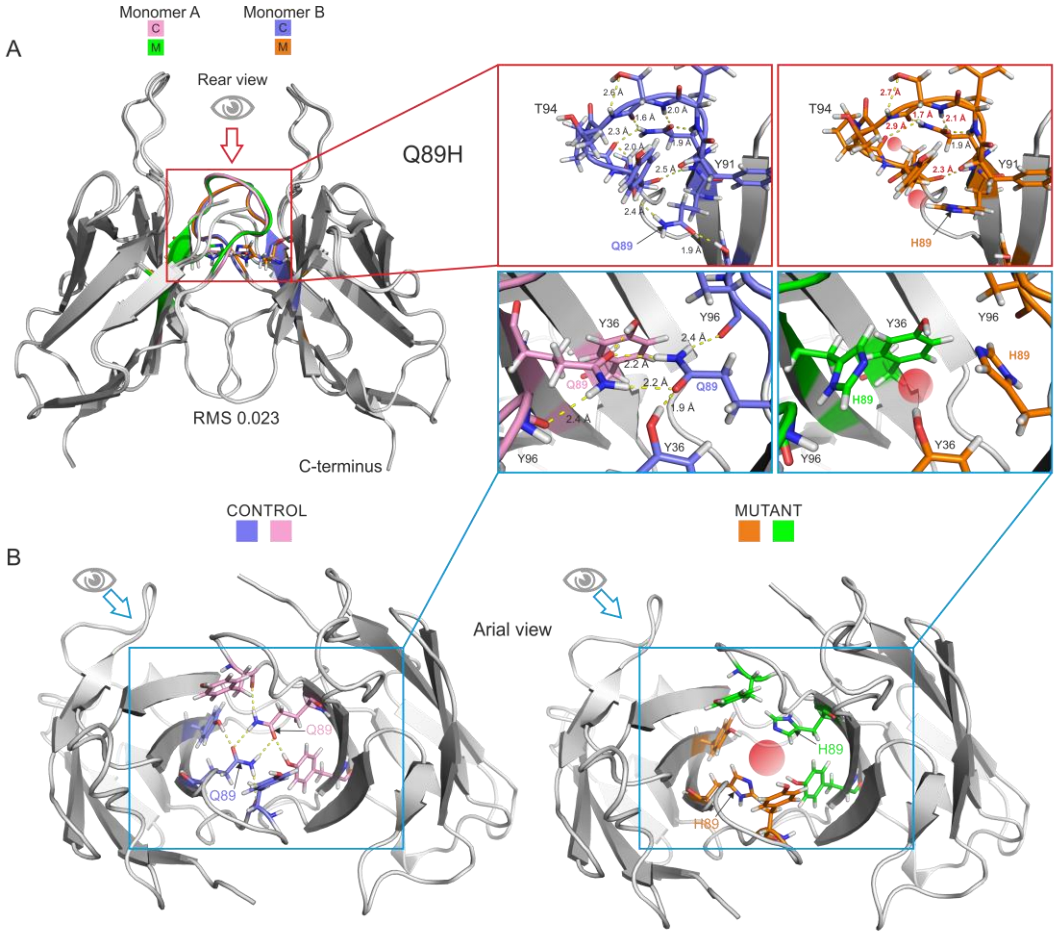


Figure 27 Computational analysis of mutation-induced structural changes of Q89H of the CDR3. Dimer structure of native LEN is shown as cartoon overlaid with mutant Q89H (A) with mutated residue side-chains shown as sticks. Top view (B) reveals dramatic loss of hydrogen bond between interfacial residues. RMS values between control LEN and mutant are indicated. Colour coded expanded regions highlight key areas of interest in greater detail. Direction of view for these regions is indicated. Control (C) monomers are shown in pink (monomer A) and blue (monomer B), with mutant (M) structures in green (monomer A) and orange (monomer B). Hydrogen bonds are shown as yellow dashed lines with lengths given; changes in lengths are shown in red text on mutant images. Red spheres indicate loss of

For Y96Q, residue-specific altered dynamics between LEN and the mutant structure extends further than the CDR3 loop containing the site of mutation (Figure 28A), giving a large RMS of 0.391, consistent with large changes in dynamics (Figure 21) (Mukherjee, Pondaven et al. 2017). Interaction between the CDR2 and 3 loops, through a hydrogen bond between Glutamic acid 55 - Threonine 94 are lost upon mutation (Figure 28A, expanded region, top and Figure 28B). This would result in both CDR2 and 3 loops having increased flexibility, explaining the excessive exchange broadening and loss of detection observed in these regions (Pondaven 2012). Further intramolecular changes are noted in several residues along the β-strands that form the core of the Ig, shown as coloured regions within Figure 28A. One of these changes is loss of another hydrogen bond between Serine 52 and Glycine 64 following mutation (Figure 28A, expanded region, bottom). There are also subtle changes in the hydrogen bond distance involving residue Y91 (Figure 28A expanded region), a residue that is part of the tyrosine cluster (DiCostanzo, Thompson et al. 2012).

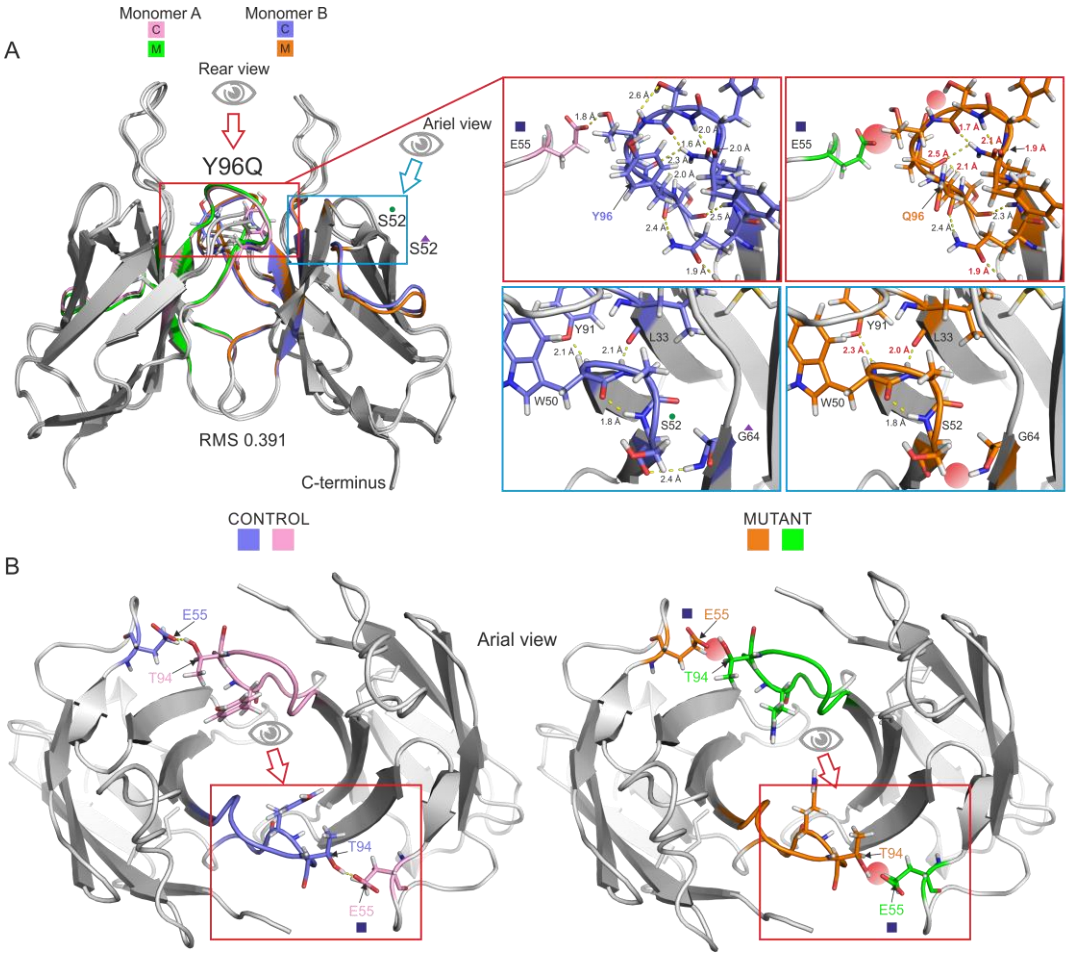


Figure 28 Computational analysis of mutation-induced structural changes of Y96Q. Dimer structure of native LEN is shown as cartoon overlaid with mutant Y96Q (A) with mutated residue side-chains shown as sticks. Top view (B) reveals dramatic loss of hydrogen bond between residues E55 and T94 as a result of the mutation. RMS values between control LEN and mutant are indicated. Colour coded expanded regions highlight key areas of interest in greater detail. Direction of view for these regions is indicated. Control (C) monomers are shown in pink (monomer A) and blue (monomer B), with mutant (M) structures in green (monomer A) and orange (monomer B). Hydrogen bonds are shown as yellow dashed lines with lengths given; changes in lengths are shown in red text on mutant images. Red spheres indicate loss of hydrogen bond.

The three mutations (T94H, Q89H and Y96Q) perturb a consistent set of residues surrounding the mutation within the CDR3 region, and interactions with surrounding loops. Y96Q loses all interactions between this loop and surrounding regions, through the loss of a hydrogen bond between Glutamate 55 and Threonine 94 of adjacent monomers, whereas in T94H a similar loss is observed between E55-T94, however Histidine 94 forms a new hydrogen bond with Tyrosine 90 on the adjacent monomer re-stabilising the CDR3 loop (Figure 26). Q89H also loses hydrogen bonds at the base of this loop (at the termini of FR3 and FR4, strands H and I respectively) (Figure 27), Taken together, disrupted hydrogen bonds appear to contribute to decreased thermodynamic stability (and ability to form fibrils under seeded conditions) for Q89H and Y96Q, whereas, in T94H the lost hydrogen bond is replaced to show increased stability and reduced amyloidogenicity. This suggests that the CDR3 loop, for this V_L domain is key to regulating thermodynamic stability and amyloidogenicity. It is also interesting to note that, while Q89H is highlighted to be part of an amyloidgenic prone region by a number of digital platforms (Amylpred2 and PASTA) (Figure 29) T94H was not predicted to be part of an aggregation prone region by Amylpred2 or AmyloidFISH. This is again consistent that this residue may prevent aggregation and stabilise both the dimer and the V_L monomer.

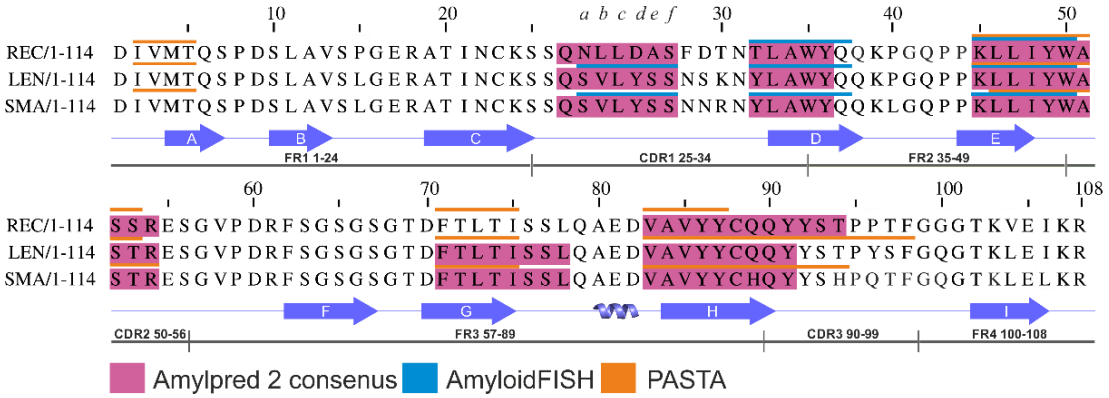


Figure 29 Identification of aggregation prone regions of benign LEN and amyloidogenic REC and SMA. A total of 11 different algorithms (under Amylpred 2) were used to predict aggregation prone regions using the primary sequence of SMA REC and LEN. Highlighted segments (Pink) show where different algorithms (see materials and methods AmylPred2) consistently highlighted regions of high aggregation propensity. AmyloidFISH (blue) and PASTA 2.0 (orange) were also used to highlight aggregation prone regions in addition. The location of CDR and FR are highlighted. Secondary structure content is also indicated.

3.2.3. Structural perturbations induced by the SMA-like mutants

For this thesis, we also wished to extend our analysis to cover the remaining three SMA-like mutants (S29N, S97T and I106L) that were not covered by such detailed structural analysis within the associated publication, but have NMR R_{ex} profiles recorded (Pondaven 2012). In a previous study each of these mutations S29N, S97T and I106L were shown to be stabilising ($\Delta\Delta G_{unf}$ -1.0 kcal/mol, -0.6 kcal/mol -0.2 kcal/mol respectively (Raffen, Dieckman et al. 1999). LEN harboring these single mutants yielded R_{ex} profiles that were almost indistinguishable from control structure (LEN). High resolution analysis of a structure harboring the S29N mutation (Figure 30) revealed that the solvent exposed residue (81.2% determined by mCSM), located within the CDR1 region shows no major perturbations with a minimal backbone

perturbation (RMS 0.103), which corresponds to the lack of any significant changes reflected in the NMR R_{ex} data (Figure 21) which was almost indistinguishable from LEN. No large differences were noted for a structure harbouring I106L (Figure 30).

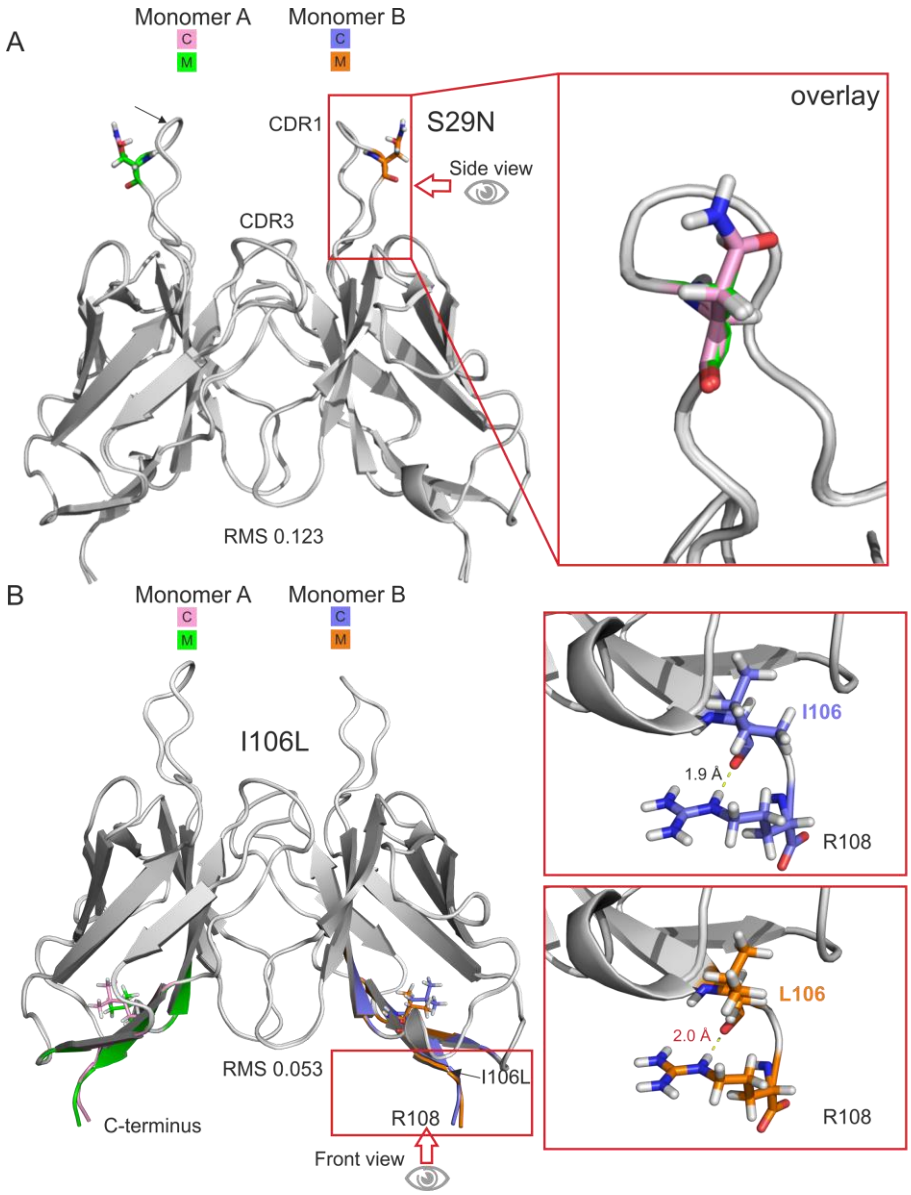


Figure 30 Computational analysis of mutation-induced structural changes of S29N and I106L Dimer structure of native LEN is shown as cartoon overlaid with mutant (S29N and I106L) with mutated residue side-chains shown as sticks RMS values between control LEN and mutant are indicated. Colour coded expanded regions highlight key areas of interest in greater detail. The direction of view for these regions is indicated. Control (C) monomers are shown in pink (monomer A) and blue (monomer B), with mutant (M) structures in green (monomer A) and orange (monomer B).

For LEN harbouring S97T, a number of interesting observations were made. Small changes in hydrogen bond length are observed at the site of mutation (Figure 31A), and a loss of interaction is noted between residues Methionine 4 and Threonine 97, occurring directly at the site of mutation. There are multiple changes to hydrogen length at the site of each mutation but the most profound change is observed between residue Y92 (this is part of the important Tyrosine cluster of the CDR3 region) where subtle changes in the CDR3 region allow for this residue to adopt new intramolecular hydrogen bonds of (1.8 Å) between Y96 and Y49 of the opposite monomer.

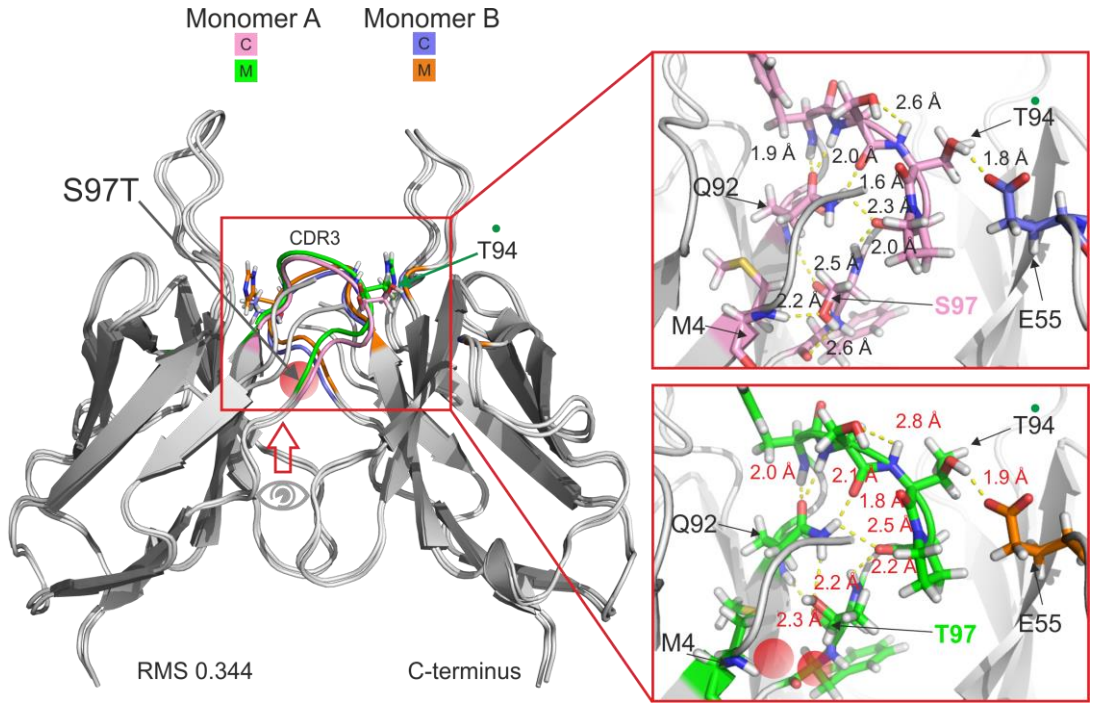


Figure 31 Computational analysis of mutation-induced structural changes of S97T and I106L. Dimer structure of native LEN is shown as cartoon overlaid with mutant (S97T or I106L) with mutated residue side-chains shown as sticks. RMS values between control LEN and mutant are indicated. Colour coded expanded regions highlight key areas of interest in greater detail. The direction of view for these regions is indicated. Control (C) monomers are shown in pink (monomer A) and blue (monomer B), with mutant (M) structures in green (monomer A) and orange (monomer B). Hydrogen bonds are shown as yellow dashed lines with lengths given; changes in lengths are shown in red text on mutant images. Red spheres indicate loss of hydrogen bond.

3.2.4. Chapter summary

Summary

In this chapter, we have used a computational approach to assess the contribution of an individual somatic mutation to the folding stability of the non-amyloidogenic V_L domain LEN. Inconsistencies between the ability of different algorithms used to calculate the destabilising and stabilising nature of a mutation (Table 8 colour coded) means that these programs have failed to provide a reliable benchmark from which effect of a mutation on the stability of LEN can be assessed. Instead, a more sophisticated structural approach was needed. An exhaustive analysis of both binding (presented last chapter) folding stability (experimental values this chapter) and structural analysis of each mutation performed in combination with previously (unpublished) NMR data has revealed that the CDR3 loop, for this V_L domain is key to regulating the stability and amyloidogenicity.

Highlights

NMR data reveals that out of all 8 SMA-like mutations of LEN only Q89H, T94H and Y96Q are responsible for enhanced main chain dynamics over the control structure (LEN). Sequence-specific NMR relaxation exchange data (mapped onto the LEN structure Figure 21) indicate that these enhanced dynamics were largely restricted to the CDR3. The dynamics of structures harbouring the remaining mutants (S29N, K30R, P40L, S97T and I106L) were not substantially altered in comparison to the control structure (LEN). Computational structure-based assessment of these three mutations (Q89H, T94H and Y96Q) indicate that each are able disrupt the hydrogen bonding

network within the CDR3 loop and surrounding regions which provides an explanation to the increased motions observed in NMR experiments. While Q89H and Y96Q have also shown to be destabilising by CD thermal denaturation experiments, the T94H mutation has previously shown to be stabilising. Further computational analysis of this mutant revealed that new intermolecular hydrogen bonds are formed which are likely to compensate for the disruption of intramolecular hydrogen bonds directly at the site of mutation. In this analysis, disrupted hydrogen bonds appear to contribute to enhanced conformational dynamics and decreased thermodynamic stability (and ability to form fibrils under seeded conditions) for Q89H and Y96Q, suggesting that these residues are key for regulating the thermodynamic stability and amyloidogenicity of LEN.

4. Predicting the stability effects of REC-like mutations and modelling the structural changes

4.1. Introduction

In chapters 2 and 3 the ability of an individual SMA-like mutation to alter the structure and stability of LEN was characterised by calculating both, changes in the binding and folding free energies (ΔG_{bind} and ΔG_{fold}) as well as evaluating the consequence of a mutation at a structural level. For the most part, were unable to indicate which program was the most accurate due to various errors of different computational methods. However, we found the most consistency in mutations Q89H and Y96Q which correlated with NMR data. Seeking to extend the investigation and trial a new, larger much data set where the results could be validated by experimental means at a later opportunity this chapter assesses the contribution of individual REC-like somatic mutations (what has not been entirely covered to date by experimental methods) to the stability and structure of the non-amyloidlogenic V_L LEN dimer and monomer.

4.2. Results

4.2.1. Predicting changes in binding energy as a result of a REC-like somatic mutation

The goal of this section is to examine the ability of each individual REC-like mutation to weaken the binding affinity and destabilise the LEN homodimer.

Changes to the binding free energy ($\Delta\Delta G_{bind}$) induced by a single mutation were calculated by *FoldX*, *BeAtMusic*, *Elaspic*, *Mutabind* and *mCSM-PPI.*

Table 10 The effects of 14 REC-like mutations on the binding stability of the homodimer LEN. Residues located at the interface are highlighted (✓). Units for energy calculations ($\Delta\Delta G_{bind}$) are expressed in kcal/mol. FoldX supports the ability of simultaneous mutations in each side of the homodimer and was performed this way. Software and their associated cutoff for a neutral mutation (indicated by +-) are as follows: FoldX +-0.46 kcal/mol; ELASPIC +-0.46 kcal/mol; MutaBind +-0.86 kcal/mol; mCSM-PPI +-0.5 kcal/mol; BeAtMuSiC (error values not reported). Values are colour coded according to their ability to stabilise (green) or destabilise (red) the structure. Neutral mutations are shown in blue.

	FoldX	mCSM: PPI	ELASPIC	MutaBind	BeAtMuSiC	Location	Interface?
L15P	0.00	0.10	0.35	1.63	-0.37	FR1	✘
S27aN	0.00	0.54	0.54	0.68	0.21	CDR1	✘
V27bL	0.00	0.96	-0.35	1.05	-0.38	CDR1	✘
Y27dD	0.00	0.98	0.25	0.81	-0.35	CDR1	✘
S27eA	0.00	0.00	0.14	0.48	0.31	CDR1	✘
N28F	0.00	0.98	0.90	0.87	-0.01	CDR1	✘
S29D	0.00	0.04	-0.07	0.3	-0.04	CDR1	✘
K30T	0.07	1.24	1.33	0.96	0.07	CDR1	✘
Y32T	2.52	0.07	1.14	1.75	0.68	CDR1	✘
T53S	-0.04	0.32	0.33	0.76	-0.01	CDR2	✘
Y96P	0.59	1.98	1.32	3.0	4.99	CDR3	✓
S97T	0.07	0.67	0.30	0.61	-0.56	CDR3	✓
Q100G	-0.06	0.00	-0.69	1.23	0.33	FR4	✓
L104V	0.00	0.71	0.67	1.39	-0.32	FR4	✘

Examining the changes in binding energies ($\Delta\Delta G_{bind}$) calculated by *FoldX* (

Table **10**) indicate that the majority of solvent accessible residues are unable to destabilise the LEN homodimer (L15P, S27aN, V27bL, Y27dD, S27eA N28f, S29D, K30T, T53S, Q100G and L104V). While, *FoldX and B*e*AtMuSiC* calculated these mutants to be largely neutral, *mCSM*, *ELASPIC* and *Mutabind* indicated some of these mutants to be quite destabilising. For *Mutabind*, all but three (L15P, K32T, Y96P) of the mutations calculated to be destabilising (V27bL, N28F, K30T, Q100G, and L104V) were scored with low confidence ($\Delta\Delta G_{bind}$ greater than 0.86 but below 1.57 kcal/mol).

Interestingly, the most consistently destabilising mutants (Y96P and Y32T (highlighted by the majority of programs) have been previously highlighted as key residues for modulating stability in light chain variable domains. Through the analysis of 17 κI Bence Jones proteins, Fred Stevens and colleagues have previously demonstrated the important contribution of the hypervariable CDR3 region to the formation of V_L-V_L homodimers (Fred J. Stevens 1980). Specifically, an aromatic, or a hydrophobic residue located at position 96 is essential for maintaining self-association. Stevens notes that, while kappa type variable domains AU and REI form Bence Jones dimers contain a Tryptophan and Tyrosine located at position 96 respectively. In contrast, AG and GAL possess a arginine (basic) at this position (Goni and Frangione 1983), and are found almost exclusively as monomeric species, suggesting that removal of the aromatic residues is able to abolish the dimerisation capability of these two V_L completely (Fehlhammer, Schiffer et al. 1975). While the substitution of a Tyrosine (aromatic) for a proline (hydrophobic) is calculated to be slightly destabilising by *FoldX* ($\Delta\Delta G_{bind}$ +0.46 kcal/mol < $\Delta\Delta G \leq$ +0.92 kcal/mol where the calculated value is 0.59 kcal/mol)

the substitution of this residue with the SMA mutation Y96Q is far more destabilising ($\Delta\Delta G_{bind}$ +1.27 kcal/mol) which agrees with the observations made by Stevens (Fred J. Stevens 1980).

Y32T in particular is shown to substantially reduce binding affinity (*FoldX* $\Delta\Delta G_{bind}$ +2.52 kcal/mol) which may suggest this mutation can destabilise the canonical LEN dimer. The observed decrease in stability is also reflected in *ELASPIC*, *MutaBind* and *BeAtMuSIC*, ($\Delta\Delta G_{bind}$ +1.14, 1.75, and 0.68 kcal/mol respectively). This gives confidence to the findings; however, *mCSM-PPI* was unable to calculate the destabilising nature of this mutation ($\Delta\Delta G_{bind}$ 0.07 kcal/mol neutral) which is likely a result of this program to not take into account symmetry related double mutants (only able to mutate one subunit at a time). To our knowledge, the ability of Y32T to perturb any intermolecular contacts has not been documented. This mutation does however occur in AL-09 and AL103 (Figure 32) and is highlighted to be part of a highly conserved cluster of tyrosine residues that comprise Y32 of the CDR1 region and Y91 and Y96 of the CDR3 region that is termed the tyrosine corner (Hemmingsen, Gernert et al. 1994).

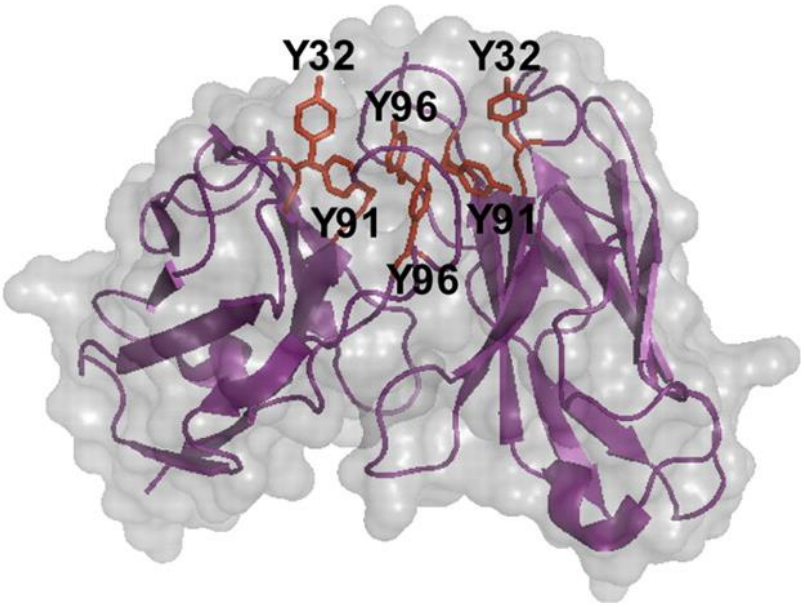


Figure 32 Location of key tyrosine residues in V_L AL-103. Location of Tyrosine 32, 91 and 96 are shown (red) as side chains in each protein chain of the AL-103 (purple) dimer structure. While Y32 is located away from the interface, Y96 occupies a key position within the dimer interface. Figure adopted from (DiCostanzo, Thompson et al. 2012)

Here, tyrosine residues that are situated at opposite ends of antiparallel β-strands participate in stabilising interactions (DiCostanzo, Thompson et al. 2012). While Y32T is not directly located within the dimer interface (Figure 32), changes to this residue may propagate between neighbouring Tyrosine 91 and 96 and cause perturbation to these residues that are part of the interface region. Similar long-range effects that are able to alter interfaces are noted for superoxide dismutase 1 (SOD1) and TTR (Johnson, Connelly et al. 2012, Broom, Rumfeldt et al. 2015). This is presumably the underlying reason to how such a destabilising effect was calculated for this mutation by *FoldX, ELASPIC, MutaBind* and *BeAtMuSiC.* Indeed high-resolution analysis of this

mutation presented later on in this chapter reveals changes involving these residues (Figure 43).

4.2.2. The effect of REC-like mutations on the dimer interface

Having highlighted the ability of a number of REC-like mutations to reduce the binding affinity of the LEN dimer, specifically Tyrosine residues 32, and Y96, we next wished to identify which mutations could possibly dictate the orientation of LEN (control dimer). To assess this, the differences in the geometrical and energetic properties of both canonical and non-canonical LEN dimers harbouring a single REC like mutation were compared (Table 11). As a control, we also include a structure of LEN that is posed into a 180 ^{o} non-canonical arrangement that is reputed to be non-physiological, as this structure was found to form a canonical dimer upon crystallisation in multiple studies (Stevens, Raffen et al. 1995, Huang, Chang et al. 1997, Pokkuluri, Huang et al. 1998).

Table 11 Interfacial analysis of LEN harbouring a single REC-like mutation. A single REC-like mutation was incorporated into the dimer of LEN (PDB ID: 1LVE). Structures were posed into two orientations (canonical ⇈ and non-canonical ⇅ with arrows indicating the direction of the monomer) and the energetic features of each interface calculated by PISA. Descriptions and their abbreviations are presented earlier in Table 3 and are taken from PISA documentation. Residues located at the interface are highlighted (✓).

A: INDIVIDUAL RESIDUE CONTRIBUTIONS															
	LEN	L15P	S27aN	V27bL	Y27dD	S27eA	N28F	S29D	K30T	Y32T	T53S	Y96P	S97T	Q100G	L104V
Ori	X-ray crystal determined canonical dimer A⇈B														
IA	747.2	746.5	753.5	687.1	721.2	730.5	720.9	692.2	725.4	729.0	745.6	653.7	723.7	729.2	727.9
iNres	24	24	22	21	22	22	21	19	21	22	24	23	20	24	23
BE	-13.6	-13.6	-12.0	-10.0	-13.2	-13.8	-12.9	-10.7	-13.0	-13.0	-13.5	-13.0	-13.6	-13.1	-10.6
ΔiG	-10.9	-10.9	-13.8	-13.3	-11.4	-11.1	-10.3	-8.3	-10.3	-10.4	-11.7	-10.4	-10.0	-10.4	-13.3
ΔG^{dis}	2.1	2.1	2.1	2.1	1.8	1.8	2.3	1.5	-0.8	1.6	1.6	1.1	2.1	1.6	1.8
HB	6	6	4	4	4	6	6	2	6	6	4	6	8	6	6
SB	0	0	0	4	0	0	0	4	0	0	0	0	0	0	0

B: INDIVIDUAL RESIDUE CONTRIBUTIONS															
	LEN	L15P	S27aN	V27bL	Y27dD	S27eA	N28F	S29D	K30T	Y32T	T53S	Y96P	S97T	Q100G	L104V
Ori	Non-canonical "flipped" dimer A⇅B														
IA	635.6	717.0	747.7	753.9	738.5	698.5	734.3	709.8	808.9	840.3	692.1	612.8	802.5	757.5	833.0
iNres	20	22	22	20	24	19	21	20	22	23	19	18	23	22	22
BE	-11.2	-12.9	-13.4	-13.0	-13.8	-12.7	-13.8	-13.3	-12.4	-13.4	-11.4	-10.1	-12.0	-11.6	-13.6
ΔiG	-10.3	-10.2	-10.8	-12.0	-11.2	-10.9	-12.9	-10.6	-10.1	-10.0	-8.1	-7.7	-9.7	-9.2	-10.4
ΔG^{diss}	-0.2	1.4	2.0	2.3	2.4	1.6	2.3	1.9	1.0	2.0	-0.1	-1.3	-0.6	0.1	2.1
HB	2	6	6	4	6	4	2	6	2	6	4	2	2	2	4
SB	0	0	0	0	0	0	0	0	4	2	4	4	4	4	4

A previous comparison between LEN dimers harbouring individual SMA-like mutations revealed that only a fraction was able to destabilise the canonical dimer of LEN (Table 5). These mutations (namely Q89H and Y96Q) were also restricted to the CDR3 region of the protein. In contrast, PISA identified that almost all of the REC-like mutations shown here (Table 11) have destabilising effects that extend further than merely the binding energy shown in

Table **10**, and are not confined to a specific CDR or FR. Interestingly, substantial differences emerge even between mutations that occur at identical positions in SMA which suggests it is not only the location but also the intrinsic properties of the mutation. For example, the canonical LEN dimer harbouring SMA-like mutation K30R had no effect on any property of the dimer interface (Table 5) and S29N had only marginally destabilising effects (IA of 742.3 $Å^2$ in comparison to 747.2 $Å^2$ of the control structure (Table 5). REC-like mutants S29D and K30T however, are able to decrease the Interfacial area, the number of interfacial residues, the binding energy and the free energy of dissociation in comparison to control structure (Table 11 A). Whilst the K30T LEN canonical model (Table 11 A) has only marginal differences in a number of energetic and geometrical properties in comparison to the control structure, it is characterised by a significantly lower free energy of dissociation (ΔG^{diss} -0.8 kcal/mol in comparison to 2.1 kcal/mol), where a negative value is indicative of an unstable structure (Krissinel and Henrick 2007, Tomovic and Oakeley 2008, Gopavajhula, Chaitanya et al. 2013, Ravn, Madhurantakam et al. 2013). This mutation is however, able to stabilise the non-canonical LEN structure (Table 11 B) which is in agreement with experimental data, where LEN harbouring the K30T mutant crystallised as a non-canonical dimer (PDB

ID: 4LVE (Pokkuluri, Huang et al. 1998). Unexpectedly, even residues situated outside of the canonical-dimer interface (assessed by PISA) (highlighted (✗) in Table 11 A) are able to alter a number of energetic and geometrical properties and inter-domain interactions (which was also reflected in some of the values presented in

Table **10**, yet were not identified by *FoldX*). This may suggest that these mutations may be able to induce downstream effects that lead to alterations within the interface, a feature that is probed further by structural analysis later.

As previously mentioned, an aromatic or a hydrophobic residue at position 96 is key in maintaining dimerisation ability (Fred J. Stevens 1980). In SMA, the tyrosine at this position is mutated to a glutamine. As calculated by *FoldX* and a number of other platforms, this mutation was particularly destabilising to the dimer interface (Table 4). Generating a structure harbouring this mutation also revealed its ability to destabilise not only the canonical LEN interface, but also the altered interface, where negative ΔG^{diss} values (-0.2 kcal/mol) indicate that this was an unstable association (Table 5). (Krissinel and Henrick 2007, Tomovic and Oakeley 2008, Gopavajhula, Chaitanya et al. 2013, Ravn, Madhurantakam et al. 2013). Here, the REC-like mutation Y96P is also able to destabilise the canonical dimer (estimated free energy of dissociation), but this is to a lesser extent than Y96Q. Unlike Y96Q, this residue is unlikely to dictate the orientation of the dimer as it is also not stable in the non-canonical structure of LEN. It is worth noting that overall most of the REC-like mutations actually disfavour the canonical structure and favour the non-canonical form, where REC crystallised in non-canonical state.

4.2.3. Changes in V_L stability as a consequence of REC-like mutations

A recent publication by Priyabrata Panigrahi *et al.,* employs the use of iRDP, (a web server that houses iMutants, iCAPS, and iStability) to elucidate stabilising and destabilising mutations by assessing their individual contributions to protein thermodynamic stability. Pertinent to investigation is their attempt to match the highly destabilising nature of L15P, which was previously shown to be destabilising ($\Delta\Delta G_{unf}$ +1.7 kcal/mol) by chemical induced equilibrium unfolding methods, and also tested positive for fibril formation (Table 12) (Raffen, Dieckman et al. 1999). Unfortunately, the digital platforms used within this study were not able to accurately predict the highly destabilisation nature of this mutation, which actually calculated the mutation to be stabilising (Panigrahi, Sule et al. 2015). We have experienced inconsistencies within other chapters, we combine these results along with current literature and structural information (later document in this thesis) to reach a more educated conclusion into the destabilising, neutral or stabilising nature of each mutation rather than taking the predicted values on face value alone. The ability of the remaining six REC-like mutations (S27aN, S27eA, S29D, Y32T, T53S, Q100G) to alter the thermodynamic stability of LEN were not investigated in the original study performed by Raffen *et al* (Raffen, Dieckman et al. 1999), and to our knowledge, have not been assessed by either experimental or computational methods to date. Here, we assessed the ability of a particular mutation to alter the folding stability (ΔG_{fold}) of the reference structure (LEN monomer) using *FoldX* and a number of other complementary digital platforms.

Table 12 Structure-based energy calculations of REC like mutations.

Monomeric LEN (PDB:ID 1LVE) was subject to the Rosetta based energy minimisation procedure and mutated using *FoldX, SDM, DUET, Eris and I-mutant 3.0.* Software was used to calculate the changes in free energy of unfolding ($\Delta\Delta G_{fold}$) and is expressed as kcal/mol. Software and their associated cutoff for a neutral mutation (indicated by +-) are as follows: *FoldX* +-0.46 kcal/mol; *mCSM*, SDM and DUET +-0.5 kcal/mol; *Eris* (not reported) I-mutant 3.0 +- 0.5 kcal/mol. Calculations of thermodynamic stability (displayed as $\Delta\Delta G_{unf}$ made from experimental methods and the ability of each mutant structure to form amyloid fibrils is experimental data extracted from the following publication (Raffen, Dieckman et al. 1999). Values are colour coded according to their ability to stabilise (green) or destabilise (red) the structure. Neutral mutations are highlighted in blue. ND = not determined.

Mutation	*FoldX*	mCSM-stability	SDM	DUET	*Eris*	I-mutant 3.0	CD - $\Delta\Delta G_{unf}$	Ability to form amyloid	*Location*
LEN	-	-	-	-		-	-	✘	-
L15P	0.82	0.59	-0.80	-0.53	6.91	-1.34	1.7	✓	FR1
S27aN	0.64	-0.11	-0.08	0.09	5.05	-1.12	ND	ND	CDR1
V27bL	0.88	-0.33	0.73	-0.15	-6.93	-1.12	-0.7	✘	CDR1
Y27dD	-0.03	-0.37	-2.9	-0.34	-0.86	-1.04	-2.7	✘	CDR1
S27eA	-0.07	-0.48	0.16	-0.35	-1.07	-0.56	ND	ND	CDR1
N28F	-0.98	1.17	0.62	-1.30	0.86	-0.17	1.7	✓	CDR1
S29D	-0.25	-0.24	0.62	0.49	-2.34	-0.21	ND	ND	CDR1
K30T	1.31	1.32	-0.40	-1.14	2.28	-0.73	1.4	✓	CDR1
Y32T	6.77	2.29	-1.07	-2.33	9.35	-1.38	ND	ND	CDR1
T53S	0.69	1.10	-0.97	-0.99	3.12	-0.45	ND	ND	CDR2
Y96P	0.10	1.18	-3.80	-1.26	6.11	-0.78	1.2	✓	CDR3
S97T	0.89	0.82	-0.15	-0.78	-1.04	-0.66	-0.6	✘	CDR3
Q100G	1.64	0.59	-3.00	-0.77	4.10	-1.13	ND	ND	FR4
L104V	0.81	1.80	-0.39	-1.81	-1.47	-1.91	ND	ND	FR4
REC	-	-	-			-	4.4	✓	-

The single mutants L15P, N28F, K30T and Y96P have previously demonstrated (experimentally) the ability to globally destabilise the LEN monomer ($\Delta\Delta G_{unf}$ +1.70, 1.70, 1.40 and 1.20 kcal/mol respectively) and tested positive for fibril formation. In agreement with these experimental values, L15P, K30T and Y96P were shown to consistently destabilise the monomer by all digital platforms. N28F however was shown to be stabilising by *FoldX*, *SDM* and *I-mutant 3.0.* In contrast to the software iRDP (mention previously) all programs described here were able to consistently highlight the destabilising nature of L15P.

For single mutation Y96P*, FoldX* was unable to predict the highly destabilising nature of this mutation ($\Delta\Delta G_{fold}$ + 0.10 kcal/mol) in comparison to experimental values ($\Delta\Delta G_{unf}$ 1.2 kcal/mol). However, *Eris* was able to highlight the destabilising nature of this mutation ($\Delta\Delta G_{fold}$ + 6.11 kcal/mol), and overall proved to be the most sensitive platform, by its ability to consistency predict stabilising and destabilising mutations in comparison to experimental data. Interestingly, the Y96P mutation was consistently shown by each platform to weaken the stability of the dimer ($\Delta\Delta G_{bind}$ + 0.39, 1.98, 1.32, 3.00, 4.99 by *FoldX*, *mCSM:PPI*, *ELASPIC*, *MutaBind* and *BeAtMuSIC* respectively (

Table **10**). This data suggests that the mutation possesses the ability to simultaneously alter the binding and folding energy.

Out of all REC-like mutants, Y32T is predicted to be the most destabilising mutation by *FoldX* (referring back to our original program used for a benchmark) with a value of $\Delta\Delta G_{fold}$ + 6.77 kcal/mol. This is shared amongst all platforms aside from SDM (which suggests Y96P is the most destabilising).

Based on the observation presented here, this suggests that Y32T is able to simultaneously destabilise binding of the dimer ($\Delta\Delta G_{bind}$ 2.50 kcal/mol by *FoldX*

Table 10) and the stability of the monomer ($\Delta\Delta G_{bind}$ 6.77 kcal/mol by *FoldX*). Analysis of changes to the folding energy of mutants that are located within the CDR1 region that have no associated experimental data (S27aN, S27eA, S29D) shows it is unlikely that they are able to destabilise the native fold of the protein sufficiently to promote aggregation.

Mutants T53S and Q100G were consistently shown to alter the folding energy by *FoldX* ($\Delta\Delta G_{fold}$ of +0.69 and 1.64 kcal/mol) and *ERIS* ($\Delta\Delta G_{fold}$ of +3.12 and 4.1 kcal/mol). Specifically, *ERIS* considers both of these mutations to be more destabilising than K30T which had a calculated $\Delta\Delta G_{fold}$ of 2.22 kcal/mol (calculated by ERIS). This mutation has shown to be destabilizing by experimental methods ($\Delta\Delta G_{unf}$ + 1.40 kcal/mol) and tested positive for fibril formation. Based on this information, (where T53S and Q100G significantly exceed the calculated destabilizing value for K30T) it is possible then, that these mutations destabilise the native fold of LEN sufficiently enough to promote formation, and are later explored on a structural level. Unfortunately, a disagreement between the programs used to calculate the free energy burden for the mutant L104V means this mutation needs to be probed further by structural analysis.

4.2.4. Modelling the effect of REC-like mutation L15P on protein structure

To provide a structural link to how certain REC-like mutations confer a profound or negligible effect on the folding free energy ΔG_{fold} (presented in Table 8) or binding free energy ΔG_{bind} (Table 12), this section performs an exhaustive analysis to assess the consequences of each mutation at atomic-level. Firstly, the REC-like P15L mutation was examined. This mutation, is a solvent accessible residue (52.80% calculated by mCSM-interface) located within a β-turn that links strands A and B. Similar to the SMA-like P40L mutation, this substitution also occurs within a highly conserved framework region, a portion of the variable domain that is not particularly susceptible to somatic mutations in comparison to CDR regions that regularly undergo somatic mutation to provide high antigen specificity (Poshusta, Sikkink et al. 2009). A previous investigation performed by Raffen and co-workers found that LEN harbouring the L15P mutation was dramatically less stable ($\Delta\Delta G_{unf}$ 1.7 kcal/mol) and tested positive for fibril formation (Raffen, Dieckman et al. 1999) which is in agreement with our own computational analysis of this mutation (Table 12).

To provide a rationale to the highly destabilising nature of this mutation, the authors drew attention to a previous study that compared McPC603, a murine κ light chain that differs in 23 residues to LEN with two other proteins of similar structure; 2IMM immunodeficiency virus-type 1 (1ACY), and Fab17-IA (1FOR) (Huang, Chang et al. 1997). The native structure of McPC603 contains an Alanine at position 15, where 1FOR contains a proline. 1ACY on the other hand contains a Leucine at this position. Alanine, proline and Leucine are the

most common residues to occur at this position (as documented in the Ig database (Hutchinson and Thornton 1994) yet Leucine at this position is the only one of these three to participate in long-range interactions which are believed to contribute positively to protein stability (Ohage, Graml et al. 1997).

To explore if the loss in thermodynamic stability seen in the L15P of LEN mutant could indeed be attributable to a loss in the stabilising interactions highlighted by Ohage *et al*., the unaltered crystal structure of LEN (1LVE.pdb) was directly compared to a structure of LEN, harbouring the L15P mutant. Surprisingly, neither display interactions that resemble those highlighted by Ohage *et al*., in the crystal structure of 1ACY. Alternative crystal packing arrangements could be an explanation to the visual differences observed between the structures, i.e. a loop may become trapped in one particular conformation (Rapp and Pollack 2005), but we also noticed a lack of interactions even when performing Rosetta based geometry optimisation of the LEN crystal structure and examining the best ranking model. We acknowledge that Rosetta is unlikely to predict dramatic changes of protein structure in comparison to a dedicated molecular dynamics suite however, the program is able simulate a 3 Å movement (Rocco Moretti 2012), and so, we would expect the hydrophobic side chain of Leucine 15 to occupy this cavity and participate in aforedescribed interactions if the conformation was deemed energetically favourable. In this scenario, it is then difficult to correlate a loss in stabilisation energy with lost interactions. This suggests that the interactions observed by Ohage and colleagues may be unique to 1ACY. Seeking an explanation to the loss in stability, and increase in aggregation propensity specific to this

mutation, the crystal structure of LEN was mutated to L15P followed by energy minimisation protocol as detailed in the materials and methods section.

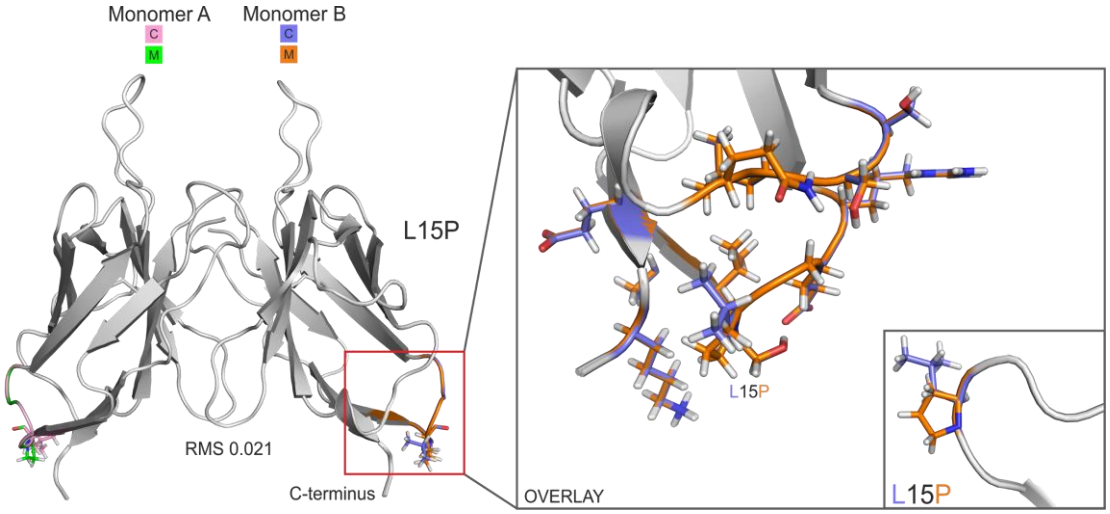


Figure 33 Computational analysis of mutation-induced structural changes of L15P. Dimer structure of native LEN is shown as cartoon overlaid with mutant L15P with mutated residue side-chains shown as sticks. RMS values between control LEN and mutant are indicated. Expanded regions show an overlay of LEN and LEN harbouring the L15P mutant in order to highlight key areas of interest in greater detail. Control (C) monomers are shown in pink (monomer A) and blue (monomer B), with mutant (M) structures in green (monomer A) and orange (monomer B). No changes in hydrogen bonding were detected.

Visual examination of the superimposed mutant and template structures (

Figure 33) reveal that the mutation has minimal effect on the monomeric fold indicated by an RMS deviation of their α-carbons of 0.021 Å. The backbone coordinates of the starting, and mutated residue occupy identical positions, and there are no perceivable changes in hydrogen bonding at the site of mutation, or between neighbouring residues. We also considered the fixed ϕ ψ values of the proline residue and used the software RAMPAGE (Lovell, Davis et al. 2003) to generate a Ramachandran plot to see if this residue fall

into a disallowed region, however it appears that the mutation does not occupy any unfavourable torsion angles and revealed no discernible differences in comparison to the control structure (no residues in outlier regions). This mutation is not part of an aggregation prone region (assessed by Amylpred2, AmyloidFISH and PASTA Figure 29), but surprisingly, Aggrescan3D suggested that the structure of LEN harbouring this mutation was less aggregation prone (mutant -0.22) than the control structure (control +1.46) LEN (Figure 34).

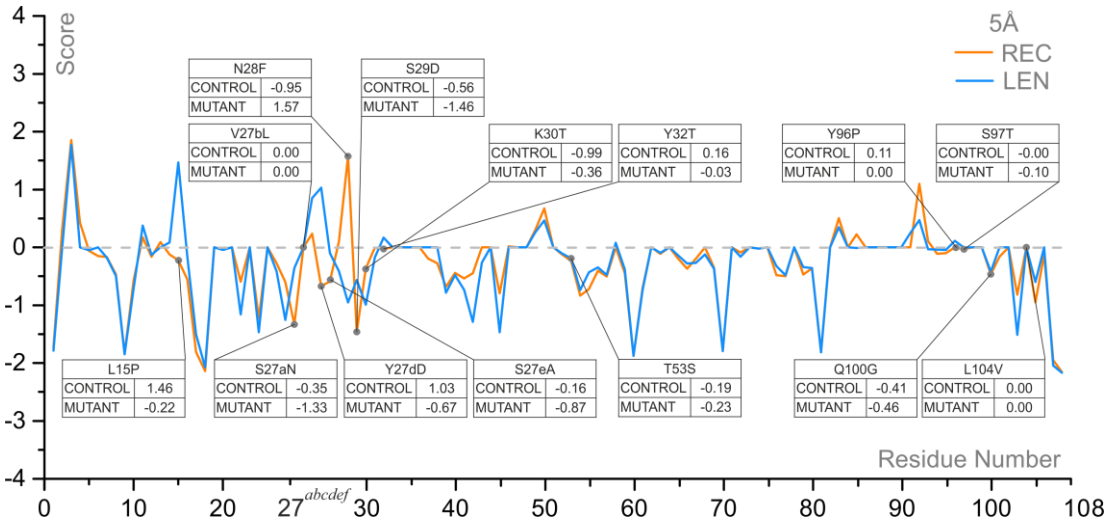


Figure 34 Aggrescan3D prediction profile. Aggrescan3D was used to assess the aggregation propensity of LEN (blue) and REC (orange). The Y axis of the profile corresponds to the prediction scores. Positive scores indicate regions that calculated to be aggregation-prone. Analysis was performed using a 5Å radius to pinpoint the individual residue contributions to overall aggregation potential.

Further inspection of the crystal structure of LEN reveals that the native Leucine forms interactions with the solvent in the crystal structure, but these are preserved following mutation. These observations collectively suggest that

the mutation is well tolerated, which is unusual given the destabilising effect of the mutation calculated by Raffen *et al* (Raffen, Dieckman et al. 1999).

To test if the presence of Glycine 16, a residue neighbouring the mutation ($_{13}$VSLGER$_{18}$) was accommodating the introduction of proline by preventing distortion of the loop, we also created a structure harbouring a G16A mutation in addition to L15P. Again, the double mutant displayed only minimal differences (data not shown). Unfortunately, the data up to this point does not account for the highly destabilising nature of this mutation. Given that the removal of a hydrophobic residue from the solvent (Leucine) and introduction of proline are both entropically advantageous, and the frequency in which prolines are found within loops and β-turns both in nature and in light chain variable domains, with proline, alanine and leucine occupying this position in 98% of V_Ls (Ramirez-Alvarado 2012) the experimentally derived values of this mutation were puzzling.

Seeking a more definitive answer, we next considered if the plasticity of this β-turn could be altered due to the intrinsic conformational rigidity of the Proline. In order to approximate the motion of this loop by simulating its motions in solution, a structure harbouring the mutation was subject to simulations over a 10-nanosecond timescale using the CABS-FLEX (Jamroz, Kolinski et al. 2013) molecular dynamics software (Figure 35).

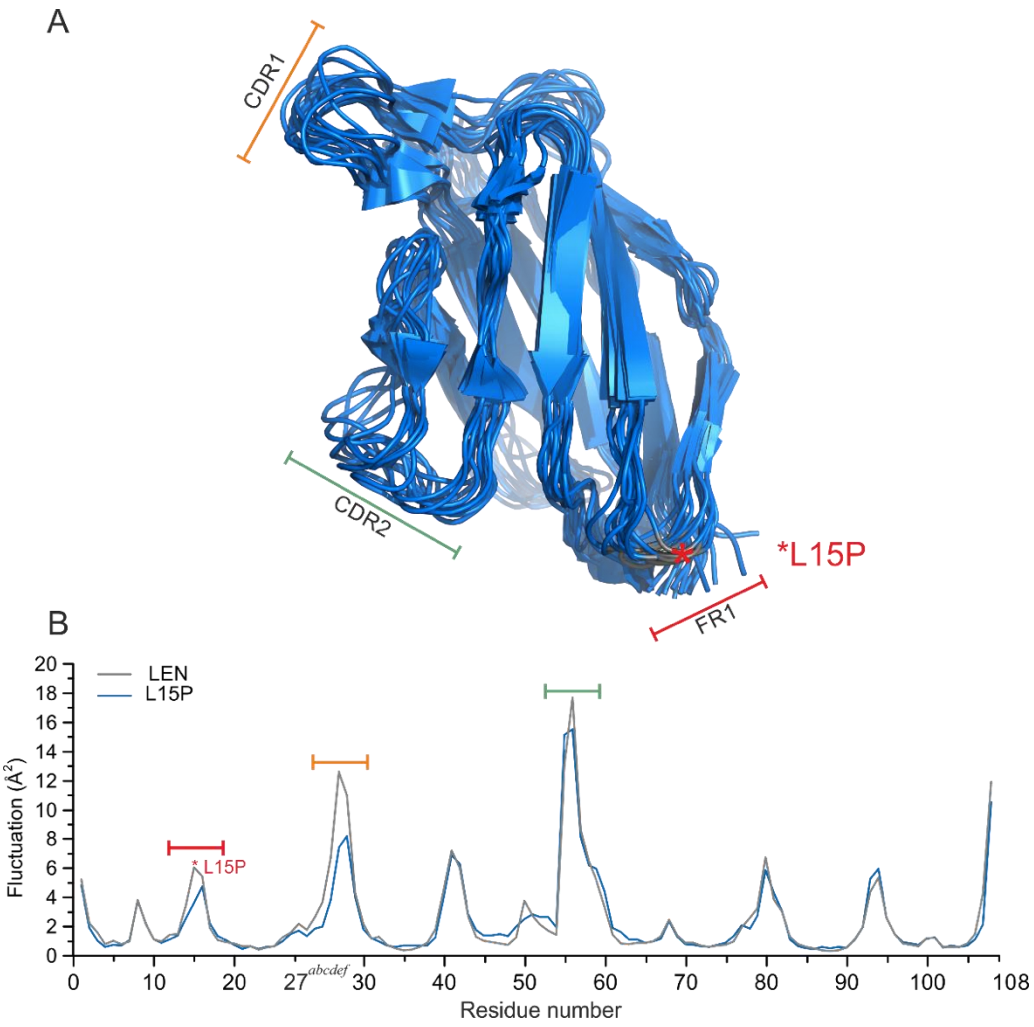


Figure 35 Residue fluctuation profile and structural ensemble of LEN containing REC-type mutant L15P calculated using CABS-FLEX. A single leucine to proline mutation in the β-turn of the FR1 region effects loop dynamics at three distinct regions; the FR1 (red), CDR1 (orange) and CDR 2 (green). Superimposition of all structures acquired from snapshots of time frame. (A) The location of the mutation is highlighted (*). (B) The root mean square fluctuations (RMSF) of *α*-carbon atoms per residue plotted as a function of residue against fluctuation in Å acquired from a 10 nanosecond timescale.

From the superimposition of alpha carbon fluctuations from control and mutant structures the data reveals that the majority of residues remain close to the starting conformation (Figure 35).

There are however, distinct differences in a minority of residues that are limited to three chain segments. Firstly, a decrease in loop mobility over the timescale is seen directly at the site of mutation (~1.5 Å), suggesting that proline is able to constrain the motions in this FR1 located β-turn. Strikingly however, the mutant structure also conveys a decrease in loop motions of the CDR1 and (L26c Y27d, S27e and N28F and CDR2 region (E55 and S56) to a greater effect of ~ 5 Å and ~ 2 Å respectively in comparison to the control structure. The regions contain residues that would be directly involved in the dimer interface. It is interesting to observe changes in loop dynamics that not only extend to regions distant to the site of mutation, but seemingly do so without severely disrupting core structural elements (as seen for the SMA-like P40L mutation (Figure 23) (Saccon, Bunton-Stasyshyn et al. 2013, Broom, Rumfeldt et al. 2015).

Although the difference in energy barriers between structures containing each mutation is unknown, it can be argued that a more flexible structure should be able to transition through the intermediate state (point b Figure 36), into different conformations more easily than a more rigid structure (Figure 36). The presence of a proline residue imposes constraints on the conformational freedom of the backbone (due to its fixed ϕ, ψ values) re-arrangement that leucine (a more flexible residue by comparison) otherwise permits. Furthermore, the best transition state (energetically) may involve disallowed angles which are not accessible with proline, yet are within the leucine mutation and so, proline at this position could be expected to present a higher energy barrier to folding in comparison to leucine.

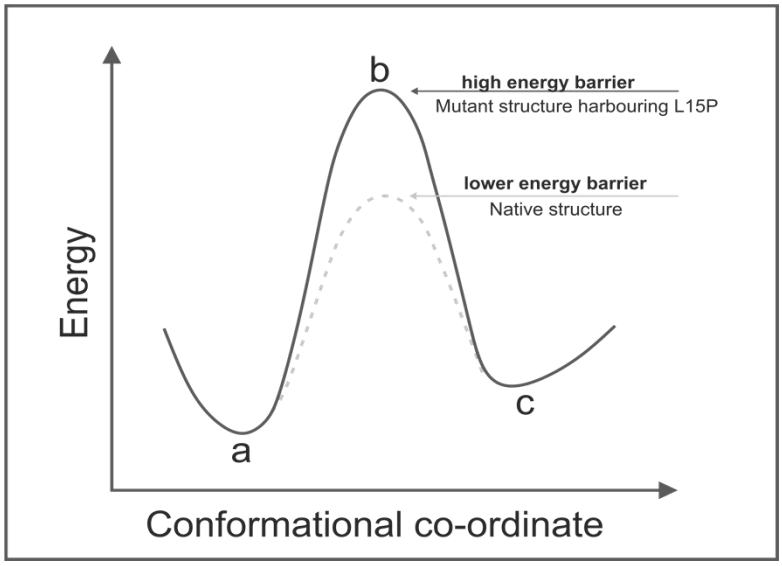


Figure 36 Schematic illustrating one-dimensional energy surface and energy barrier between two protein conformations. In order to adopt a different conformation, the protein must traverse an energy barrier that separates two separate conformations. The figure shows two minima between two folded states. In order for the protein to transition from one state (left, a) to another conformation (right, c) it must transition across the energy barrier (middle, b).

In addition to a change in thermodynamic stability as a consequence of a substitution, the location of the mutation also made us consider the potential impact on fibrillation kinetics. Previously, Marina Ramirez-Alvarado and co-workers noted that the intermolecular packing contents of the V_L BRE seem to resemble the structure of the mature amyloid fibril (Ramirez-Alvarado, De Stigter et al. 2007). Similarly, an analysis of the crystal packing of λ6 protein 6aJL2 containing a P7S mutation performed by Alejandra Hernández-Santoyo and colleagues noted that strand B and the N-terminus participate in intermolecular interactions that allow for V_L monomers to stack (Figure 37 A). The P7S mutation in this study appears to override an evolutionary developed

structural motif, termed the β-bulge which prevents intermolecular interactions that may lead to self-association (Hernández-Santoyo, del Pozo Yauner et al. 2010).

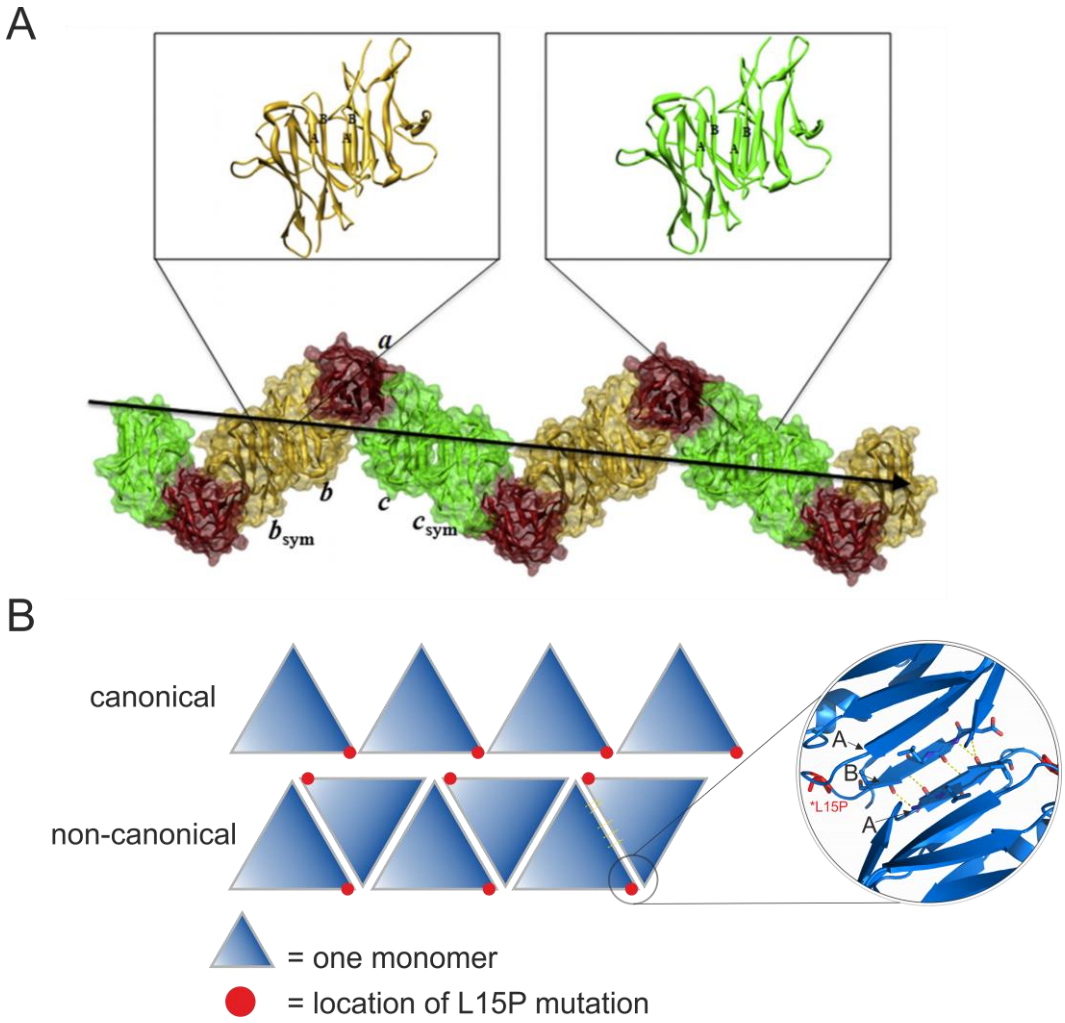


Figure 37 Schematic of V_L monomer stacking. A) Crystal-packing of 6aJL2 mutant harbouring P7S mutation. Expanded region shows non-canonical orientation of dimer with strands A (harbouring the P7S mutation) and strand B labelled. Image adopted from (Hernández-Santoyo, del Pozo Yauner et al. 2010) B) shows proposed model of stacking in LEN harbouring L15P mutation. In the canonical dimer β -strands A and B do not form any intermolecular contacts. In the non-canonical dimer, monomers are able to stack. The location of the L15P mutation (highlighted red) may facilitate β -strand interactions in a similar manner to P7S. Expanded region shows β -strand interactions between LEN monomers harbouring mutation.

Given the ability of this mutation to subdue the motions of this loop, it may present a complementary binding interface that can facilitate β-β strands interactions, which was not existent before the L15P mutation. Although speculative, the location of this mutation and its ability to form fibrils with marginally destabilising effects in comparison to other mutants (Y96 was far more destabilising and had a more profound affect in the loss of intermolecular contacts) could suggest it has a local structural effect that leads to self-association, rather than any large global changes that leads to protein unfolding and subsequent aggregation. Taken together these observations provide a more detailed explanation into the structural perturbations induced by the L15P mutation than existing studies.

4.2.5. Modelling the effect of REC-like mutations on protein structure: CDR1

Next, we address three of the mutations located within the CDR1 loop region; S27aN, V27bL, and Y27dD. The CDR1 region of REC has accumulated the largest number of somatic mutations than any other region, but out of all the REC like mutants of LEN that occur within this region discussed further on in this chapter, Y27dD and V27bL are the only ones so far to experimentally demonstrate increased folding stability of $\Delta\Delta G_{unf}$ -2.7 and -0.7 kcal/mol respectively (illustrated in Table 12). For S27aN, such empirical values of stability have not been experimentally determined. *In silico* analysis of the mutant structure reveals that the serine to asparagine substitution results in small changes in hydrogen bond lengths that are located directly around the mutation site. For example, the NH of S27a is hydrogen bonded to the OH

group of the nearby Y27d at a distance of 2.0 Å in the control structure and 2.2 Å in the mutant (Figure 38).

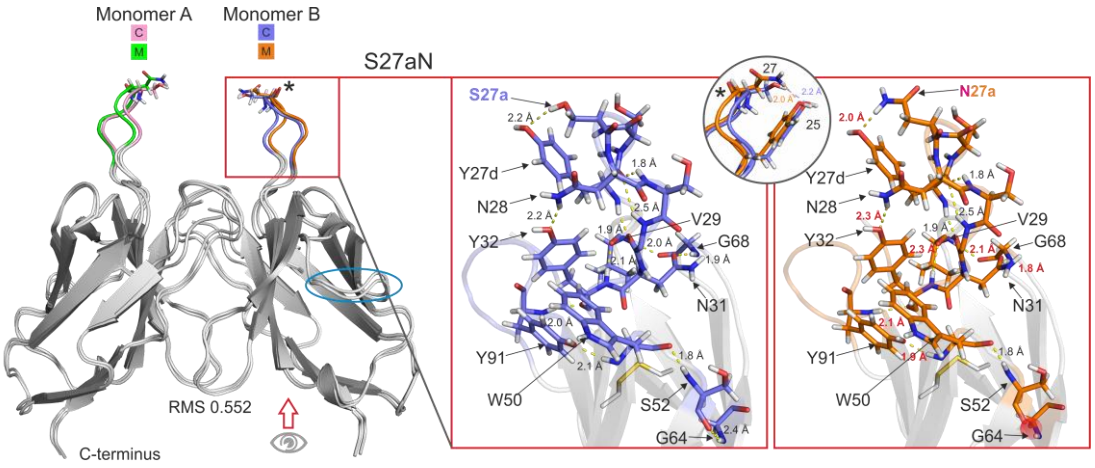


Figure 38 Computational analysis of mutation-induced structural changes of S27aN. Dimer structure of native LEN is shown as cartoon overlaid with mutant (S27aN) with mutated residue side-chains shown as sticks. RMS values between control LEN and mutant are indicated. Colour coded expanded regions highlight key areas of interest in greater detail. The direction of view for these regions is indicated. Control (C) monomers are shown in pink (monomer A) and blue (monomer B), with mutant (M) structures in green (monomer A) and orange (monomer B) Asterisk indicates location of mutation. Perturbation of distal loop that corresponds to residues S52-G64 is highlighted (blue oval). Hydrogen bonds are shown as yellow dashed lines with lengths given; changes in lengths are shown in red text on mutant images. Red spheres indicate loss of hydrogen bond.

It is interesting that such minor structural alterations to the light chain variable domain caused by a single mutation (such as that illustrated in Figure 38) can potentially alter the hydrogen bonding. However, single mutations in variable domains have shown to dramatically change hydrogen bonding networks without inducing any large structural changes (Pokkuluri, Raffen et al. 2002). It is important to highlight that the cut-off for identifying a hydrogen bond in the software used within this thesis is set at 3.5 Å. This is to prevent losing any

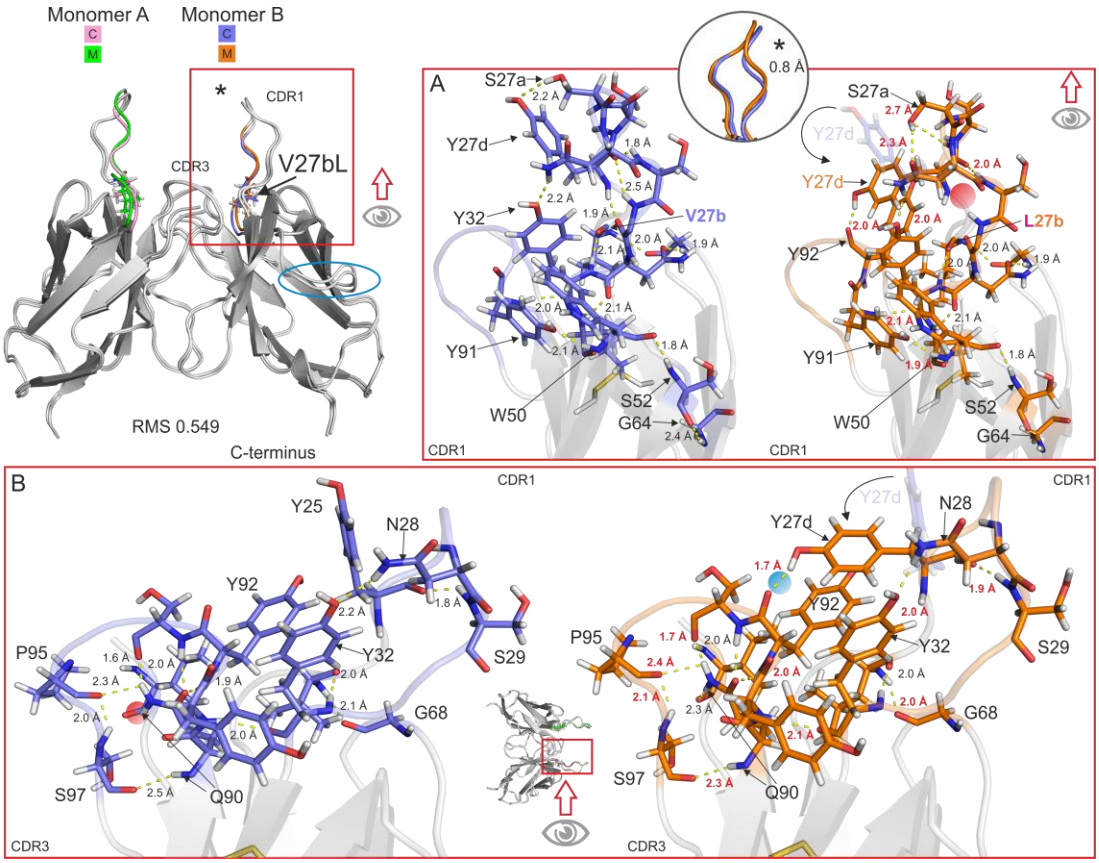
information where the mean length between donor and acceptor in nature is 3.0 Å (Dannenberg 1998). It is also important to consider that the computational methods described within this work provide us with hypothesised changes in hydrogen bonding, the techniques of NMR, X-ray crystallography; CD (if there are large structural alterations) and neutron scattering would permit these to be confirmed experimentally.

Visual inspection of the mutant structure allowed us to map changes in residues that propagate further than the site of mutation where a change in hydrogen bond length between the CO group of Tryptophan 50, and the NH group of Serine 52 located on the top of sheet C'' of the CDR2 region respectively in the control (2.2 Å) and mutant (1.8 Å) is observed. However the bond between the CO group of Serine 52 and the NH group of Glycine 64 is lost, which most likely accounts for the perturbations of Arginine 54-Phenylalanine 62 that lay on the loop connecting these two regions (Figure 38 blue oval).The circular inset (*) shows the extent of the perturbation at the site of mutation (values taken from OH group of S27a to the OH group of Y27d – illustrated Figure 38) indicating that the slightly larger size of Asparagine side chain is easily accommodated in this solvent accessible loop of the CDR1 region. Given the displacements in hydrogen bond length at the site of mutation, the ability to perturb a region located away from the mutation site combined with our computational predictions of folding and binding free energy changes in the protein (

Table **10**), slight decrease to the binding energy in a structure harbouring this mutant (Table 11), and slightly more substantial changes in folding energies (Table 12), it is likely that this mutation is a moderately

destabilising mutation that would have a negligible effect on the folding and stability of the protein. In addition to thermodynamic stability changes, Aggrescan3D suggests that the mutation decreases aggregation propensity of the protein, indicated by a value of -1.33 (in comparison to control of -0.35) (Figure 34).

Next, we inspected the structural differences between the control and REC-type V27bL mutant. The conservative substitution of a Valine to Leucine residue results in a rather minute displacement of the CDR1 loop (Figure 39 panel A), and only small movement of residues within the neighbouring CDR3 region (Figure 39 panel B), most likely to accommodate the slightly larger size of the Leucine side chain. Again, this mutation is accompanied with slight perturbations in hydrogen bonds between residues local to the mutation site, and alike the S27aN mutation a loss in hydrogen bonding between S52 and G64 causes movement of the loop that is located between β-strands D and E.


Monomer A
Monomer B
C
M
CDR1
CDR3
V27bL
RMS 0.549
C-terminus
A
S27a
Y27d
Y32
V27b
Y91
W50
S52
G64
CDR1
0.8 Å
L27b
Y92
B
Y25
N28
P95
S29
G68
S97
Q90
CDR3

There is however a slightly more dramatic rearrangement involving Tyrosine 27d. Movement in the CDR1 loop as a result of the Valine to Leucine substitution results in a loss of hydrogen bonding with S27a, leaving the Y27d to adopt an alternative rotamer, where the residues side chain forms a hydrogen bond with the CO group of the main chain of Y92, a residue located within the F-G loop of the CDR3 region (within close proximity of residues involved within the tyrosine cluster). The close proximity of these aromatics may facilitate pi stacking between the residues (Figure 39) which may tether the CDR1 loop to the CDR3 preventing any increased motions which may expose aggregation prone regions, an important consideration given that the

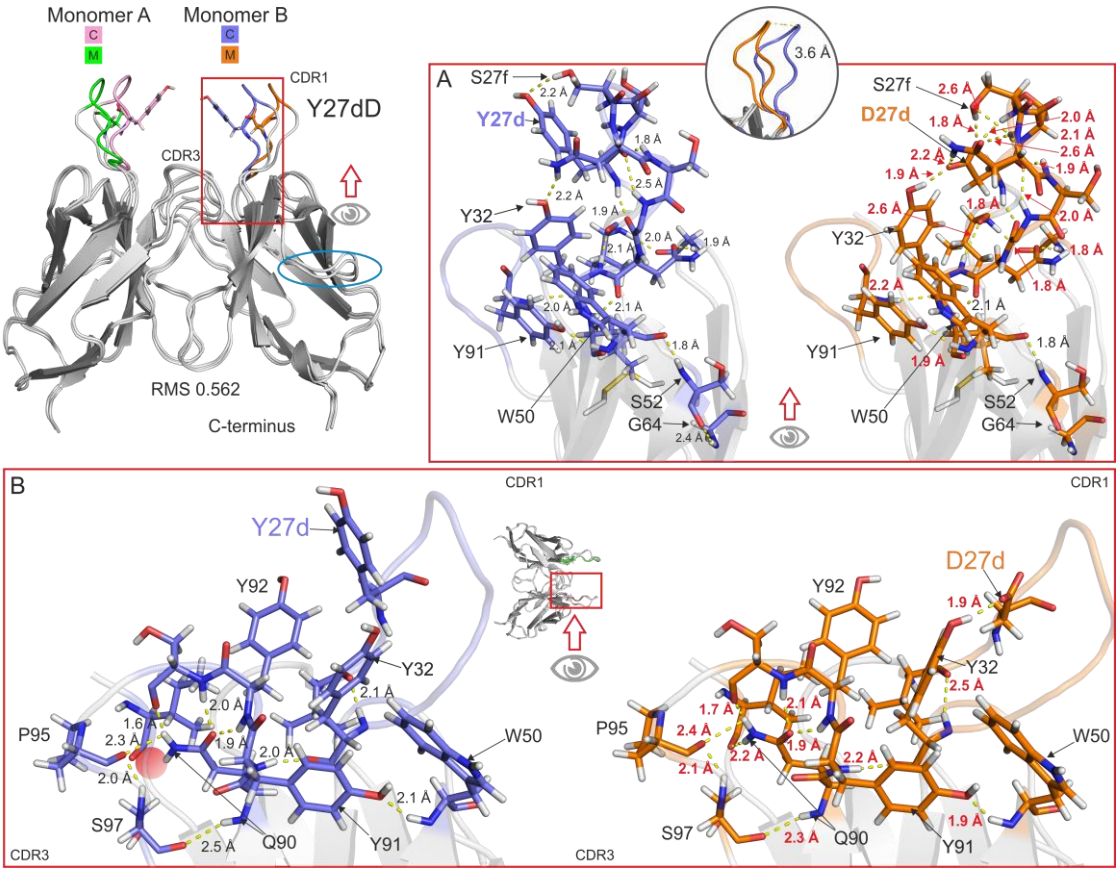
Figure 39 Computational analysis of mutation-induced structural changes of V27bL. Dimer structure of native LEN is shown as cartoon overlaid with mutant (V27bL) with mutated residue side-chains shown as sticks. RMS values between control LEN and mutant are indicated. Colour coded expanded regions highlight key areas of interest in greater detail. Bottom panel shows a side view of the same region of the dimer structure that has been rotated by 90^{o} on the x-axis. The direction of view for these regions is indicated. (C) monomers are shown in pink (monomer A) and blue (monomer B), with mutant (M) structures in green (monomer A) and orange (monomer B) Asterisk indicates location of mutation. Perturbation of distal loop that comprises residues S52-G64 is highlighted (blue oval). Hydrogen bonds are shown as yellow dashed lines with lengths given; changes in lengths are shown in red text on mutant images. Red spheres indicate loss of hydrogen bond. In this case, blue sphere indicates formation of new hydrogen bond concerning Tyrosine 25.

CDR1 region in its entirety and a portion of the CDR3 region are predicted to be part of an aggregation prone stretch (Figure 29). The resulting Leucine also establishes new hydrogen bonds between residues G68. These observations may account for the slight increase in stability shown in thermal denaturation

experiments where a structure harbouring the V27bL mutation was stabilised by -0.7 kcal/mol (Table 12). The most pronounced stabilising effects of any of the mutations within this study is Y27dD. The substitution of an aromatic Tyrosine for the charged Aspartic acid leads to a large addition of hydrogen bonds directly at the site of mutation (Figure 40) which also extends, again to the CDR3 region. Such a dramatic increase in the number of hydrogen bonds is likely to account for the significant increase in folding stability ($\Delta\Delta G_{unf}$ -2.7 kcal/mol) demonstrated previously (Table 12) (Raffen, Dieckman et al. 1999).

To add greater confidence to these results, the predictions here are also supported by a former crystallographic study, where a crystal structure harbouring the Y27dD mutation (along with a M4L substitution) also revealed an increase in hydrogen bonding between S27f (Figure 40) (Pokkuluri, Raffen et al. 2002).

Of particular interest is that the three mutations covered here (S27aN, V27bL, and Y27dD corresponding to Figure 38 Figure 39 Figure 40 respectively) each possess the ability to induce structural changes that propagate from the site of mutation to residues that are distant both in sequence and structure. Such a “cascade” mechanism is described for other REC-like mutations further on in this study, but unique to these three mutants is their ability alter structural elements within the CDR3 region, which was previously identified to be a key area from analysis of all SMA-like mutations, and also their ability to cause movements in the loop located between β-strands D and E. Together, these changes largely account for an increased RMS over other mutations that occur within a similar region of the protein.


Monomer A
Monomer B
C
M
CDR1
Y27dD
CDR3
RMS 0.562
C-terminus
A
S27f
Y27d
Y32
Y91
W50
S52
G64
3.6 Å
D27d
B
CDR1
CDR3
P95
S97
Q90

An additional LEN to REC mutation that introduces an aspartic acid into the hypervariable CDR1 region is S29D. The small polar side chain Serine 29 forms hydrogen bonds with several surrounding residues (Figure 41) which are not disrupted once replaced with an aspartic acid, indicated by a low RMS of 0.103. There are no experimentally derived values to how this mutation alters the stability and aggregation potential of LEN. However, based on our own Aggrescan 3D result (Figure 34) and changes to stability (Table 12) it is likely

Figure 40 Computational analysis of mutation-induced structural changes of Y27dD. Dimer structure of native LEN is shown as cartoon overlaid with mutant (Y27dDL) with mutated residue side-chains shown as sticks. RMS values between control LEN and mutant are indicated. Colour coded expanded regions highlight key areas of interest in greater detail. Bottom panel shows a side view of the same region of the dimer structure that has been rotated by 90^{o} on the x-axis. Control (C) monomers are shown in pink (monomer A) and blue (monomer B), with mutant (M) structures in green (monomer A) and orange (monomer B). Perturbation of distal loop that comprises residues S52-G64 is highlighted (blue oval). Hydrogen bonds are shown as yellow dashed lines with lengths given; changes in lengths are shown in red text on mutant images. Red spheres indicate loss of hydrogen bond.

that this mutation is stabilising.

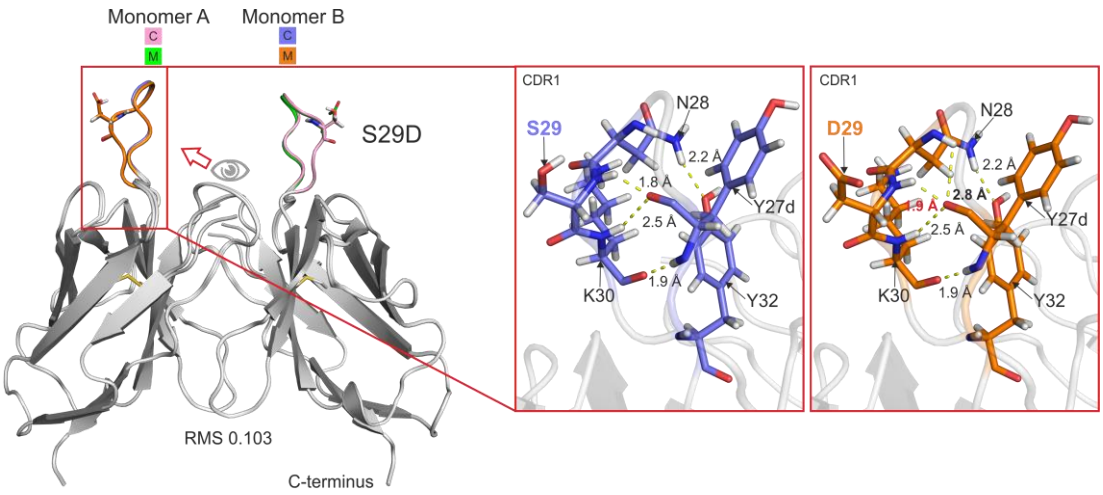


Figure 41 Computational analysis of mutation-induced structural changes of S29D. Dimer structure of native LEN is shown as cartoon overlaid with mutant (S29D) with mutated residue side-chains shown as sticks. RMS values between control LEN and mutant are indicated. Expanded regions highlight key areas of interested in greater detail. Control (C) monomers are shown in pink (monomer A) and blue (monomer B), with mutant (M) structures in green (monomer A) and orange (monomer B). Hydrogen bonds are shown as yellow dashed lines with lengths given; changes in lengths are shown in red text on mutant images.

A comparison between the control and another REC-type mutant of the CDR1 region, S27eA (Figure 42), shows there are very minor structural changes induced by the substitution (RMS 0.223). Alike S27aN, no experimental values have been achieved for this mutation, although based on the rather insignificant changes, the mutation is likely to be neutral, in its ability to alter the binding free energies (

Table **10**) and was found to be slightly stabilising (for the monomer) in our own computational analysis of ($\Delta\Delta G_{fold}$ -0.252, 0.246, 0.62, 0.493, 2.34 kcal/mol by *FoldX*, *mcSM-stability*, *SDM*, *DUET* and *ERIS* respectively) (Table 12).

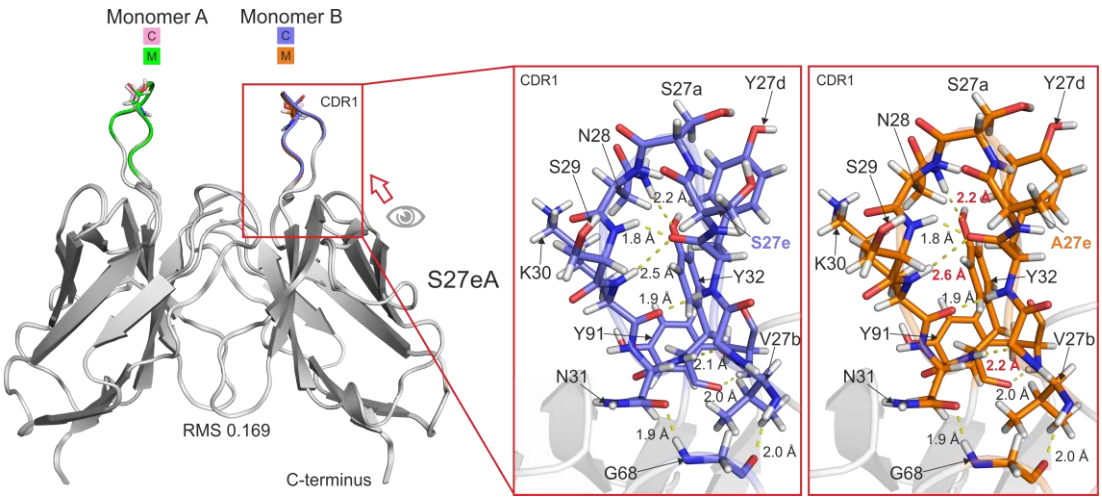


Figure 42 Computational analysis of mutation-induced structural changes of S27eA. Dimer structure of native LEN is shown as cartoon overlaid with mutant (S27eA) with mutated residue side-chains shown as sticks. RMS values between control LEN and mutant are indicated. Expanded regions highlight key areas of interested in greater detail. Control (C) monomers are shown in pink (monomer A) and blue (monomer B), with mutant (M) structures in green (monomer A) and orange (monomer B). Hydrogen bonds are shown as yellow dashed lines with lengths given; changes in lengths are shown in red text on mutant images.

In contrast to the aforementioned CDR1 region mutations (S27aN, V27bL, S27eA, Y27d, S29D), the remaining mutants of the same region; N28F, and K30T demonstrate destabilising effects of +1.37, 1.45, kcal/mol respectively ($\Delta\Delta G_{fold}$ experimental calculations). Unfortunately, there was inconsistency between the results acquired for N28F in our own computational analysis (which showed stabilising and destabiling outcomes) yet, changes to the folding free energy for K30T were largely in agreement with experimental data (Table 12) where each program was able to calculate the destabilising nature of this residue. In addition, Y32T has no accompanying experimental data, but was shown largely to be highly destabilising (more so than K30T by our own computational analysis ($\Delta\Delta G_{fold}$ +6.7 , 2.29, 1.07, 2.33, 9.35, 1.38 kcal/mol

(Table 12) and was also shown to affect the binding energy of the dimer ($\Delta\Delta G_{bind}$ + 2.52 1.14, 1.75, 0.68 kcal/mol) (

Table **10**). For N28F the decrease in stability ($\Delta\Delta G_{unf}$ +1.7 kcal/mol shown in Table 12 is likely to be attributable to the non-conservative replacement of asparagine to an aromatic moiety. It is also worth mentioning that this mutation granted a significant increase in aggregation propensity (-0.53 for control and 1.57 for the mutant structure, as assessed by Aggrescan3D (Figure 34). As illustrated in Figure 43 the side chain of Asparagine 28 participates in hydrogen bonding of 2.2 Å with Y32, a residue located near β-strand C of the CDR3 region and is part of the aforementioned Tyrosine cluster (DiCostanzo, Thompson et al. 2012). Upon mutation, the interaction between these two regions is lost Figure 43. This also causes alterations to the hydrogen bond length, and angle between residue Y32, and Y91, another residue of the tyrosine cluster that was mentioned previously (Figure 32) (DiCostanzo, Thompson et al. 2012). Equally, this observation is mirrored in Y32T where there are alterations in bond length and angle involving Y91. Despite the distal location of the residue to the dimer interface, perturbation to the interactions between residues that occur within the CDR3 region, which is involved within the dimer interface presumably account for the changes seen to the binding free energies documented earlier Table 12. K30T however, does not display the downstream changes involving the tyrosine cluster in comparison to the other two mutations. Instead, this mutation causes local changes.

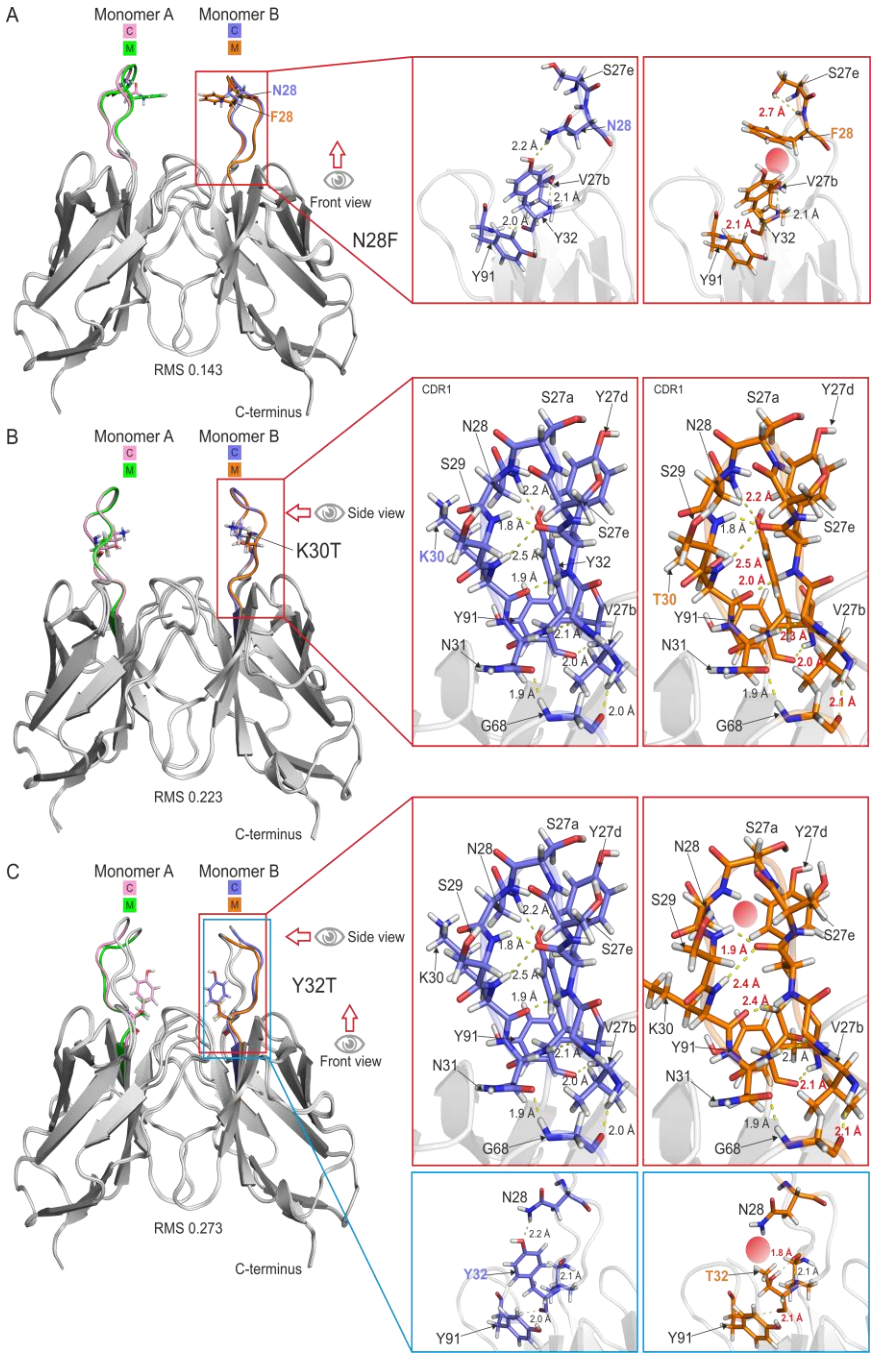

A
Monomer A
Monomer B
N28
F28
Front view
N28F
RMS 0.143
C-terminus
S27e
V27b
Y32
Y91
2.2 Å
2.1 Å
2.0 Å
2.7 Å
B
Side view
K30T
RMS 0.223
CDR1
S27a
Y27d
S29
K30
T30
N31
G68
1.8 Å
2.5 Å
1.9 Å
C
Y32T
RMS 0.273
2.4 Å

Figure 43 Computational analysis of mutation-induced structural changes of N28F, K30T and Y32T. Dimer structure of native LEN is shown as cartoon overlaid with mutant (N28F, K30T and Y32T) with mutated residue side-chains shown as sticks. RMS values between control LEN and mutant are indicated. Colour coded expanded regions highlight key areas of interest in greater detail. The direction of view for these regions is indicated. Control (C) monomers are shown in pink (monomer A) and blue (monomer B), with mutant (M) structures in green (monomer A) and orange (monomer B). Hydrogen bonds are shown as yellow dashed lines with lengths given; changes in lengths are shown in red text on mutant images. Red spheres indicate loss of hydrogen bond.

4.2.6. Modelling the effect of REC-like mutations on protein structure: FR4 and CDR3

Next, the structural alterations caused by REC like substitution T53S were explored. Inconsistencies between platforms that assessed the changes to the binding free energy as a consequence of this mutation motivated us to explore this mutation at a structural level (

Table **10**). Visual inspection reveals that the conservative substitution of threonine to serine at position 53 results in a loss of hydrogen bonding that extends far from the site of mutation both in sequence and structure (Figure

44). Specifically, changes at the site of mutation lead to perturbations that extend to the CDR1, CDR3 and FR region (residues used as example highlighted in Figure 44).

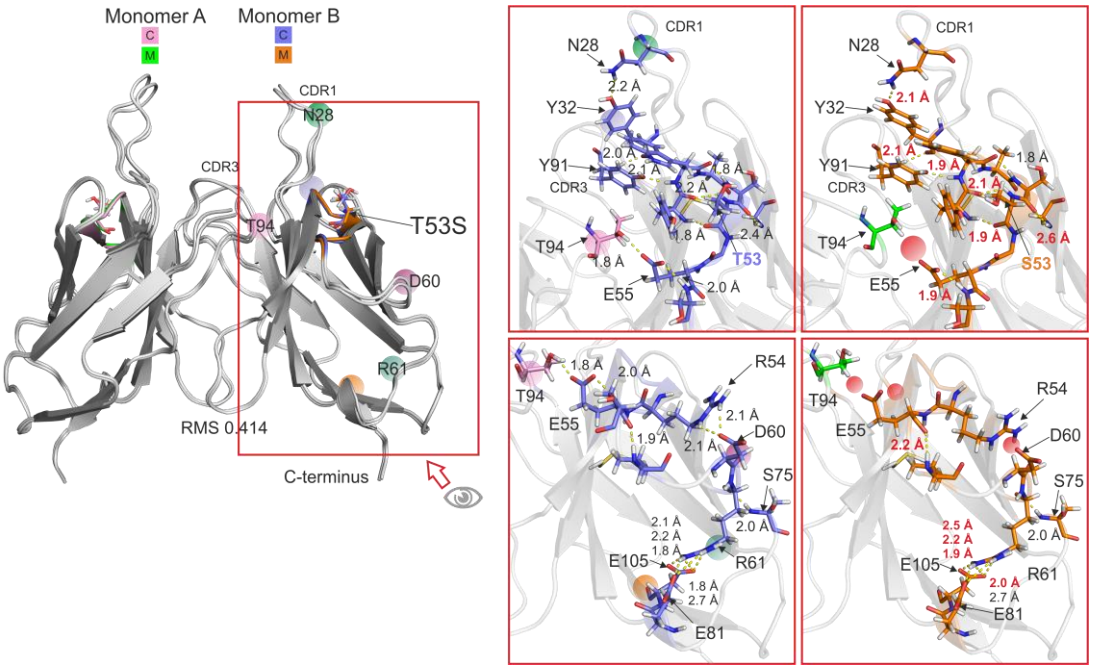


Figure 44 Computational analysis of mutation-induced structural changes of T53S. Dimer structure of native LEN is shown as cartoon overlaid with mutant (T53S) with mutated residue side-chains shown as sticks. RMS values between control LEN and mutant are indicated. Colour coded expanded regions highlight key areas of interest in greater detail. The direction of view for these regions is indicated. For clarity, regions of the structure have additional labelling and have been shaded by a coloured sphere (green, pink, blue purple, green orange) which match zoomed region. Control (C) monomers are shown in pink (monomer A) and blue (monomer B), with mutant (M) structures in green (monomer A) and orange (monomer B). Hydrogen bonds are shown as yellow dashed lines with lengths given; changes in lengths are shown in red text on mutant images. Red spheres indicate loss of hydrogen bond.

While this mutation causes a loss in interchain hydrogen bond, (between E55 and residue T94), the interfacial analysis of the dimer harbouring this mutation also showed on minimal loss in binding energies (BE of 13.5 kcal/mol in comparison to control of -13.6 kcal/mol) (Table 11). Therefore, it is unlikely that it would be able to confer any dramatic loss in dimer stability.

There was much more consistency between programs when calculating the ability of this mutation to alter the folding energy, where in every case ($\Delta\Delta G_{fold}$ = 0.691, 1.107, 0.97. 0.993, 3.12 and 0.45 kcal/mol assessed by *FoldX*, *mCSM*, *SDM*, *DUET*, *ERIS* and *I-mutant 3.0* respectively) was calculated to be destabilising (Table 12).

Given such profound changes to the intrachain contacts that propagate throughout the V_L monomer, and the calculated values, this residue is likely to alter the stability of the monomer. It is interesting to note that, the accumulation of non-conservative mutations at specific structural locations are suggested to be key determinants for light chain amyloidogenicity (rather than

the quantity of non-conservative mutations) based on a recent investigation (Poshusta, Sikkink et al. 2009, Ramirez-Alvarado 2012). Yet, the observations here present a unique case where a conservative mutation is able to induce quite dramatic effects that may confer increased aggregation potential.

As documented previously (chapter 2 and 3), the SMA like mutation Y96Q resulted in a significant decrease in both binding and folding energies. Based on the PISA analysis of canonical and non-canonical dimers harbouring this mutation (Table 5), the low energetic values suggested that this mutation could in fact destabilise both interfaces (negative ΔG^{diss}). Examination of structures harbouring this mutation revealed dramatic changes in NMR R_{ex} value (Figure 21), and both a rearrangement in intra molecular hydrogen bonds, and a loss of intermolecular contacts (Figure 28). This residue then, may dictate the dissociation of the dimer into aggregation prone monomers, and in this state, make them more liable to unfolding.

Here, we further examined the impact of the Y96P mutation (Figure 45) and found that even though the substitution occurs in an identical position; the mutation is better tolerated than Y96Q. This is not only evidenced by our previously calculated changes to inter and intramolecular energies, but there are much fewer changes to the structure of LEN harbouring this mutation.

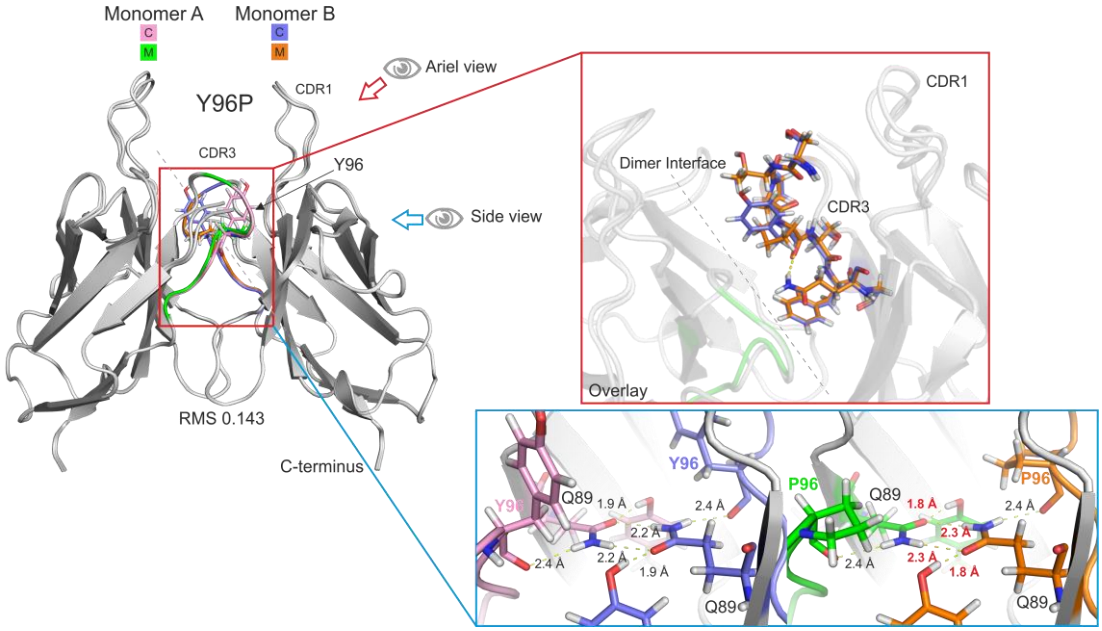


Figure 45 Computational analysis of mutation-induced structural changes of Y96P. Dimer structure of native LEN is shown as cartoon overlaid with mutant (Y96P) with mutated residue side-chains shown as sticks. RMS values between control LEN and mutant are indicated. Colour coded expanded regions highlight key areas of interest in greater detail. Bottom panel shows view of structure rotated by ~45^{o} on Y-axis Control (C) monomers are shown in pink (monomer A) and blue (monomer B), with mutant (M) structures in green (monomer A) and orange (monomer B). Hydrogen bonds are shown as yellow dashed lines with lengths given; changes in lengths are shown in red text on mutant images.

The remaining two mutations, Q100G and L104V (S97T also a mutation of SMA and was covered in previous chapter) are located in the highly conserved FR4 region. Neither mutation has any associated experimental derived values of binding or folding free energies. Introducing mutant Q100G into the structure of LEN was accompanied by mostly low values for $\Delta\Delta G_{bind}$ (-0.06, 0.007, 0.692 kcal/mol by *FoldX*, *mCSM*, *ELASPIC*, where MutaBind predicted this mutation to be destabilising 1.23 kcal/mol) despite its interfacial location (Table 12). PISA interfacial analysis suggested this mutation also had only a

minor effect on the binding energy of the canonical interface (BE 13.1 kcal/mol in comparison to control of BE 13.6 kcal/mol) (Table 11). Visual inspection of this structure (Figure 46) revealed no discernible changes between interfacial contacts; collectively suggesting that the mutation is unlikely to alter the stability of the dimer. This mutation was however, consistently computed to reduce folding free energies (Table 12) and was calculated to have a more destabilising effect than K30T (which tested positive for fibril formation) (Raffen, Dieckman et al. 1999).

One consideration for this glutamine to glycine mutation is the increased conformational freedom, which arises through the less restricted phi/psi angles of the protein backbone, potentially resulting in a more favourable entropy change for the unfolded state. In addition, inspection of the control structure reveals that Gln in this position participates in hydrogen bonding between residues of neighbouring β-strands which contribute to overall stability. Upon substitution to a glycine these hydrogen bonds are no longer formed (Figure 46) which is presumably due to the small size of the residue (lack of side chain). Interactions between residues Thr102 of β -strand G and Tyrosine 86 of β -strand F (Figure 46) are also altered. This loss of hydrogen bonding in a highly conserved FR4 region of the protein that rarely undergoes mutation, may increase chain flexibility and also remove the enthalpic contribution that these hydrogen bond provide to overall stability. The net result of these effects in the model is not dramatic, however these observations in combination with the computed $\Delta\Delta G_{fold}$, suggests that it is possible that this mutation has a destabilising outcome.

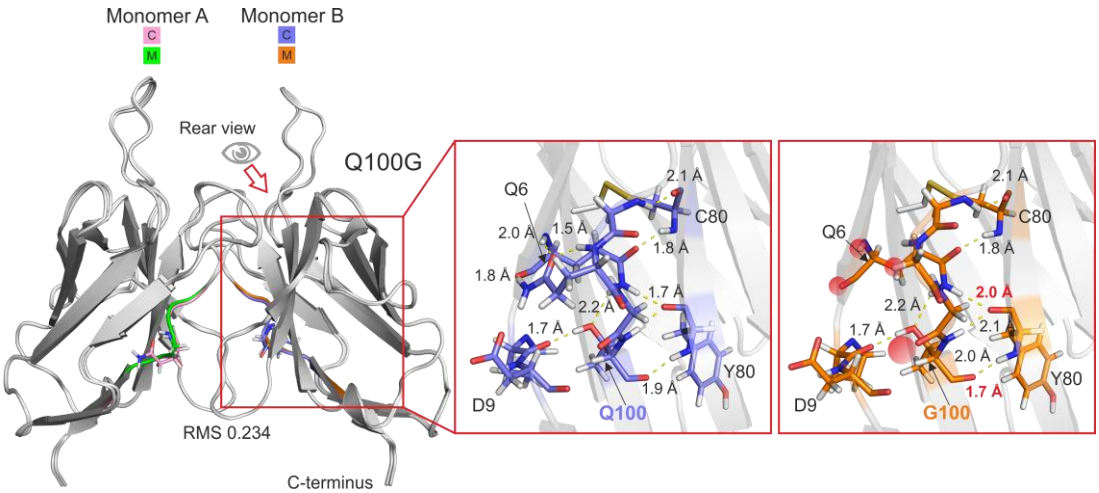


Figure 46 Computational analysis of mutation-induced structural changes of Q100G. Dimer structure of native LEN is shown as cartoon overlaid with mutant (Q100G) with mutated residue side-chains shown as sticks. RMS values between control LEN and mutant are indicated. Expanded regions highlight key areas of interest in greater detail. Control (C) monomers are shown in pink (monomer A) and blue (monomer B), with mutant (M) structures in green (monomer A) and orange (monomer B). Hydrogen bonds are shown as yellow dashed lines with lengths given; changes in lengths are shown in red text on mutant images. Red spheres indicate loss of hydrogen bond.

Finally, we inspect the modelled structure of LEN L104V. The inconsistencies in calculated $\Delta\Delta G_{bind}$ values acquired for this mutation ($\Delta\Delta G_{bind}$ + 0.713, 0.678, 1.39 by mCSM, ELASPIC and Mutabind yet was shown to be neutral by *FoldX*, and stabilising (-0.32 kcal/mol) by BeAtMuSIC) (

Table **10**) again prompted us to assess the structural changes of this mutation in order to provide a more sophisticated understanding into the consequences of this mutation merely than taking the calculations at face value.

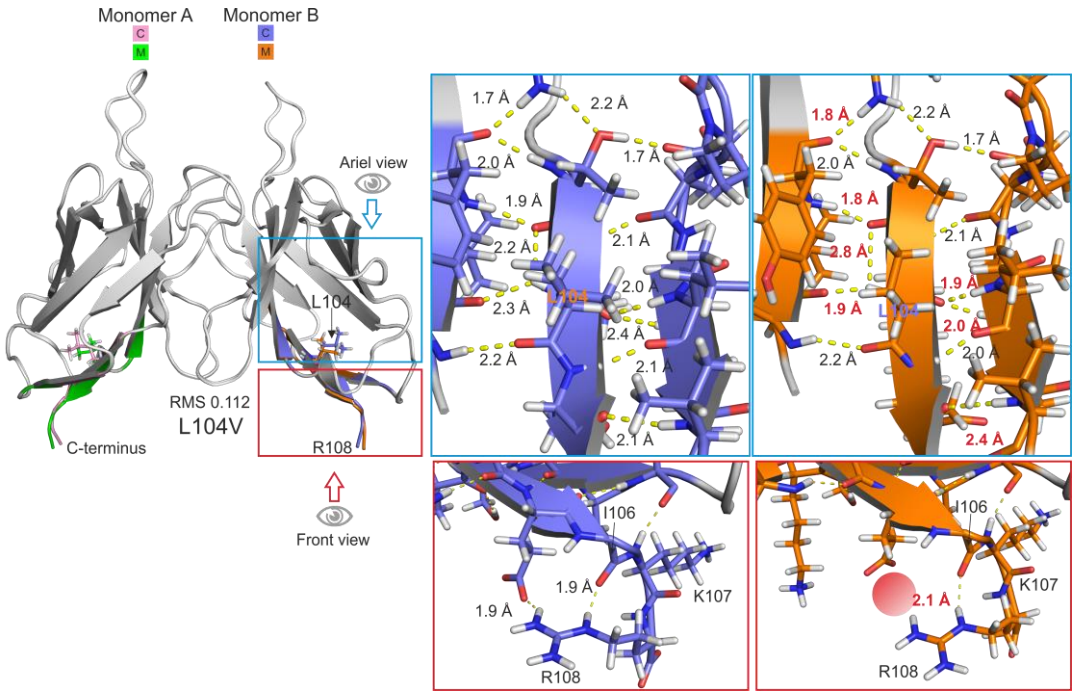


Figure 47 Computational analysis of mutation-induced structural changes of L104V. Dimer structure of native LEN is shown as cartoon overlaid with mutant (L104V) with mutated residue side-chains shown as sticks. RMS values between control LEN and mutant are indicated. Expanded regions highlight key areas of interested in greater detail. Control (C) monomers are shown in pink (monomer A) and blue (monomer B), with mutant (M) structures in green (monomer A) and orange (monomer B). Hydrogen bonds are shown as yellow dashed lines with lengths given; changes in lengths are shown in red text on mutant images. Red spheres indicate loss of hydrogen bond.

Inspection of the control and mutated structure shows that there are no major perturbations between the structures at the site of mutation, and overall the structure remains largely unaltered (RMS 0.112). Given the location of the mutation, it is unlikely to play any significant part in weakening the dimer, there are however noticeable losses and alterations to the hydrogen bonding network between β-sheets. Based on these observations and the computational values ($\Delta\Delta G_{fold}$ 0.810, 1.805, 1.813, 1.91 of FoldX,

mCSMstability, Duet and I-mutant 3.0) it is possible that this mutation has a destabilising effect.

4.2.7. Main findings and summary

Summary

Using a number of individual platforms to assess the binding and folding free energies as a consequence of REC-like mutations, has highlighted destabilising and stabilising mutations that not only complement *in vitro* experiments performed by other groups, but also highlight new residues that have previously been uncharacterised. Complementing the calculated inter and intra molecular free energy changes with an associated structure may help reveal how mutations cause a loss, or gain of intra-domain/protein contacts which can be linked to a change in stability in novel mutations of other V_L previously uncharacterised.

Highlights

In silico analysis of each REC-like mutation revealed that Y32T and Y96P, residues located within the CDR1 and CDR3 region were the two most strongly destabilising mutations that led to large reduction in binding energies in a number of software (Table 11). Changes to $\Delta\Delta G_{bind}$ from the remaining 12 REC-like mutations were largely minimal by comparison. In the previous chapter, Y96Q was able the only SMA-like residue to disrupt the energetic features of both canonical and non-canonical dimers significantly enough that that the structure was indicated as unstable based on the predictions by PISA. By comparison, Y96P discussed in this chapter was unable to destabilise the interface to the same level (Table 11). This prediction agrees with the

observation made by *Stevens et al*., and reaffirm the fact, that an aromatic or hydrophobic residue at position 96 is integral to maintaining a stable dimer (Fred J. Stevens 1980). Unusually, Y32T, a residue located away from the dimer interface was also able to alter $\Delta\Delta G_{bind}$. We suggest that its close proximity to Y91 and Y96, aromatics that are part of a conserved tyrosine cluster may account for the observed changes. As the amyloid formation in AL amyloidosis is recognized to be a two-state process where dimer dissociation precedes aggregation, it is likely that Y96 is a key residue in this process and may shift the equilibrium from dimers to more aggregation prone monomers in solution.

Earlier findings have shown the L15P mutation was able to reduce the thermodynamic stability ($\Delta\Delta G_{unf}$ +1.7 kcal/mol) and induce fibril formation (Raffen, Dieckman et al. 1999) but molecular level rationale for this finding has gone unknown. Interestingly, no changes were observed in our own computational analysis; however the nature of the residue (fixed phi, psi angles of the protein) and position of this mutation also made us consider local alterations rather than global destabilising effects where we found this solvent exposed residue to possibly allow for intermolecular β-strand interactions to form.

For the work conducted by *Davis et al.,* who assessed the changes in stability of many of the REC-like mutations in conditions where the light chain variable domain is largely in the monomeric state, the results in this chapter complement these original findings with high resolution structural models, but also reveal that Q100G and T53S, two mutations that were previously unobtainable by experimental methods are likely to be destabilising to the

monomer, but not the dimer. A significant loss in hydrogen bonding at highly conserved β-sheet regions may be critical to maintaining stability and the fold of the monomer, allowing it to transition from the native Ig domain and sample states that are more aggregation prone more easily. Overall, these results suggest that it is likely that the CDR3 region is a key mediator in dimer stability, where mutations that occur in the conserved framework regions play a more significant role in the stability of the light chain monomer. It is also important to consider not only the location of the mutation, but also the nature – conservative to non-conservative substitutions.

5. The recombinant expression and purification of the light chain variable domains; SMA and LEN

5.1. Introduction

The 114-amino acid immunoglobulin light chain variable domain SMA and LEN were originally isolated from patients suffering from either multiple myeloma or light chain amyloidosis (Stevens, Raffen et al. 1995). SMA is amyloidogenic *in vivo* but LEN adopts a stable dimer, displaying amyloidogenic properties under destabilising conditions only *in vitro*. A recombinant *E. coli* protein expression system was previously established for these V_L proteins, employing lysosyme cell disruption and purification using a multi-step chromatographic strategy of strong anion and cation exchange followed by gel filtration (Stevens, Raffen et al. 1995). For LEN, yields were reported to be around 10 mg/L, while SMA was reported to be less. *Rognoni et al.,* 2013 reported an optimised procedure for obtaining light chain variable domains through recombinant expression and refolding from inclusion bodies (Rognoni, Lavatelli et al. 2013). After numerous trials, we found that this method was not successful for SMA and LEN resulting in a lower recovery from refolding, and elution with many co-contaminants. In addition, refolding is renowned to result in variable isomerisation undesirable for subsequent structural analysis (Vallejo and Rinas 2004, Berkmen 2012, Singh, Upadhyay et al. 2015). We also found that previous reports for expression of these proteins have sparse details, use expensive or out-of-date equipment and cell lines (e.g. JM83). We present here an alternative strategy exploiting periplasmic expression of the two light chain proteins SMA and LEN that improve on previous methods, employing a modern cell line optimised for the synthesis of toxic proteins. (Stevens, Raffen et al. 1995, Khurana, Souillac et al. 2003). Periplasmic expression can result in suboptimal yields and incomplete removal of peptide leader sequences

(Rognoni, Lavatelli et al. 2013), however we show that in our system we have improved yields, comparable with those obtained through the more complex refolding process and have complete removal of the leader sequence confirmed by mass spectrometry. We propose a simplified purification process and avoid the use of lysozyme which can cause complications in purification. We confirm with CD, multi-dimensional NMR and SEC-MALLS that the proteins produced here have secondary structure consistent with other V_Ls.

5.2. Materials & Methods

5.2.1. Plasmids and cloning

The LEN and SMA genes (Table 13) were synthetically produced by Life Technologies and inserted into pOPINO plasmids by the Oxford Protein Production Facility. pOPINO comprises a signal sequence based on OmpA prior to the protein of interest (POI), followed by a lysine residue and polyhistidine tag with ampicillin resistance (Figure 48). To remove the polyhistidine tag at the C-terminus, a premature stop codon was introduced directly upstream of the oligonucleotide containing lysine and 8 histidines by mutagenesis using the Site-directed, Ligase-Independent Mutagenesis (SLIM) method (Chiu, March et al. 2004) to generate the plasmid named LEN and SMApOPIN_ompAstop. All plasmids were sequenced prior to use (Source Bioscience). A detailed procedure for the generation of the plasmids is described in the following sections.

Table 13 Biochemical properties of SMA and LEN. Amino acid sequence of the V_L domains were acquired from the Amyloid Light Chain Database (ALBase Boston university) using the patient ID's as search. Isoelectric point (pI) and molecular weight were derived using Expasy (Wilkins, Gasteiger et al. 1999)

Variable domain	Molecular weight (Da)	Isoelectric point (pl)	Amino acid sequence
SMA	12735.19	7.96	DIVMTQSPDSLAVSLGERATINCKSSQSVLY SSNNRNYLAWYQQKLGQPPKLLIYWASTRE SGVPDRFSGSGSGTDFTLTISSLQAEDVAV YYCHQYYSHPQTFGQGTKLELKR
LEN	12640.08	7.92	DIVMTQSPDSLAVSLGERATINCKSSQSVLY SSNSKNYLAWYQQKPGQPPKLLIYWASTRE SGVPDRFSGSGSGTDFTLTISSLQAEDVAV YYCQQYYSTPYSFGQGTKLEIKR

5.2.2. Site directed mutagenesis

To remove the polyhistidine tag at the C-terminus, a premature stop codon was introduced directly upstream of the oligonucleotide coding for this region. SLIM protocols for a PCR-mediated mutagenesis were used to introduce the premature stop codon. This procedure uses a form of Inverse PCR, where the primers used to amplify the template DNA strand are in the reverse direction (Ochman, Gerber et al. 1988) A total of four primers; two forward and two reverse are used in the reaction. Each primer pair contains a long tailed primer and a short set of primer (F_T, R_T and R_S, $R_{L)}$ where the tailed primers are designed to carry the mutation on their complementary sequences, located at the 5' end. A full description for this procedure can be found in the original publication (Chiu, March et al. 2004). Primers used for the mutagenesis procedure were designed in Bioedit (Hall 1999) and synthesised (Sigma Aldrich)

Table 14 Biochemical properties of primers used for mutagenesis. Modifications to introduce a stop codon to prevent the transcription of nucleotides encoding for a C-terminal polyhistidine tag were performed using the SLIM mutagenesis procedure. Two sets of primers (F_T, R_s and R_T and R_L described below) were used in the Inverse

PCR reaction (SLIM) where the tailed primers harbour the mutation. The stop codon "taa" located in the tailed (reverse and long) is underlined.

Primer	Nucleotide sequence (5' to 3')	Length	GC content (%)	Melting temperature (In ^{o}C)	Information
Forward long 5'	TAAACGCtaaAAACATCACCATCACCATCAC	31	38	59	Forward primer for site-directed mutagenesis engineered to introduce a premature stop codon
Forward short 5'	CACCATCACCATCACTAAGTGATT	24	41	54	Forward primer for site-directed mutagenesis engineered to introduce a premature stop codon
Reverse long	CACCAAACTGGAAATTAAACGCtaaAAACATC	32	34	58	Shown in reverse complement
Reverse short	GTCAGGGCACCAAACTGGAAATT	23	47	51	Shown in reverse complement

To start the SLIM procedure, KOD Hot Start polymerase (0.5 µL (0.02 units/µL^{-1}) Millipore) was added to a mixture of 0.5 µL (5 ng) plasmid; 2.5 µL (10 mM) Betaine; 2.5 µL (0.2 mM final concentration) dnTPs (Novagen); 1.5 µL (1.5 µM) MgSO4; 2.5 µL 10x buffer KOD buffer (Novagen) and nuclease-free water (final volume 24.5 µL), and the reaction was performed using a mastercycler personal (eppendorf) with a polymerase chain reaction (PCR) programme consisting of: 25 cycles of denaturation 98 ^{o}C, 2 min 30 s; one cycle 98 ^{o}C for 25 seconds, annealing 50 ^{o}C, 35 s and extension 68 ^{o}C, 7 m before the temperature was held at 4 ^{o}C. A 5 µL sample was retained for analysis by agarose gel electrophoresis. PCR samples were purified by a PCR Clean-Up kit (GenElute Sigma-Aldrich) according to manufacturer's guidelines, employing a cooling centrifuge where necessary. Template DNA was removed

from the mixture (50 µL) by the addition of DpnI (1 µL) and 6 µL 4-CORE 10x buffer (New England Biolabs) and the reaction volume made up to 60 µL with nuclease-free water and incubated at 37 ^{o}C for 90 min. Digested product and remaining restriction enzyme were removed from the mixture (GenElute Sigma-Aldrich) according to manufacturer's guidelines with a minor modification in the elution volume (35 µL). 25 µL of buffer (25 mM Tris pH 8 and 150 mM) were added to reach a final volume of 60 µL. The reaction mixture was subject to an annealing step of the generated overhangs using a PCR programme consisting of 2 cycles of denaturation (98 ^{o}C, 3 min), extension (65 ^{o}C, 5 min) and cooling (30 ^{o}C, 1 min) before storage (- 20 ^{o}C). The newly generated constructs, named SMA_ompAstop and LEN_ompAstop were used as integration plasmid for the production of recombinant V_L. 1 µL (100 ng/µL) of each plasmid was transformed into *E.coli* XL1 cells for plasmid stock preparation or *E.coli* C41 cells for protein production. DNA sequencing and assessment of purity was performed as previously described.

5.2.3. Agarose gel electrophoresis

Agarose gels cast at 0.8% were used to confirm that PCR reactions and DNA digestions were successful. Agarose (0.6g Bioline) was dissolved in 75 mL tris-acetate-EDTA (TAE) by heating, supplemented with 1 µg/mL of ethidium bromide, and the gel loaded with 5 µL of PCR product containing 2 µL of loading buffer (6x gel loading dye purple New England Biolabs) and run at 90 v for ~ 60 min. A 1 kb DNA ladder (Hyperladder 1kb).

5.2.4. Transformation of bacterial cells

1 µL (100 ng/µL) of the plasmid; SMApOPIN_ompAstop or LENpOPIN_ompAstop were transformed into ~50 µL of *E. coli* XL1 (DE3) using the available protocol supplied by Addgene (Addgene 2017). Transformation occurred via heat shock at 42 ^{o}C for 45 s followed by, a 5 min 4 ^{o}C incubation step before the transformation mixture were incubated in an outgrowth step for ~ 45 min in SOC medium (Super Optimal broth with Catabolite repression) before plating onto ampicillin (100 µg/mL) containing LB-agar plates to select for plasmid-bearing cells and incubated for 14 hours in a stationary incubator at 37 ^{o}C.

5.2.5. Plasmid DNA purification

A single colony harbouring the pOPINO plasmid was used to inoculate 5 mL of LB broth containing 100 µg/mL ampicillin. Cells were grown for ~ 16 h at 37 ^{o}C with agitation (200 rpm) before pelleting by centrifugation (4000 x *g* for 5 min). The plasmid was extracted and purified from the host cell using the QIAprep Spin Miniprep kit according to the manufacturer's protocols (QIAGEN). Plasma DNA was sequenced (Source bioscience) and stored at - 20 ^{o}C in DNA free water at ~ 500 µg/µL concentrations (determined spectrophotometrically using a NanoDrop ND-1000 Thermo scientific).

5.2.6. Glycerol stock preparation

The LEN or SMA (pOPIN_ompAstop) plasmids were transformed into *E. coli* C41 cells and incubated overnight at 37 ^{o}C. A single colony was used to inoculate 5 mL of LB supplemented with ampicillin (100 µg/mL final

concentrations) and the culture grown for ~16 h at 37 ^{o}C and 180 rpm. The culture was diluted 50:50 (v/v) with sterile 50% v/v glycerol and stored in 2 mL cryo valves for storage at - 80 ^{o}C.

5.2.7. Expression of light chain variable domains SMA and LEN

For protein expression, LB agar plates containing ampicillin (100 µg/mL) were streaked with *E. coli* C41 (DE3) cells transformed with either the LEN or SMA plasmid (pOPIN_ompAstop) and grown overnight at 37 ^{o}C. A single colony was used to inoculate 50 mL of Luria Broth (LB) supplemented with 100 µg/mL ampicillin and grown ~16 h at 37 ^{o}C with agitation (200 rpm). This culture was used inoculate 1 L of LB media at a starting optical density (OD_{600}) of 0.06 - 0.1. The culture was incubated at 30 ^{o}C with shaking (110 rpm) until an OD_{600} ~ 0.75 – 0.85 was achieved. The addition of 1 mM (final concentration) isopropyl β-D-1-thiogalactopyranoside (IPTG), was added to the culture and the culture incubated for no longer than 16 h at 30 ^{o}C for LEN, and 25 ^{o}C for SMA, (SMA was found to aggregate at higher temperatures) with shaking (110 rpm). IPTG, (an analog of allolactose) induces protein expression by releasing the lac repressor from the lac operator (which normally inhibits transcription of the lac operon) (Bell and Lewis 2000). Upon its release two changes occur; 1) T7 polymerase is transcribed and translated and 2) a conformational change allows the polymerase to subsequently bind to the T7 promoter region (located directly upstream of the target sequence as shown later in Figure 49). Binding of this polymerase leads to the subsequent transcription and translation of DNA located directly downstream of the promoter. Within the pOPINO expression system used in this study, this includes the ompA leader sequence

and our protein of interest (the immunoglobulin light chain variable domains SMA or LEN). After IPTG induction cells were harvested by centrifugation (3360 x *g* for 10 min at 4 ^{o}C). For NMR experiments, the 50 mL overnight culture was centrifuged (3360 x *g*) and the pellet resuspended in 1 L of M9 minimal media (90 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.56 mM NaCl and 18.7mM NH_4Cl (final concentrations) pH 7.4 substituted with (0.2% v/v) ^{15}N glucose for the remainder of the growth and induction period.

5.2.8. Osmotic shock treatment

Osmotic shock was used to liberate the recombinant V_Ls from the periplasmic space of the host *E.coli* cell. Cell pellets were resuspended in a hypotonic osmotic shock solution (TES buffer) comprising 200 mM Tris, 5 mM EDTA, and 200 g w/v sucrose pH 8. Bacterial pellets were resuspended in 100 mL TES buffer (at 4 ^{o}C) and incubated on ice for 30 min, with inversion at intervals of ~ 5 min to prevent sedimentation. Pellets were centrifuged again, at a higher centrifugal speed of 8000 x *g* for 10 min at 4 ^{o}C, to sediment the pellet now in sucrose. It is important to note here that lower speed does not result in a firm pellet. The supernatant was discarded and the pellet rapidly resuspended in MilliQ (35 mL/L of initial culture, 4 ^{o}C) supplemented with one protease inhibitor tablet (cOmplete, Mini Protease Inhibitor Cocktail, ROCHE) acting as a hypertonic solution. Solutions were again incubated for 30 min on ice before centrifugation at 23, 000 x *g*, 30 min at 4 ^{o}C to remove cellular debris. Periplasmic proteins were found to be released from the periplasmic space in this final stage.

5.2.9. Dialysis of SMA and LEN and isoelectric precipitation

Hypertonic fractions containing LEN, or SMA were loaded into a 3500 Molecular weight cut-off (MWCO) presoaked dialysis membrane (Pierce) and the 35 mL dialysed against 3 L of 10 mM sodium acetate pH 5.0 at 4 °C for ~ 36 h with 3 buffer changes. This resulted in the precipitation of a large proportion of host cell contaminants (assessed by SDS-PAGE) which were removed by centrifugation at 8000 x *g* for 15 min at 4 °C. The supernatant containing the V_L of interest was used for further purification.

5.2.10. Purification of light chain variable domains

A combination of cation exchange and size exclusion was used as a chromatographic step to purify the V_Ls to high levels of homogeneity. For each V_L, supernatants following dialysis and centrifugation steps were loaded directly onto dedicated 5 mL HiTrap SPFF columns (GE Healthcare) mounted to an ÄKTA purifier chromatography system (GE Healthcare) at a flow rate of 0.75 mL per minute. A post load wash consisting of three column volumes (CV) of 10 mM acetate buffer, pH 5.0 was made before LEN was eluted using a 0-250 mM NaCl gradient over 120 mL. For SMA, which contained more contaminants than LEN, the protein was eluted with 5 CV's of 10 mM Tris pH 8. At this point, the purity degree of LEN fractions was deemed to be 95 % pure, as judged by SDS-PAGE (Figure 55) and reverse-phase high-performance liquid chromatography (RP-HPLC). At this stage SMA containing fractions displayed minor higher molecular contaminants which were not removed by microfiltration. For further purification of SMA, fractions were pooled and concentrated down to a volume of ~ 50 µL (from 4 L growth

culture) using a 0.5 mL 10 kDa (MWCO) filter (Millipore) and applied to a HiLoad 16/60 Superdex 75 prep grade (GE Healthcare life sciences) size exclusion column pre-equilibrated with 20 mM Tris-HCl pH 7.5, 150 mM NaCl. Samples were injected through a 100 µL loop, flushed for 3 sample loop volumes and the chromatographic profile recorded at a flow rate of 1 mL/min. Proteins were eluted isocratically (0% B, over 1 column volume) and the purity degree evaluated by SDS-PAGE (Figure 55) and RP-HPLC. Pure proteins (SMA and LEN) were concentrated and filtered into phosphate buffered saline (PBS) using a 10 kDa MCWO filters. Protein concentration was quantified spectrophotometrically (UV_{280nm}) using the theoretical protein extinction coefficients ε of 0.1% of 1.71 and 1.82 for SMA and LEN respectively which were adopted from publications that first characterised these proteins (Raffen, Dieckman et al. 1999, Kim, Cape et al. 2001, Khurana, Souillac et al. 2003, Qin, Hu et al. 2007). Samples were stored at > 3 mg/mL at 4 °C where they showed no signs of degradation over a 12-month period (assessed by UV_{280nm}, no visible precipitation, no degraded product on SDS-PAGE). Typical yields from 1 L of culture were ~ 10 mg for SMA and ~ 50 mg for LEN.

5.2.11. Gel electrophoresis, RP-HPLC and mass spectrometry

The expression and purification of LEN and SMA were analysed by SDS-PAGE using 12% Tris-Tricine gels in a Bio-Rad gel electrophoresis system. Samples were solubilised in 4x Laemmli sample buffer (Laemmli 1970) for 5 min at 90 ^{o}C prior to loading. The Pierce unstained protein MW marker (Life technologies) was loaded as molecular mass markers in electrophoresis studies. Gels were run for 60 min at 165 V, stained with Coomassie Brilliant Blue G-250 0.25% (w/v) and destained with H_2O, methanol, and acetic acid in

a ratio of 45/45/10 (v/v/v). Identification of the intact recombinant protein and the removal of the N-terminal ompA tag was confirmed by mass spectrometry. Protein samples were dialysed into 50 mM ammonium bicarbonate using HPLC grade water and spectra acquired using an ESI-Q-TOF micro spectrometer (Micromass).

5.2.12. Circular dichroism measurements

CD was performed on a JASCO J1100 spectropolarimeter (JASCO UK, Ltd). Far UV-CD spectra (250-180nm) were acquired using a 0.2 mm cuvette, at 4 ^{o}C using 10 μM proteins in 5 mM phosphate buffer, pH 7.5. Secondary structure content values were acquired using BeStSel (Micsonai, Wien et al. 2015).

5.2.13. Size Exclusion Chromatography with multi-angle Light Scattering

Immunoglobulin variable light chains typically exist as homodimers. The dimerisation ability of the recombinant proteins were characterised by Size Exclusion Chromatography - Multi-Angle Laser Light Scattering. Purified SMA and LEN at concentrations of 1 mg/mL were applied directly to a HiLoad 16/60 Superdex 75 attached to ÄKTA pure fast protein liquid chromatography (FPLC) system equilibrated in 10 mM Tris-HCl pH 7.5, 150 mM NaCl. A DAWN 8+ and optilab T-rex Helios 8 (WYATT) scattering detector was directed downstream flowpath of the SEC column. As a control lysozyme was run under identical conditions. Chromatograms were acquired at a flow rate of

0.75 mL/min at 25 °C. Data was analysed using ASTRA v6.1 software (WYATT).

5.2.14. RP-HPLC of SMA and LEN

Analytical RP-HPLC was used as a technique for impurity profiling of samples containing recombinant variable domains. For both proteins, 10 µL samples were centrifuged at 10,000 *x g* for 5 minutes, and applied to a Phenomenex Aeris Widepore C4 column (150 x 2.1 mm) equilibrated in 0.08% Trifluoroacetic acid (TFA) attached to a Dionex ICS3000 HPLC system. Proteins were eluted with a linear gradient of 5-65% acetonitrile in 0.08% TFA: 0-40% over 60 min.

5.2.15. Two-Dimensional (^{1}H ^{15}N) NMR experiments

To determine that the recombinant V_Ls had correctly folded, two-dimensional heteronuclear correlation experiments with ^{1}H ^{15}N enriched LEN were performed. Spectra of 180 µM LEN was acquired at 30 ^{o}C in 20 mM sodium phosphate, 100 mM NaCl and 10% (v/v) D_2O for lock using a Bruker AVANCE III 600 MHz, spectrometer (Bruker BioSpin) equipped with a cryogenically cooled 5 mm ^{1}H [^{15}N ^{13}C] probe. Acquired spectra were processed using Topspin 3.1.7 (Bruker) and analysed using the Collaborative Computational Project for NMR ccpNMR software v2.2.2. Backbone resonances of LEN were assigned by transferring the Biological Magnetic Resonance (BMRB) deposited assignment of LEN (BMRB ID: 16463 (Mukherjee, Pondaven et al. 2009)

5.3. Results and Discussion

5.3.1. Plasmid generation

A system for the production of the recombinant V_Ls SMA and LEN was established previously (Stevens, Raffen et al. 1995), however, this study has been difficult to replicate in recent years due to the use of out-of-date cell lines and specialised equipment. Here, we employ periplasmic expression of our POI using the ompA leader peptide fused to the amino termini (MKKTAIAIAVALAGFATVAQA) of each individual V_L, where the protein is targeted by the SEC translocase pathway to the oxidising compartment of the periplasmic space (Figure 48). It is here where the ompA signal sequence is cleaved by a signal peptidase which leaves an unmodified amino terminus. This strategy has several attractions; the prokaryotic periplasm contains lower quantities of endogenous proteases and contaminating bacterial proteins (Park and Lee 1998) which negates the use of many initial purification steps. This compartment also contains the foldases; disulfide oxidoreductase (DsbA) and disulfide isomerase (DsbC) that are localised to the periplasmic space (Dartigalongue, Nikaido et al. 2000, de Marco 2009, Nozach, Fruchart-Gaillard et al. 2013, Goemans, Denoncin et al. 2014) and assist in correct folding and disulfide bond formation.

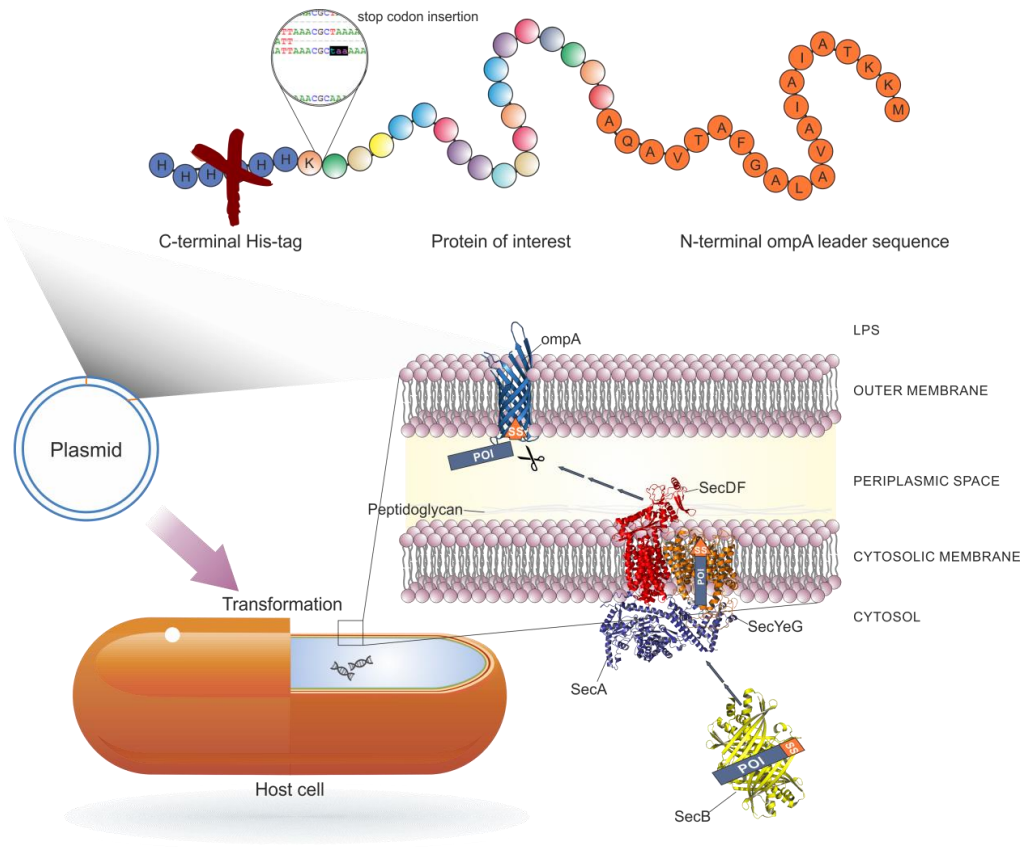


Figure 48 Simplified representation of the post-translational Sec translocase and ompA bacterial export mechanism of fusion protein. The protein of interest (POI), is expressed fused to an authentic N-terminal signal sequence (SS orange) that is required for the translocation of the protein to the periplasmic space. Following synthesis, the nascent chain is transported by cytoplasmic chaperone SecB (yellow PDB ID; 1QYN) chain to SecA ATPase (blue PDB ID: 3DIN) that is directly bound to the SecYEG (orange PDB ID: 3DL8) translocation channel which spans the membrane and facilitates the successful transport across the cytosolic membrane. The membrane bound SecDF (red PDB ID; 3AQP) protein complex orchestrates the process, and utilising transmembrane proton-motive force (PMF) facilitates the final stages of translocation. Now in the periplasmic compartment, the OmpA signal sequence associates with the membrane spanning, outer membrane protein A (navy PDB ID: 1BXW) and is cleaved by a signal peptidase before its export into the growth media. The protein of interest, now absent of its leader sequence is left in the periplasm where it can be isolated experimentally by osmotic shock extraction methods.

We employed the pOPINO plasmid from OPPF (linearised plasmid shown in Figure **49**), which encompasses an inducible T7 promoter, ampicillin resistance cassette and an N-terminal ompA leader sequence allowing for diffusion of the protein into the periplasmic space. However, in early experiments we found that the His tag seemed to prevent the successful translocation of the POI into the periplasmic space.

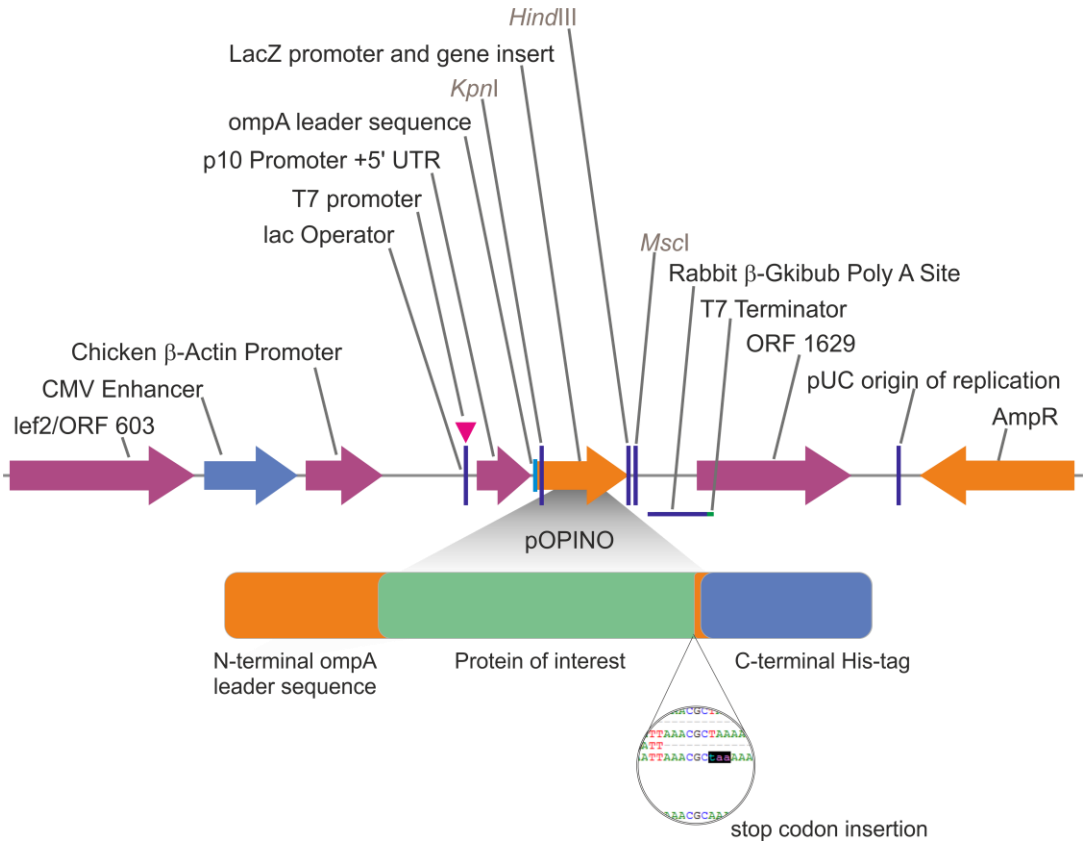


Figure 49 Schematic representation of the pOPINO (Linearised). In this construct, the C-terminal polyhistidine-tag (blue) was removed by insertion of the stop codon “taa” (circular insert) into the recombinant DNA sequence of each construct (shown in expanded region of promoter and gene insert. Other key features of the plasmid are highlighted.

We therefore introduced a premature “taa” stop codon directly upstream of the nucleotides encoding the his-fusion tag using the SLIM mutagenesis procedure (Chiu, March et al. 2004). The success of the procedure was determined by successful amplification of the template DNA, analysed by agarose gel electrophoresis (Figure 50) and sequencing. Bands migrating to ~ 5500 base show successful amplification of the PCR product. Sequencing of the generated plasmids confirmed the successful incorporation of a “taa” stop codon.

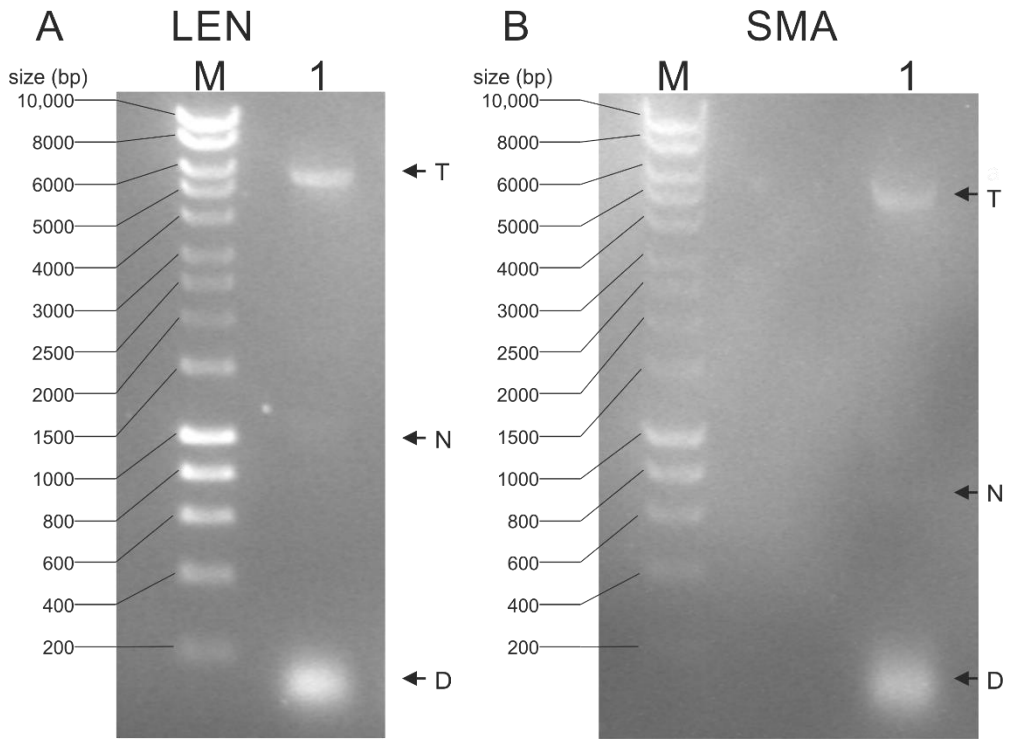


Figure 50 Agarose gel electrophoresis analysis of SLIM PCR products. The 1% agarose gels were cast with in the presence of ethidium bromide and photographed. Lanes M – Quick-load 1 kb ladder; 1 – linearised construct. The labelled arrows denote Template (T) non-specific PCR product (N) and the dye front (D).

5.3.2. Recombinant expression of SMA and LEN

This modified plasmid was then used to transform *E.coli* C41 (DE3) cells and protein expressed with IPTG induction until an OD 600 of ~0.8 was reached, with shaking at 110 rpm at 30 ^{o}C for LEN and 25 ^{o}C for SMA (reduced temperature to prevent aggregation). Osmotic shock using sucrose was used to liberate proteins from the periplasmic space and SDS-PAGE analysis showed detectable levels of soluble protein expression, with a monomeric band of approximately ~13 kDa corresponding to the calculated theoretical molecular weights (MW) of LEN and SMA (Table 13) (Figure 51 A and B). The liberated protein mixture was then dialysed into 10 mM sodium acetate buffer pH 5, with multiple changes over a 36 hour period to remove sugar and lower pH. This resulted in the precipitation of a large amount of host cell contaminants (Figure 51 C).

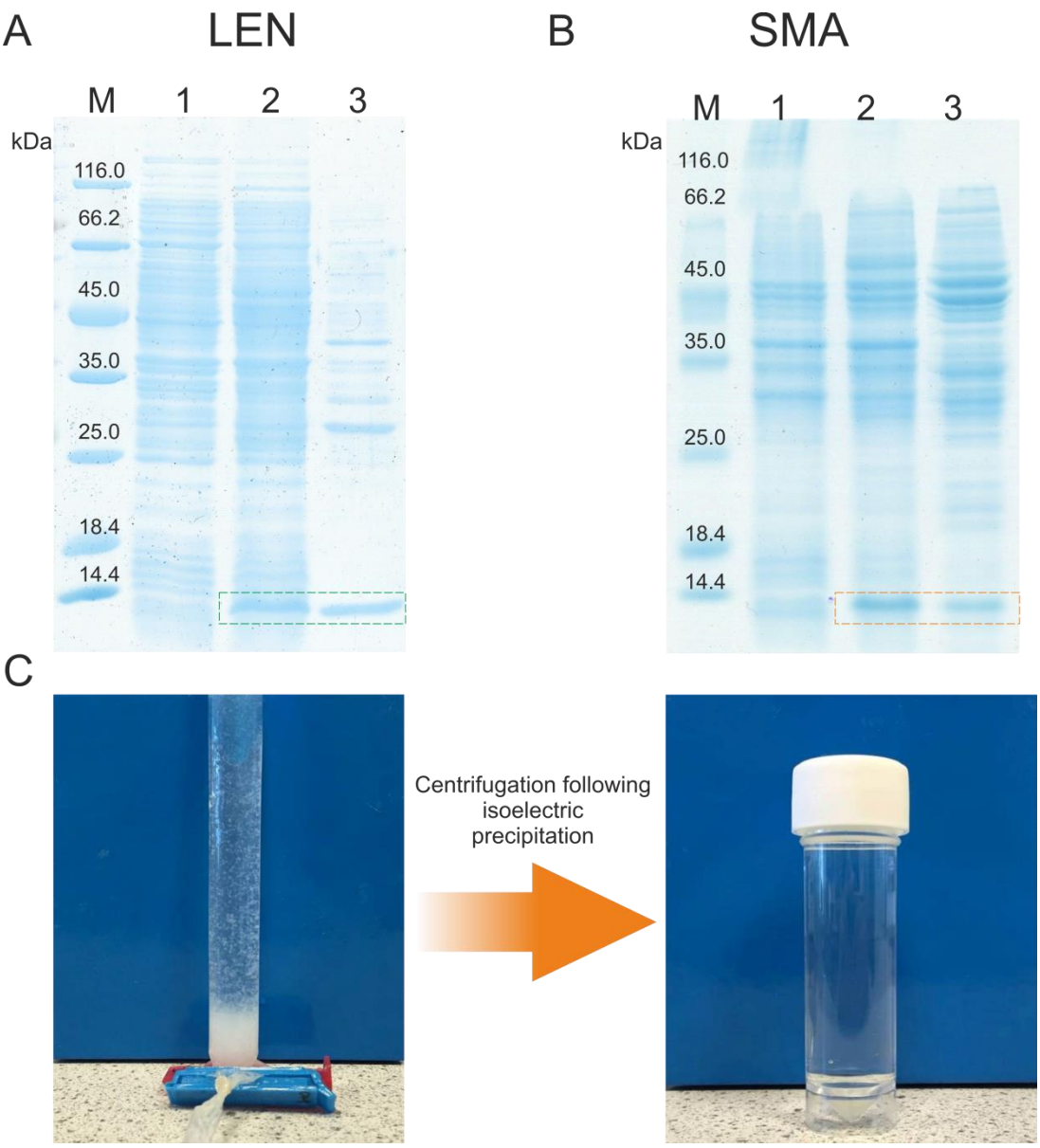


Figure 51 SDS-PAGE analysis of the expression and isolated of LEN and SMA. Both V_Ls were expressed and isolated from the periplasmic space of the host cell using osmotic shock. The success of the procedure was assessed by SDS-PAGE (Panels A and B). The gel lanes are marked as follows: Lane M Pierce™ Unstained Protein MW Marker; Lane 1- Uninduced total bacterial proteins; Lane 2 – IPTG Induced total bacterial protein extract; Lane 3 - the hypertonic solution. The target proteins LEN and SMA are indicated (dashed box). (C) A number of host cell contaminants were then removed using an isoelectric precipitation step.

Following removal of these precipitated contaminants cation exchange chromatography was used, purifying recombinant protein to high levels of homogeneity by a single-step process. A linear salt gradient in the mobile phase eluted LEN as a single peak with high purity (Figure 52) where host organism contaminants remained bound.

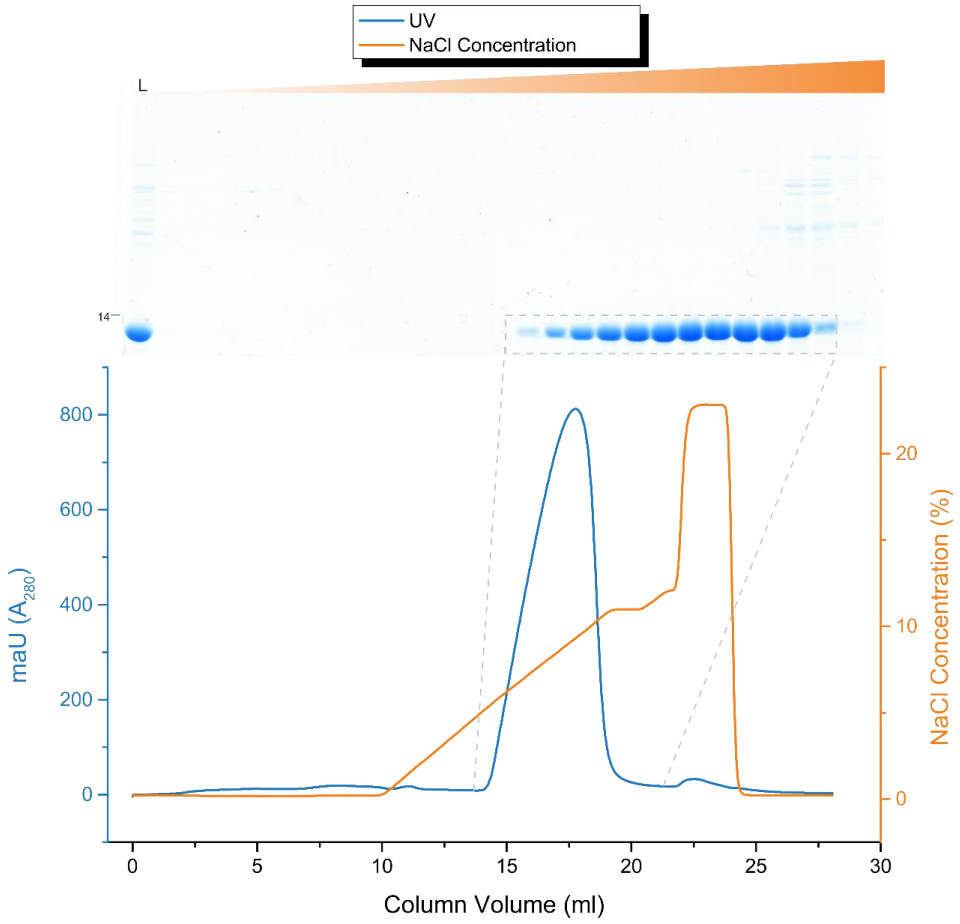


Figure 52 Representative chromatogram and SDS-PAGE analysis of the purification of LEN by ion exchange chromatography using a gradient elution. The LEN in sodium acetate pH 5.0 was applied to a 5 ml HiTrap SP HP column (GE healthcare life sciences) at a flow rate of 0.75 ml/min. A post load wash of 5 CV sodium acetate pH 5.0 was followed by the elution of recombinant V_L using a salt gradient of 0-250 mM in a volume of 28 ml over a run time of 30 min. The purity degree of fractions (2 ml) corresponding to the elution peak were assessed by SDS-PAGE. Lane "L" shows load fraction, the remaining lanes are fractions 1-30 fractions that correspond to the elution profile. Target proteins were isolated as a single band which corresponds to the increased UV trace shown on the ÄKTA FPLC chromatogram.

5.3.3. Recombinant expression of SMA

Expression in C41 cells, following a 16 h IPTG induction at an OD 600 of ~0.8, with 110 rpm shaking at 25 ^{o}C produced soluble SMA in the periplasmic space liberated by osmotic shock as described for LEN above. SMA required expression at a lower temperature than LEN to prevent aggregation (mentioned earlier). In addition, expression exceeding 16 h resulted in the formation of SMA in inclusion bodies or SDS-resistant oligomers (experimentally verified by western blot). As for LEN, isoelectric precipitation removed contaminating proteins prior to cation exchange chromatography. SMA was eluted from the S column using Tris pH 8 (Figure 53). Analysis of all fractions by SDS-PAGE confirmed the presence of SMA and host organism contaminants in the eluates, thus requiring further purification steps.

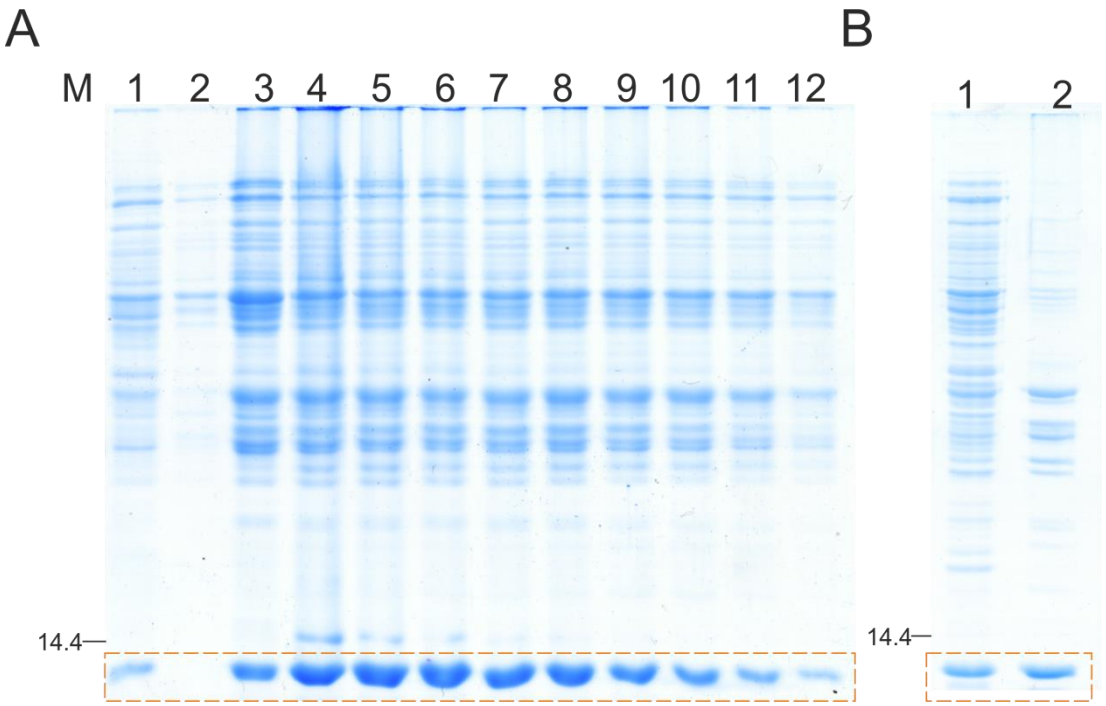


Figure 53 SDS-PAGE analysis of the purification of SMA by ion exchange chromatography using a pH step elution. SMA that had been dialysed into sodium acetate pH 5.0 was applied to a 5 mL HiTrap SP HP column (GE healthcare life sciences) at a flow rate of 0.75 mL/min. No loss of protein on the cation exchanger was detected (Panel A, Lane 2, flow through). A post-load column was followed by elution of SMA using 10 mM Tris pH 8 (Panel A, Lanes 3-12). Panel B shows before (Lane 1) and after (Lane 2) further dialysis into sodium acetate buffer pH 5.0 that resulted in the removal of a large number of host cell contaminants. Target protein is highlighted (orange box). Estimated position for the 14.4 kDa marker based on gels shown in Figure 50 is shown.

Prior to gel filtration, fractions containing SMA were again dialysed into sodium acetate pH 5.0 which resulted in a further crash of contaminants (Figure 53) before passing through a 0.22-micron syringe filter to remove any aggregates, concentrated using a 10 kDa MWCO with multi exchanges of buffer (to match column equilibration buffer) before being applied to a size exclusion column equilibrated in 10 mM Tris-HCl pH 7.5, 150 mM NaCl. Figure 54 shows the successful separation of the contaminants to leave pure SMA.

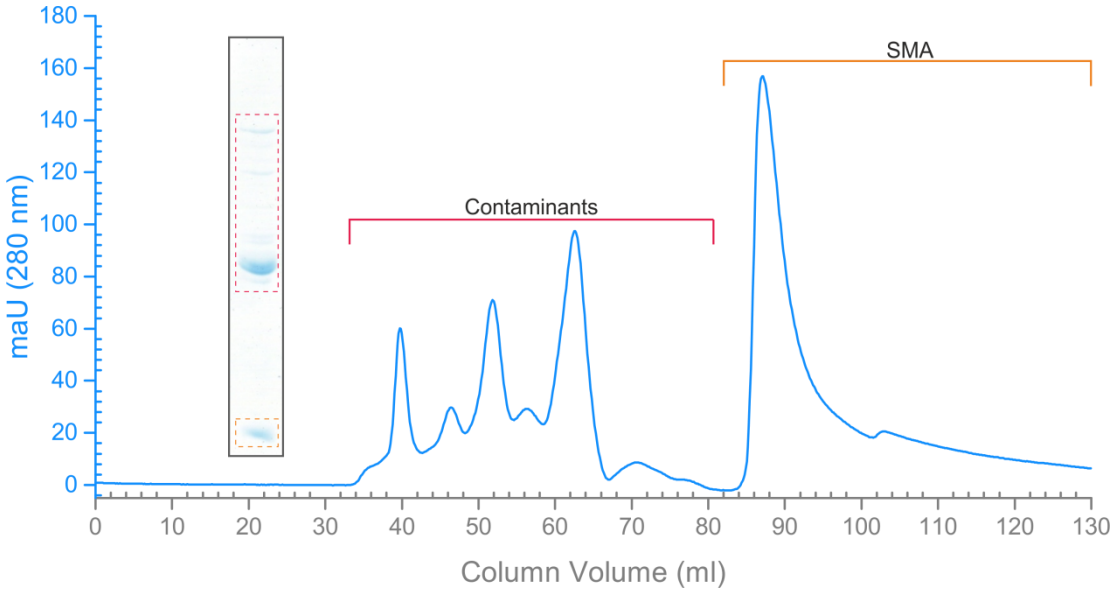


Figure 54 Final purification strategy for SMA isolated from C41 cells. SDS-PAGE analysis of SMA following ion exchange chromatography revealed upon concentration, higher molecular weight contaminants. SEC chromatography was used as a polishing step to separate the recombinant protein from host cell contaminants. Protein fractions were dialysed into 10 mM Tris-HCl pH 7.5, 150 mM NaCl, passed through a 0.22 micron filter, then concentrated using a 10 kDa MWCO (millipore) before application to a prepacked Superdex 16/60 superdex 75 (GE healthcare). As shown, the protein elutes at ~ 82ml and displays the typical asymmetric peak characteristic of the V_Ls which indicates the simultaneous dissociation and binding of the protein and the column.

5.3.4. Confirmation of recombinant Immunoglobulin light chain identity

The end purity degree of LEN and SMA containing fractions following all chromatographic procedures were assessed by SDS-PAGE and RP-HPLC, where the chromatogram reveals a single peak for each protein (LEN and SMA) indicating high levels of purity and no other visible contaminants.

The isolated proteins were confirmed to be products of recombinant expression, by mass spectrometric analysis (Figure 55).

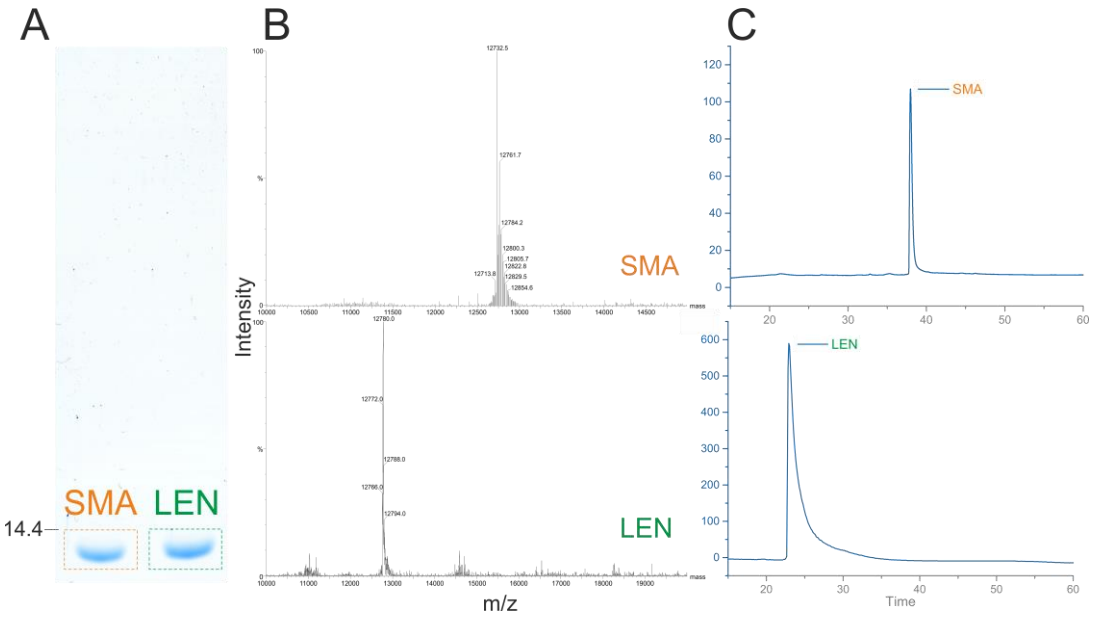


Figure 55 SDS-PAGE analysis and mass spectra of LEN and SMA. V_Ls that had been expressed and undergone purification using ion exchange chromatography and gel filtration to isolate it from contaminating *E.coli* proteins were dialysed into 5 mM sodium phosphate pH 7.5 before analysis. (A) Purity degree of SMA and LEN was assessed by SDS-PAGE analysis and deemed >95% pure, where the proteins can be seen migrating as a single band below the estimated position for the 14.4 kDa marker based on gels shown in Figure 50. (B) The molecular ion peaks with the highest intensities correspond to the theoretical molecular weight of (^{15}N) monomeric protein LEN (m/z 12780) and monomeric SMA. Both spectra confirm that the ompA leading sequence has been successfully removed following translocation into the periplasmic space. Predictions were performed by peptidemass (Wilkins, Lindskog et al. 1997). All mass spectrometry was performed by Dr Mark Wilkinson on a micromass Q-Tof micro (waters) in positive ion mode. (C) Reverse-phase HPLC profile of purified LEN and SMA. UV absorption was measured at 280 nm. Purity degree for each protein was estimated to be >98% and continued for use in fluorescent labelling and cell internalisation assays documented in chapter 6. The differences in elution time are a combination of lower protein concentration and column availability.

5.3.5. Analysis of protein secondary structure

Affirmation of correct protein folding for LEN was performed by NMR spectroscopy. LEN was expressed in minimal media producing isotopically labelled protein and 2D ^{1}H ^{15}N Heteronuclear single quantum coherence spectroscopy (HSQC) spectra acquired under conditions matching those described previously (Mukherjee, Pondaven et al. 2009). The resonance frequencies provided by amides of the protein backbone were compared to the assigned chemical shift of LEN performed by our collaborators (BioMagResBank accession: 16463) and confirms that the protein was correctly folded (Figure 56).

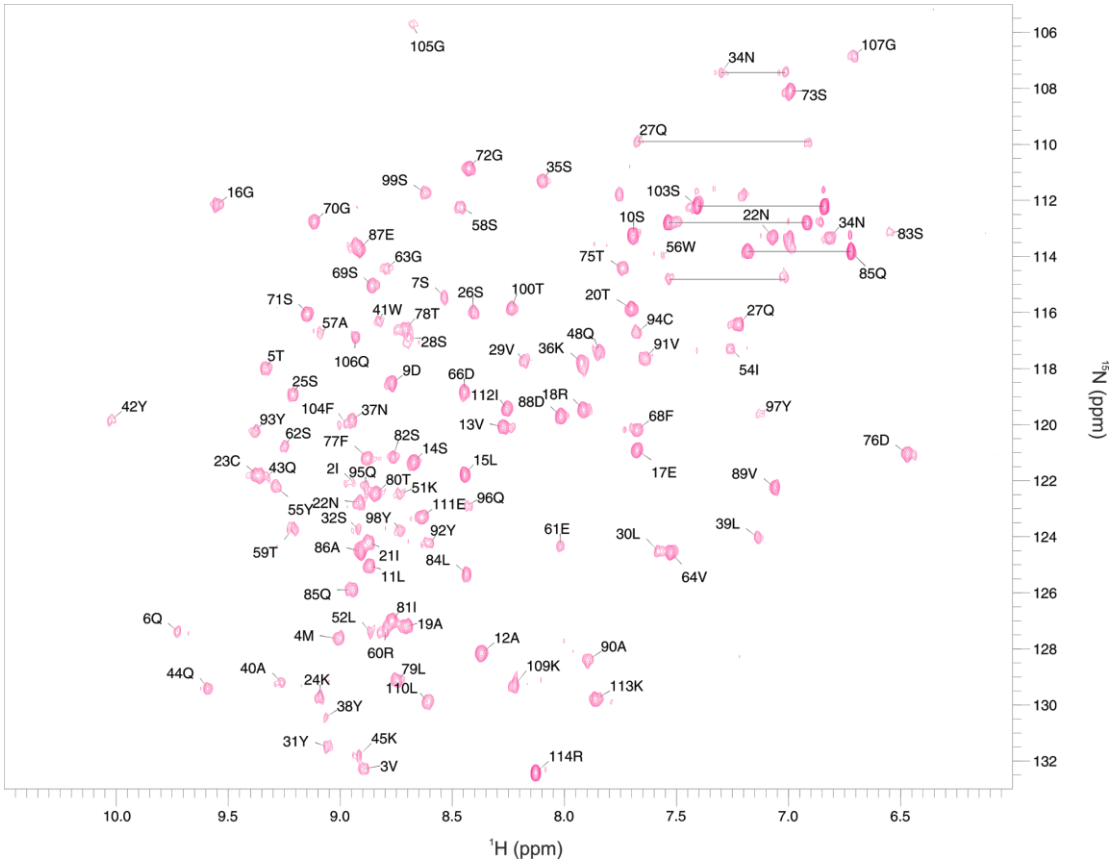


Figure 56 ^{1}H-^{15}NHSQC spectra of LEN. Spectra of the recombinant LEN (180 µM) obtained at 600 MHz, 30 ^{o}C in 20 mM sodium phosphate, 100 mM NaCl, and 10% (v/v) D_2O. Backbone chemical shifts are assigned using the resonances of deposition 16463 within the BioMagResBank.

Matching the experimental conditions to the deposited BMRB allowed for all (114 amino acid) residues of LEN backbone resonances to be able to be assigned unambiguously. Unfortunately, full assignment of SMA is not possible due to signal overlap, described by our collaborator who attempted assignment previously in unpublished work (Pondaven 2012). As the spectra indicate the protein is correctly folded, a sample of LEN from the same protein batch was used as a reference for SMA in circular dichroism experiments, used to determined secondary structure (Figure 57).

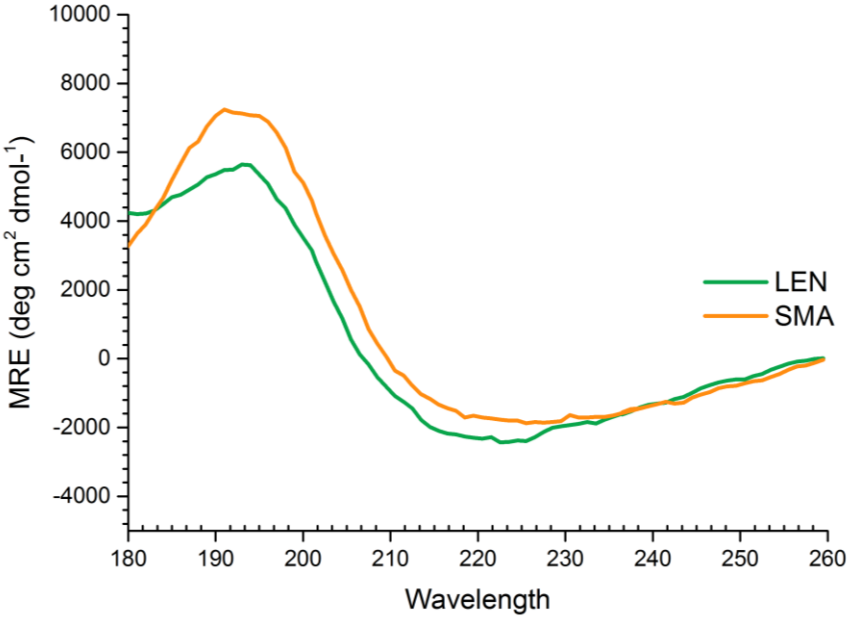


Figure 57 Far UV CD spectra of V_L confirms correctly folded recombinant protein, displaying native β-sheet structure. CD spectra obtained for LEN (green) and SMA (orange) indicate a minimum around 220 nm. BEST-SEL analysis of CD spectra indicates the protein to possess an Ig like fold through its fold recognition software (Micsonai, Wien et al. 2015). CD experiments were performed with 20 μM protein in 5 mM sodium phosphate (pH 7.5) at 4°C. Units are presented as mean residue ellipticity.

5.3.6. Size exclusion chromatography with multi-angle light scattering

To experimentally assess the ability of recombinant LEN and SMA to form dimers, their oligomerisation state was assessed by SEC-MALLS (Figure 58). Both proteins had near identical retention times and had calculated molecular weights of 24.32 kDa (LEN) and 25.46 kDa (SMA) that correspond well to the theoretical calculated molecular weights of (25.28 kDa and 25.46) respectively. This confirms that each recombinant protein retained their dimerisation capability.

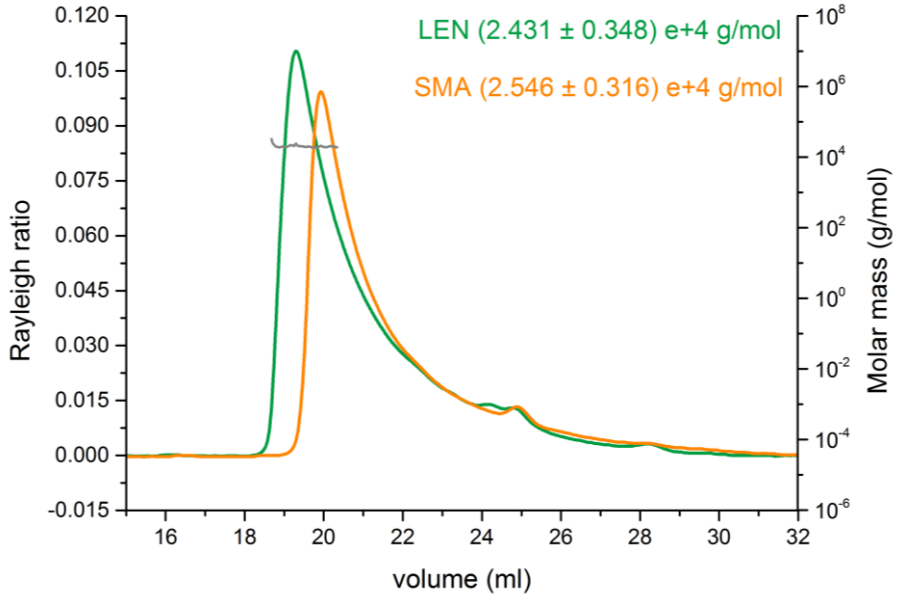


Figure 58 SEC-MALS chromatogram confirming the dimerisation state of LEN and SMA Both V_Ls elute as single asymmetrical peaks at a volume of ~18.5 mL and ~19.5 mL on a Superdex gel filtration column (GE Healthcare) showing that a single species is present. The unique shape of the peak is consistent with other V_Ls characterised by gel filtration based methods (Fred J. Stevens 1980, Stevens and Schiffer 1981, Raffen, Stevens et al. 1998). Based on the calculated molecular weight of 24.31 kDa (LEN) and 25.46 kDa (SMA) both proteins are homodimers, corresponding to the theoretical molecular weights of dimers. Concentration of each protein was 1 mg/ml.

5.3.7. Hairpin molecular A10 binding to SMA and LEN

Having determined the fold and purity degree of each recombinant V_L, samples were subject to a number of binding experiments using the β-wrapin AS10. Where amyloidogenic proteins display the high propensity to adopt a quaternary structure rich in β-sheet, the AS10 molecule is designed to arrest the aggregation cascade by binding to β-hairpins, the structural motif commonly present in amyloid structures (Shaykhalishahi, Mirecka et al. 2015). The β-wrapin AS10 has previously demonstrated the ability to bind to specific regions that display β-sheet propensity within monomeric islet amyloid polypeptide (type 2 diabetes) amyloid b peptide (Alzhiemer's disease) and alpha synuclein, the 140 kDa intrinsically disordered protein implicated in Parkinson's disease and abolish their aggregation (Shaykhalishahi, Mirecka et al. 2015). For demonstration purposes, a structure showing the binding of AS69 (a β-wrapin that differs from AS10 by just one mutation Le34Val (Shaykhalishahi 2015) to a region of alpha-synuclein (residues 37-54) is shown (Figure 59).

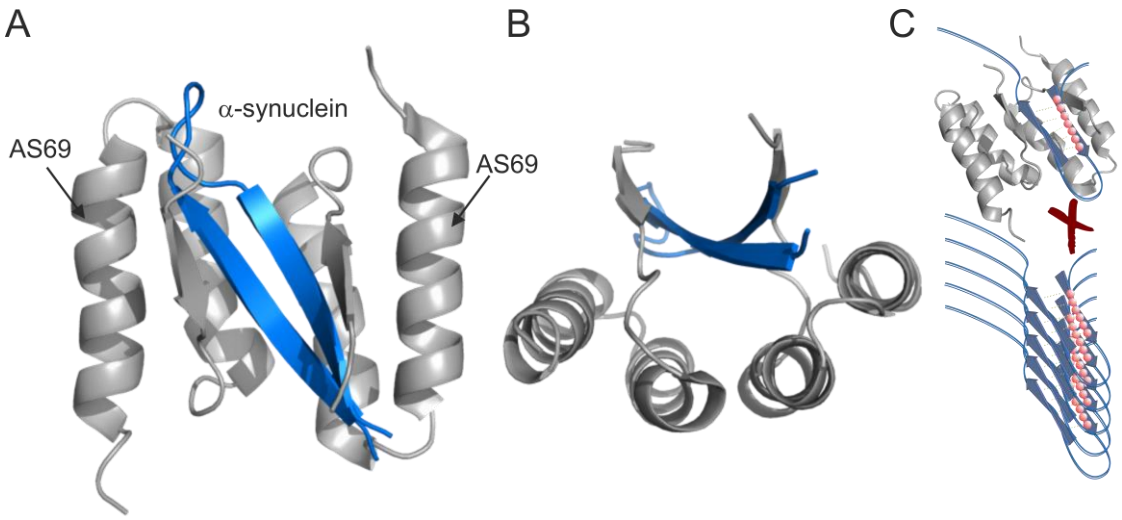


Figure 59 shows β-wrapin AS69 in complex with a region of α-syn that adopts a β-turn. A) Shows NMR structure (PDB ID: 4BXL) of α-syn region in complex with two AS69 subunits. B) Shows structure with 90 ° rotation on the x-axis. C) Shows cartoon representation of how the binding of AS69 may sequester the free β-hairpin monomer, thus preventing the association of free monomers required for elongation and extension of the amyloid fibre. Figure was based on the publication (Mirecka, Shaykhalishahi et al. 2014) with modifications.

Here, we investigate the ability of AS10 to bind the light chain variable domains of this study. Experiments were performed by Wolfgang Hoyer (University of Dusseldorf). The results of the preliminary study are shown below in the HSQC-NMR spectra (Figure 60) where unfortunately, in this experiment there was no evidence of binding. Binding would be demonstrated by a chemical shift where the LEN control (blue) would demonstrate peaks in different parts of the spectrum to the peaks (red) that were incubated with AS10.

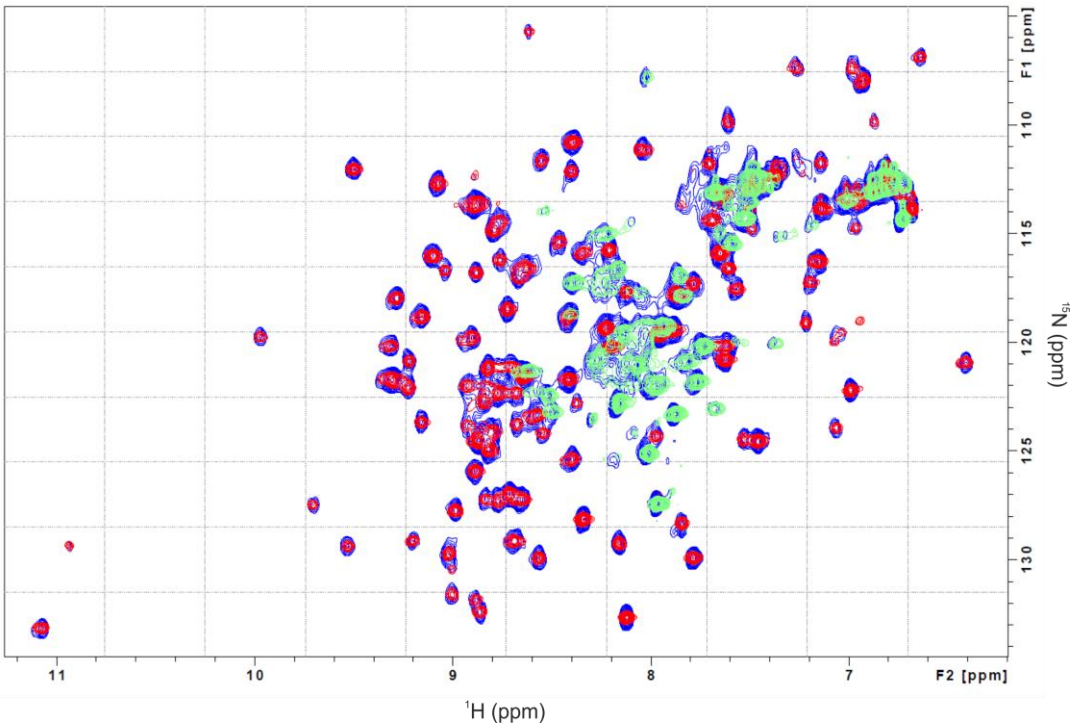


Figure 60 ^{1}H-^{15}N HSQC spectra of SMA. Spectra was obtained at 800 MHz, 30 ^{o}C in 20 mM sodium phosphate, 100 mM NaCl, and 10% (v/v) D_2O. Overlayed spectra are as follows: LEN no inhibitor (control) is shown in blue, LEN plus inhibitor is shown in red. Inhibitor is shown in green.

5.3.8. Main findings and summary

Summary

We present here a periplasmic expression method, with purification employing a controlled osmotic shock procedure free of lysozyme or other chemical lysis methods that disrupts only the periplasmic space leaving the cytoplasmic space undisturbed. Furthermore, the addition of an isoelectric precipitation step dramatically reduces the level of contaminating host cell proteins reducing chromatographic steps, required for large-scale production and improving purity and yield. Purity was deemed > 95% as confirmed by SDS-PAGE, and reverse phase HPLC (documented in chapter 6) with correct protein

sequences, and full cleavage of the ompA leader sequence confirmed using mass spectrometry. Evaluated protein yields were ~ 50 mg/L for LEN and ~ 10 mg/L for SMA with little variations between preparations which improve significantly on previous work for these proteins. In addition, we perform secondary structure analysis by circular dichroism spectroscopy, and use SEC-MALLS to confirm the ability of recombinant products to dimerise. We believe the more detailed methods described here will be applicable to other V_L domains that may not be amenable to refolding techniques previously proposed (Rognoni, Lavatelli et al. 2013) to produce high quality protein in sufficient quantities for functional and structural analysis when a protein source is no longer available from the AL diagnosed patient. Sources of SMA and LEN will be continued in further experiments for use with collaborators (including Wolfgang Hoyer at the University of Dusseldorf and UCB slough).

Highlights

In this chapter, the variable domains SMA and LEN have been successfully expressed and purified. Both SMA and LEN were successfully isolated from the periplasmic space of *E.coli* expression hosts and shown to be correctly refolded by CD and NMR. SEC-MALLS experiments have revealed that the recombinant protein possesses the ability to form dimers. SDS-PAGE analysis of each recombinant protein has revealed that the proteins were purified to high levels and are now suitable for further biochemical analysis.

6. Cellular internalisation of light chain variable domains

6.1. Introduction

The accumulation of amyloid plaques is a pathological hallmark of many conformational disorders. What remains part of an ongoing investigation is whether the initial site of fibrillation takes place intracellularly or extracellularly (Meng, Fink et al. 2008). Recent publications have highlighted the presence of intracellular fibrillar deposits in Alzheimer's and Parkinson's disease (Glabe 2001, Ross and Poirier 2004, Bosco, Fowler et al. 2006, LaFerla, Green et al. 2007, Friedrich, Tepper et al. 2010, Jucker and Walker 2013) providing evidence to support an intracellular hypothesis. For AL amyloidosis, detection of upregulated p38 mitogen-activated protein kinase (MAPK) levels found in patient extracts and animal models which is indicative of oxidative stress and cellular apoptosis has been the initiating factor for a small number of studies aiming to shed light on the ability, and consequences of V_L internalisation (Shi, Guan et al. 2010, Mishra, Guan et al. 2013, Spencer and Engelhardt 2014, Grogan, Dispenzieri et al. 2017). In a selection of studies using primary rat cardiac fibroblasts, HL-1 mouse cardiomyocytes and AC10 and AC16 human ventricular cardiomyocytes both soluble V_Ls, full length LC and fibrillar forms of V_L have been confirmed to internalise under a proposed mechanism of pinocytosis (Levinson, Olatoye et al. 2013, McWilliams-Koeppen, Foster et al. 2015). Tissue specific uptake has also been recognised where, renal mesangial cells; appear to facilitate the active transport of V_Ls using a receptor based mechanism (Teng, Russell et al. 2004, Levinson, Olatoye et al. 2013). It

is an interesting concept that, depending on the mechanism of internalisation, the V_L may be subject to acidic components of the endosome and lysosome (Hu, Dammer et al. 2015). While low pH has been shown to accelerate light chain aggregation (Ramirez-Alvarado 2012, Blancas-Mejía and Ramirez-Alvarado 2013), it is yet unclear if the conditions present in these compartments promote the partial unfolding required for aggregation to occur. Such mechanisms for other amyloidogenic precursor proteins however, has been noted previously (Su and Chang 2001). A summary of these mechanisms is shown below (Figure 61).

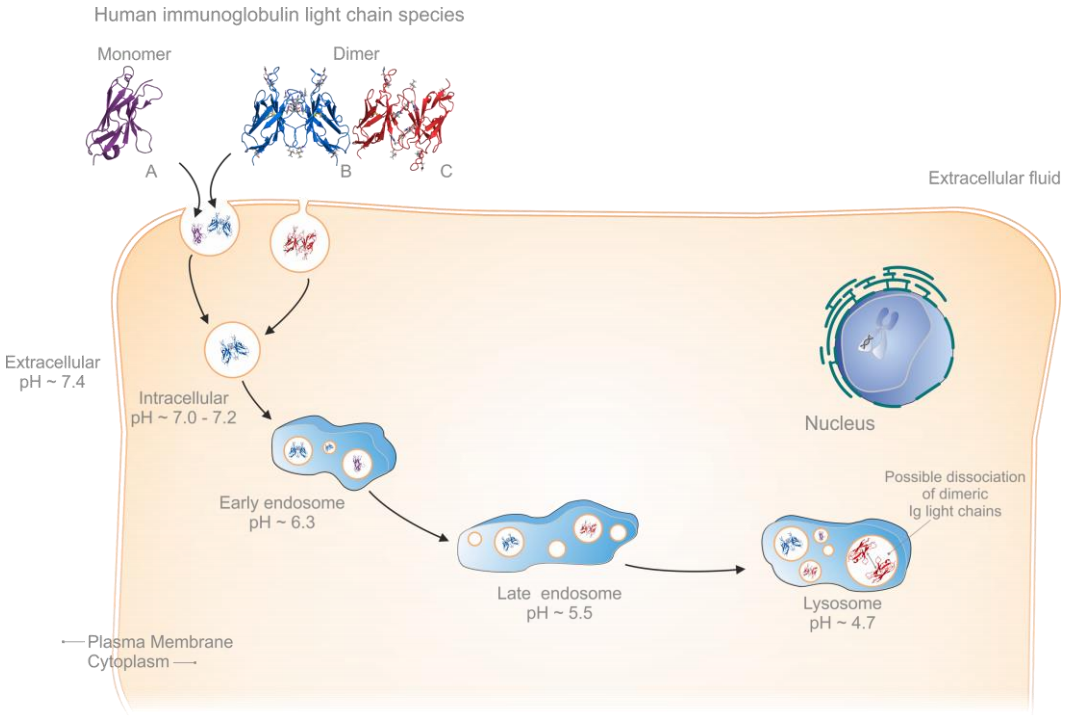


Figure 61 Possible mechanism of V_L internalisation. Monomeric or homodimeric V_L of different orientations (A B C) may be internalised. Proteins internalised by endocytosis which may proceed through a microtubule-mediated endosomal and lysosomal pathway. This process is accompanied by stages of pH acidification (illustrated).

In this chapter, we aim to explore the current working hypothesis, and investigate the mechanisms of V_L internalisation and associated cytotoxicity surrounding the kappa IV light chain variable domains SMA and LEN which has not been addressed before. We use the H9c2 rat cardiomyoblast cell line, and monitor the internalisation of Fluorescein isothiocyanate (FITC) labelled V_Ls by fluorescence microscopy. Cytotoxicity assays reveals that SMA possesses greater cytotoxic effects in an unusual dose dependant manner. In the attempt to devise a therapeutic strategy, a study using the proteasome inhibitor MG132 is described herein.

6.2. Materials & Methods

6.2.1. Preparation of VLs SMA and LEN

Recombinant V_Ls SMA, and LEN were expressed and purified as reported previously in chapter 5. Pure V_Ls were dialysed against 10 mM PBS pH 7.4 before concentration in a 0.5 mL 10 kDa MWCO (Millipore) after rinsing in MilliQ and PBS to a 10 mg/mL stock solution.

6.2.2. Immunoblot for detection of V_Ls

Recombinant LEN and SMA were resolved by gel electrophoresis (60 min at 165 V) and transferred (60 min at 100 v) onto polyvinylidene difluoride (PVDF) membranes (0.45 µm, Hybond-P Amersham bioscience) using a Bio-Rad gel electrophoresis system. For immunoblots, both proteins were single blotted onto PVDF membranes and air dried. Both membranes were blocked in Tris-buffered saline (TBS) supplemented with 0.1% Tween 20 and 5% non-fat dried milk for 4 h at room temperature with gentle agitation. For the detection of

recombinant V_Ls (SMA and LEN), membranes were incubated with primary monoclonal anti-human kappa light chain (1:1000 K4377 Sigma) in TBS, 5% nonfat dry milk and 1% Tween 20 overnight at 4^{o}C with gentle agitation. Following the primary incubation period, Horseradish peroxidase (HRP) conjugated secondary antibody (#A10668 goat anti-mouse IgG Sigma-Aldrich 1:5000) was incubated with the membrane in the presence of 5% w/v nonfat dry milk, and 0.1% Tween 20 for 1 h with gentle agitation. Enhanced chemiluminescence (ECL) methods were used to detect bound antibodies according to manufacturer's protocols (Immobilon Western Chemiluminescent HRP). Membranes were exposed to Amersham Hyperfilm (GE Healthcare) for ~ 60 seconds. As a negative control for the reaction the intrinsically disordered 14.46 kDa protein alpha synuclein was used.

Immunoblots associated with the MG132 pharmacological inhibition assay were processed under identical conditions as described above, but include the use of Nrf2 and GAPDH antibodies (cell signalling) both at dilutions of 1:1000. Total protein concentration of cell lysate in these assays was determined by a BCA assay performed according to manufacturer's guidelines (Pierce).

6.2.3. Immunoglobulin V_L labelling by FITC

In a similar method to the procedure of labelling light chains with Oregon green described previously (Levinson, Olatoye et al. 2013), purified LEN and SMA were labelled with FITC, a fluorescent probe that is said to bind selectively to all solvent ε- and N-terminal amines exposed amines (Jobbagy and Kiraly 1966, Hermanson 2013) were each dialysed into 10 mM PBS pH 7.4 overnight

with multiple changes using a 3000 MWCO (Slide-A-Lyzer) Mini Dialysis unit (Thermo Scientific). Membranes were presoaked in MilliQ to remove the storage glycerol solution. Labelling of V_L proteins with Fluorescein isothiocyanate (FITC) (Sigma) was performed according to manufacturer's guidelines with minor modifications to protein concentration (20 μM of each V_L protein used per reaction). Following incubation, samples were buffer exchanged into PBS by application to a NAP-5 column (GE biosciences) to remove a proportion of unbound FITC and Dimethyl sulfoxide (DMSO) present from the FITC storage solution. Buffer exchange was followed by diafiltration into 10 mM Tris pH 7.4 using a 10,000 MWCO centrifugal filter (Millipore). Around ~10 series of dilution and concentration were performed until both the dialysate appeared clear, and showed no measurement at 495 nm (assessed by nanodrop), indicating there was no unbound FITC remaining. The amount of bound FITC was determined experimentally (nanopore) according to the manufacturer's guidelines. Protein samples were only used when a FITC labelling ratio of 1 and above was achieved as this proved to be easily detectable under the microscope.

6.2.4. Cell culture

H9c2 rat cardiomyocytes were kindly gifted from Dr Parveen Sharma. For use, cells stored in 90% fetal bovine serum (FBS) and 10% DMSO were rapidly thawed (37 ^{o}C (~ 1 min), and the vial contents diluted with in Dulbecco's Modified Eagle's Medium (DMEM) (Invitrogen) supplemented with 10% FBS, and 5% penicillin. Cells were pelleted by centrifugation (*1000 x g* for 5 min) the supernatant discarded, and the pellet resuspended in 7 mL DMEM with the above supplements. Cells were transferred to an uncoated T-25 culture flask

(Corning) and cultured as a monolayer in a humidified atmosphere incubator at 37 ^{o}C, 95% air and 5% CO_2 and nourished at 2-3 day intervals. At the desired confluency (70-80%), cells were rinsed twice with PBS sub cultivated by trypsinisation and transferred into an uncoated T-75 culture flask (Corning), or prior to internalisation experiments, seeded into 24-well plates (Corning, Costar) containing a 12 mm (diameter) 0.16-0.19 mm (depth) glass coverslip (Academy sciences) at a density of 40,000 cells/well (500 µL per well). Cells were left to adhere for 16 hours prior to experimentation.

6.2.5. Protein internalisation assay

Before protein administration, cell confluency was estimated by microscopy. Recombinant V_L proteins were administered at a confluency judged to be below ~ 50% to prevent overgrowth. In this experiment, different concentrations (indicated at results section) of soluble recombinant SMA and LEN were administered in PBS. Negative controls (PBS alone) that were free of FITC labelled V_L domains were also performed. Plates were imaged 24 hours after V_L addition.

6.2.6. Slide preparation

After 23 hours, cell media was replaced with fresh media containing lysotracker (RED DND-99 Thermofisher final concentration of 50 nM) for a final one hour incubation period. Slides were rinsed twice in PBS (room temperature for 10 mins) incubated for an additional hour in PBS (4 ^{o}C for 30 min) fixed in 4% paraformaldehyde (PFA) for 30 min at room temperature and quenched (33 mM glycine in PBS).

For antibody based detection methods, cells were rinsed in PBS (twice, 15 min, room temperature) permeabilised (0.2% tween 20, 0.5% Triton X-100 for 10 min at 4 ^{o}C) and blocked (5% bovine serum albumin (BSA), 0.2% Triton X-100 in PBS for 30 min, room temperature) before incubation with the primary antibody (KP53 1:50 (Sigma), in 5% BSA 0.2% Triton X-100 in PBS overnight at 4 ^{o}C).

For detection of internalised V_Ls, cells were washed in permeabilisation buffer (3 x 15 min) and incubated with secondary antibody (anti-mouse conjugated to Alexa Fluor 488 green 1:1000 Abcam). A further PBS wash step (10 min, room temperature) was followed by incubation of PBS containing Hoechst stain (33258 (SIGMA) 1:5000) and Phalloidin (1:250) for 20 min at room temperature. A final wash step (PBS, twice, 10 min room temperature) was conducted prior to mounting. For internalisation assays using FITC labelled proteins, cells following incubation with V_Ls were fixed (4% PFA, 30 min, room temperature) the reaction quenched (33 mM glycine in PBS 10 min, room temperature) and washed in PBS (2 x 15 min, room temperature).

In all studies, cells were mounted in the presence of ProLong Antifade (Thermo), sealed using lacquer, stored at 4 ^{o}C and visualised within 48 hours of fixing to prevent fade and maximise signal intensity. At all possible stages, cells were kept in the dark to avoid light exposure. Control slides containing cells that were free of V_Ls were included for all assays and used to match exposure levels on the 488 nanometer channel.

6.2.7. Confocal microscopy

V_L internalisation experiments were imaged using an Axio observer z1 microscope (Zeiss) equipped with ApoTome and a 40x Plan-Neoflaur oil

immersion objective (Zeiss). Wavelengths of 488_{nm}, 568_{nm} and 680_{nm} were used to visualise FITC/Alexafluor conjugated antibody, Phalloidin/ LysoTracker-Red and Hoechst respectively. For Z-stack, ~20 images were taken at 0.28 µm intervals. Analysis of Z-slice and 3D reconstruction of images was performed using Zeiss Zen blue v2.3. Exposure levels were kept consistent between experiments to allow for direct comparison with control slides.

6.2.8. Cell toxicity assay and statistical analysis

The toxicity of V_L in cell culture was assessed by using a cell counting kit-8 (CCK-8). The CCK-8 assay is a colorimetric assay, and uses a water-soluble tetrazolium salt that in a viable cell is reduced by dehydrogenases to produce a formazan dye. The quantity of this dye is proportional to the number of living cells present in the culture, and so cells treated with different concentrations of light chain variable domains can be directly compared to cells free of light chain (treated with CCK-8 solution only) which serve as a control (description adopted from manufacturer's technical manual (Dojindo laboratories (Kumamoto, Japan). Rat H9c2 cardiomyocytes were seeded into a 96-well plate at a density of 8,000 cells per well in a total of 80 µL DMEM media that was free of phenol red. After 24 hours, cells were treated with SMA or LEN at concentrations 1, 5 and 10 µM for an additional 24 hours. As a control, live cells were cultured in media substituted with 20% V_L storage buffer (to match protein samples, 10 mM Tris pH 7.4), or as control of dead cells, media supplemented with 1% Triton (final concentration). All culture volumes totalled 100 µL. Following the incubation period, CCK-8 solution was added at 10 µL per well and incubated at 37 ^{o}C for 2 hours before the absorbance measured

at 450 nm. Wells containing V_Ls were performed in triplicate. For live and dead cell controls, experiments were repeated six times. The results are expressed as the standard error of the mean. Group differences were assessed by a one-way analysis of variance (ANOVA) in conjunction with dunnett's post-hoc analysis. Results are expressed as the mean ± s.e.m and any differences deemed statistically significant (P value <0.05) are indicated (*). Statistical analyses were performed using PRISM (v7).

6.2.9. Pharmacological inhibition assay

At the desired confluency Rat H9c2 cardiomyocytes were seeded into 24-well plates (Corning, Costar) containing a 12 mm (diameter) 0.16-0.19 mm (depth) glass coverslip, (Academy sciences) at a density of 40,000 cells/well (500 µL per well). Cells were left to adhere for 16 hours prior to experimentation. Cells were treated with 5 µmol/L MG132 for 2 hours (for western blot experiment for detection of nrf2 to confirm efficiency of inhibitor) or for 24 hours for the inhibition assay, which was also conducted in the presence of SMA at 10 µM.

6.3. Results

6.3.1. Immunoblot

Both SMA and LEN were expressed, purified and their secondary structure determined previously (Chapter 5). To confirm the ability of kappa-specific antibody KP53 (Sigma) to bind both SMA and LEN, immunoblots containing V_Ls were incubated in the presence of the antibody and developed using ECL based methods. Successful Immunoreactivity was noted for both SMA and LEN (Figure 62). Here, SMA was loaded in a crude state, designed to evaluate

the presence of a higher molecular weight SMA species previously observed in periplasmic extraction experiments presented in Chapter 5 (Figure 51). This was however, later purified by SEC, confirmed by SDS-PAGE electrophoresis and RP-HPLC.

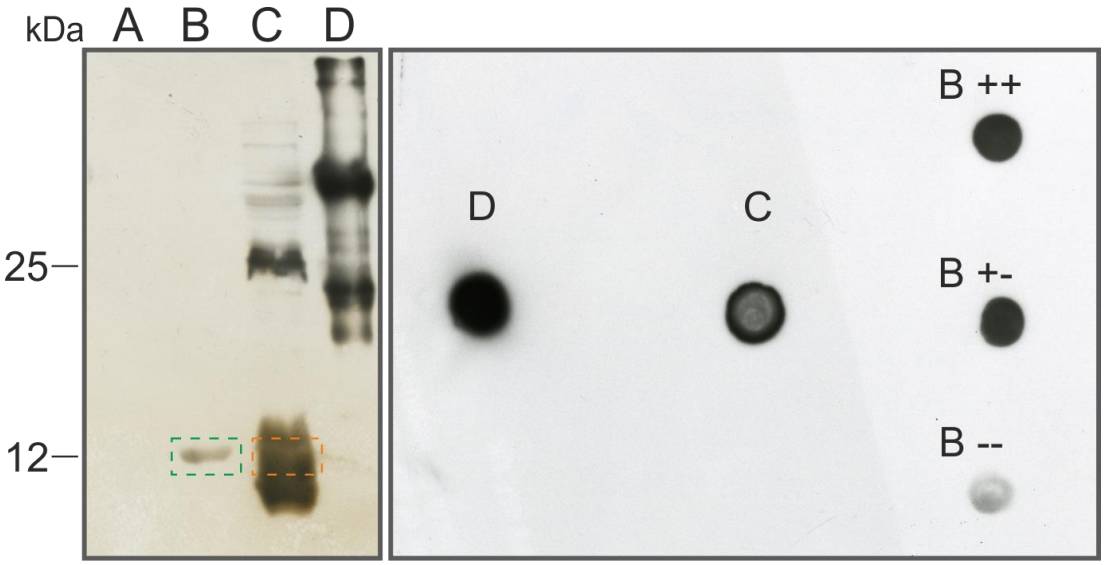


Figure 62 Validation of antibody specificity to recombinant V_L domains using Immunoblotting procedures. Left hand panel shows western blot of alpha-synuclein (14.46 kDa, Lane A) used a negative control, LEN (Lane B), crude state of SMA (Lane C) and a mixture of kappa and lambda antibody (Sigma I5381) derived from mouse serum used as positive control. Right hand panel shows immunoblot of identical samples. Samples of LEN were blotted in native (--), in the presence of beta-mercaptoethanol (+-) and the presence of beta-mercaptoethanol and heating at 90 °C (++) for 10 minutes.

Interestingly, the immunoreactivity of LEN was influenced by the denatured state of the protein (Figure 62) where -+ and ++ indicate the presence or absence of beta-mercaptoethanol, and heating respectively. This suggests that the epitope (which is not documented in manufacturer's notes) is more accessible in a denatured state. Positive reaction of crude SMA fraction suggests the presence of SDS-resistant higher molecular species. This was only apparent in liquid growth cultures that exceeded 16 hour growth time

post-IPTG induction. Overall, the results indicate that the antibody is able to detect both V_Ls without detecting the negative control (alpha synuclein). As a positive control, we used a mixture of Ig acquired from mouse serum (Sigma I5381) were immunoreactivity was detected.

6.3.2. Antibody based Immunofluorescence

Current studies investigating the trafficking of V_Ls have used mouse, human and rat derived cells. Here, we use the H9c2 rat cardiomyocyte cell line originally isolated from rat ventricular tissue (Kimes and Brandt 1976) to assess the ability of kappa light chain variable domains SMA and LEN to internalise. The cell line is commonly used in cardiovascular research (Lenco, Lencova-Popelova et al. 2015, Peter, Bjerke et al. 2016, Witek, Korga et al. 2016), and has been found to mimic cell signalling pathways found in animal-derived primary cardiomyocyte cell lines (Watkins, Borthwick et al. 2011, Kuznetsov, Javadov et al. 2015) supporting the validity of using such cell line as a model system.

In an initial study designed to evaluate the ability of commercially available antibody (KP53) screened in the immunoblot (Figure 62) to detect internalised LEN in cell culture, H9c2 cells were incubated with purified LEN at two different concentrations for 24 hours. Cells were then fixed, probed with the appropriate antibodies and visualised using 2D immunofluorescence microscopy. LEN was administered at two different concentrations; low LEN at 0.05 mg/mL (3.96 μM) and High LEN at 3 mg/mL (237 μM) final concentration which correspond to the concentrations where monomeric and dimeric species are in their abundance (i.e very little dimeric LEN exists at 3.96 μM), as

discussed in the following publication (Qin, Hu et al. 2007). Primary antibody was administered at dilutions of 1:50 and 1:100 to determine the appropriate concentration, and any internalisation detected by species-specific Alex Fluor488 conjugated secondary antibody.

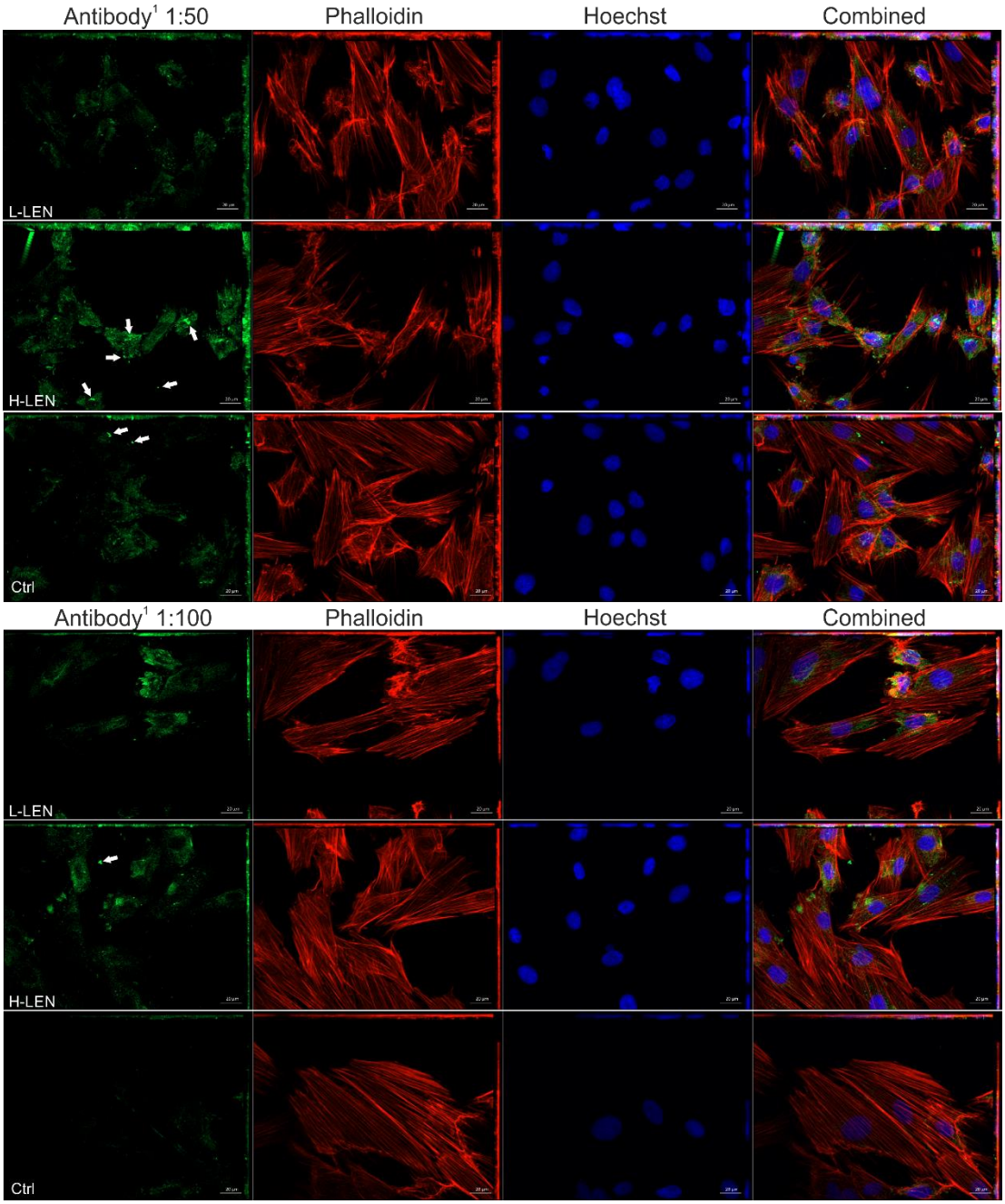


Figure 63 V_L internalisation into evaluated by anti-kappa antibody KP53. H9c2 cardiomyocytes were incubated in the presence of LEN at concentrations that correspond to monomeric and dimeric species (H-LEN 3.0 mg/ml and L-LEN 0.05 mg/ml) for 24 hours. Cells were permeabilised and internalised proteins detected using a primary anti-human kappa light chain antibody (Sigma) with an Alexa Fluor 488 conjugated secondary antibody. Detection of signal was made by maximum intensity projections of z-stack. A control experiment (Ctrl) that was free of LEN was conducted. All slides were treated in an identical manner. Combined shows all channels (blue hoechst - nuclei, red phalloidin - f-actin and green – secondary antibody used to detect LEN). White arrow used to indicate intense fluorescent pixels. Scale bar is 20 μm.

Analysis of confocal images reveal that LEN cultured cardiomyocytes using primary antibody dilution of 1:50 have only marginal enhanced brightness in comparison to control slides (Figure 63). While discrimination between the fluorescence (green) in control and LEN cultured cells is improved in experiments using lower primary antibody dilutions (1:100) which may indicate the presence of internalised LEN, the long exposure times (~1.2 ms) used to reach observable levels of fluorescence and presence of intense fluorescent pixels (white arrow) similar to that visualised by (McWilliams-Koeppen, Foster et al. 2015), may indicate aggregated fluorescent antibody). Overall, this made discriminating between cellular auto fluorescence and internalised V_L quite difficult. Confocal images also reveal a diffuse, granular staining rather than the punctate cellular staining pattern observed in other studies of similar nature (Levinson, Olatoye et al. 2013) which made us question again, the validity of using antibody detection methods for visualisation of internalised V_L. Overall, we concluded that the secondary antibody in this crowded intracellular environment appears to exhibit large levels of cross reactivity, and was found to be overall, rather insensitive for the detection of V_L.

6.3.3. Purity analysis of SMA and LEN before labelling

To overcome the observed pitfalls in antibody based detection methods, each V_L was directly labelled with FITC, a dye that is commonly used in protein labelling procedures (Koniev and Wagner 2015). An important consideration here is the purity degree of the material to be labelled. As FITC binds to all solvent exposed amines (The and Feltkamp 1970, Hermanson 2013), the recombinant V_L protein sample has to be free of all host bacterial cell impurities that may be undesirably labelled. As the variable domains would

also be used in cell cytotoxicity assay (described later on in this chapter) we also took this opportunity to ensure that samples were free of bacterial contaminants that may give false positives (contaminants themselves causing cytotoxic effects rather than the protein of interest). The purity of each sample was previously assessed by SDS-PAGE and RP-HPLC (Figure 55). Both techniques show that no contaminants are present.

6.3.4. Detection of internalised FITC-labelled LEN by Immunofluorescence

H9c2 cardiomyocytes were incubated with concentrations of FITC labelled LEN at 1, 5, and 10 μM (final concentrations) for 24-hours. Based on the observations of *Qin et al.,* LEN is mostly monomeric at these concentrations (K_d of LEN is 10 μM. 88% of the protein is monomeric at 40 μM and 10% of the protein monomeric at 237 μM) (Souillac, Uversky et al. 2002, Souillac, Uversky et al. 2003). For an accurate comparison between samples incubated with or without FITC labelled LEN, all images were taken using identical exposure levels.

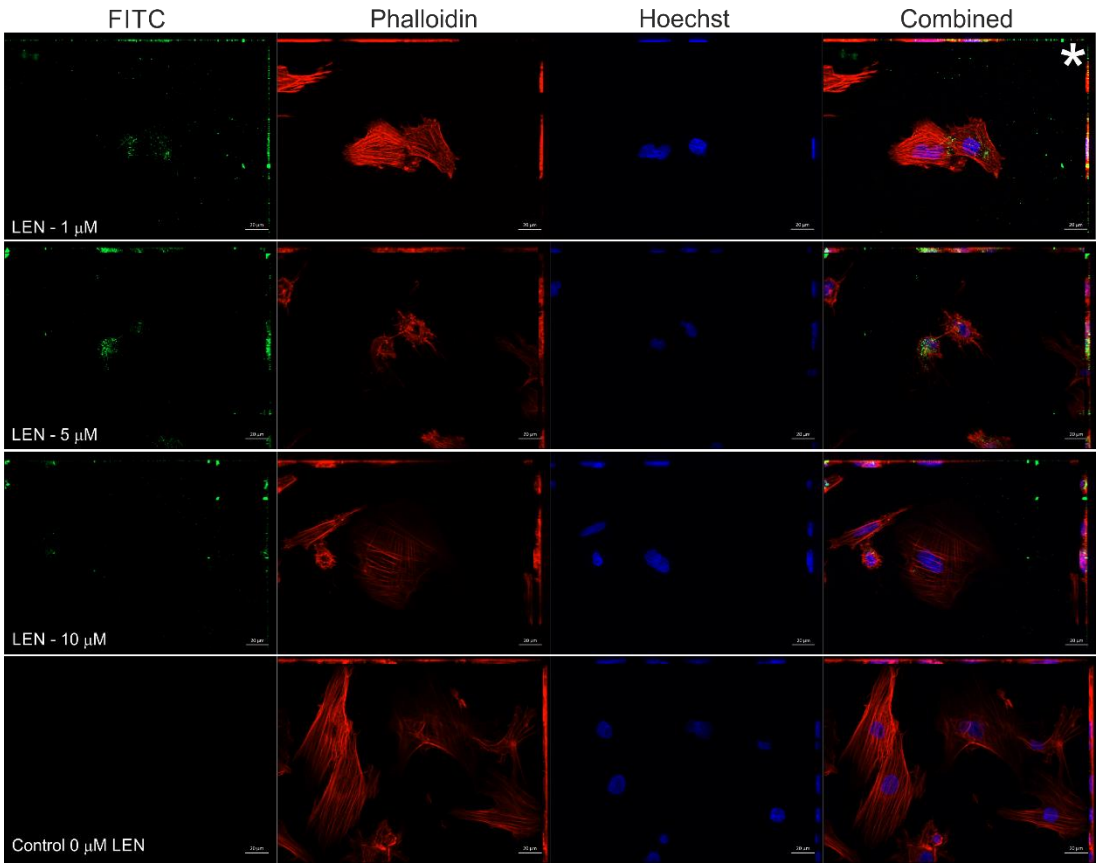


Figure 64 Internalisation of FITC conjugated LEN monitored by immunofluorescent microscopy. In three independent experiments, FITC labelled LEN (green) at concentrations of 1, 5 and 10 µM (indicated) were incubated with rat H9c2 cardiomyocytes for 24 hours. Detection of the FITC signal was made by maximum intensity projections and Z-stack analysis. A control experiment (Ctrl) that was free of LEN was conducted. Combined shows all channels (blue hoechst - nuclei, red phalloidin - F-actin and green – FITC labelled LEN). Scale bar is 20 µm. Asterisk indicates image that was taken for further processing.

In contrast to antibody based methods (Figure 63), the detection of FITC conjugated LEN by confocal microscopy was much improved (Figure 64). Notably, there was no detectable fluorescence in control slides even under identical exposure levels. FITC signals were clearly distinguishable against the background and this fluorescence was detected under much lower exposure time (350-600 millisecond exposure using Apotome (Zeiss)) to our previous experiment (high exposure times >1.2 sec).

Confident with this procedure, a number of interesting observations were made. Firstly, FITC conjugated LEN appears to be internalised in all experiments, with no visible differences between the different concentrations used. The distribution of LEN appears to be sparse, where some regions of the slide were completely absent of signal in the 488 channel. Such an observation is consistent with the analysis of AL-09 in HL-1 cardiomyocytes (Levinson, Olatoye et al. 2013). Although, it is worth noting that in comparison to the three-colour confocal images published in this study, the quantity of internalised LEN, on visual inspection alone is noticeably lower than that of amyloidogenic AL-09. One striking observation is seen in H9c2 cells that were incubated with LEN at 5 μM. In one field of view, strong staining is observed in a cell that gives the appearance of apoptotic staging. This observation is based on the cells unusual morphological features, where it does not display the typical striated morphology of F-actin filaments that other cells within the culture display and instead shows blebbing of the membrane, which is a feature of apoptosis (Noritake, Aki et al. 2012) (Figure 64). Surrounding cells (past the field of view for this image) did not display any green signal. As we did not see any changes to cell viability in the presence of LEN administered at this concentration (later presented in this chapter as a CCK8 ssay) this may indicate that LEN has greater affinity for cells undergoing apoptotic staging. At this magnification, overall cell morphology appears largely unaffected by the addition of LEN, where stained F-actin (red- Phalloidin) has a similar structure to control slides with the exception of the aforementioned 5 μM LEN slide.

6.3.5. Internalised FITC-LEN: Z stack and 3D reconstruction

Next, we wished to determine the subcellular origin of the FITC signals. Here, optical sections of the experiment using 1 µM FITC Labelled LEN (marked asterisk taken from previous figure (Figure 64) was analysed further.

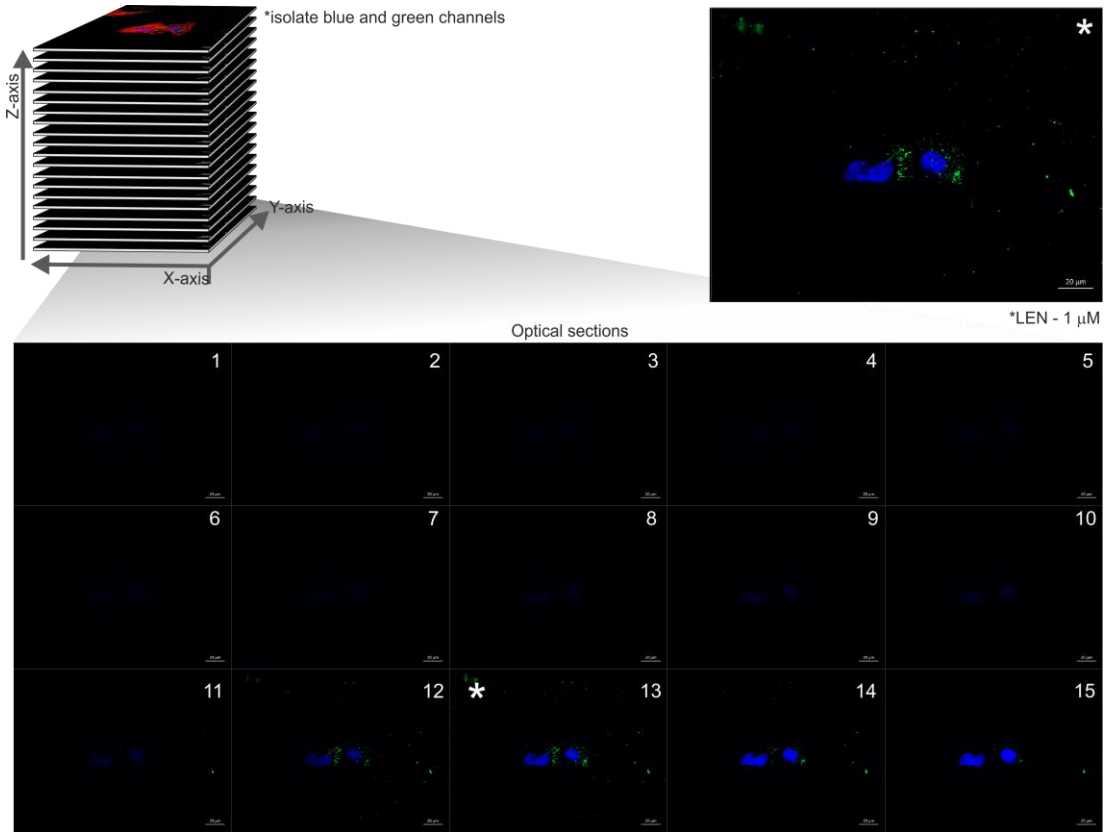


Figure 65 Subcellular localisation of Internalised FITC conjugated LEN assessed by analysis of Z-stack. Optical sectioning of complete z-stacks reveals FITC-labelled LEN (green) is on the same focal plane as the cell nucleus (hoescht – blue) indicating the V_L is inside the cells and not surface bound. Top right panel shows enlarged image of z-slice 13 marked asterisks. Scale bar is 20 µm.

As illustrated in the two-colour z-stack image (Figure 65), the most intense FITC (green fluoresce) signal is found on the focal planes (12,13, and 14) that are occupied by the cell nucleus (hoescht – blue) suggesting that the V_L is not surface bound, and the observed signal originates from inside the cell.

Although extracellular fluorescent material is present, the majority of green signal is intracellular (illustrated in z-slice 13). To complement this analysis, we again used the 1 μM LEN 2D confocal image shown in Figure 64, but here performed three-dimensional reconstruction of the entire z-stack.

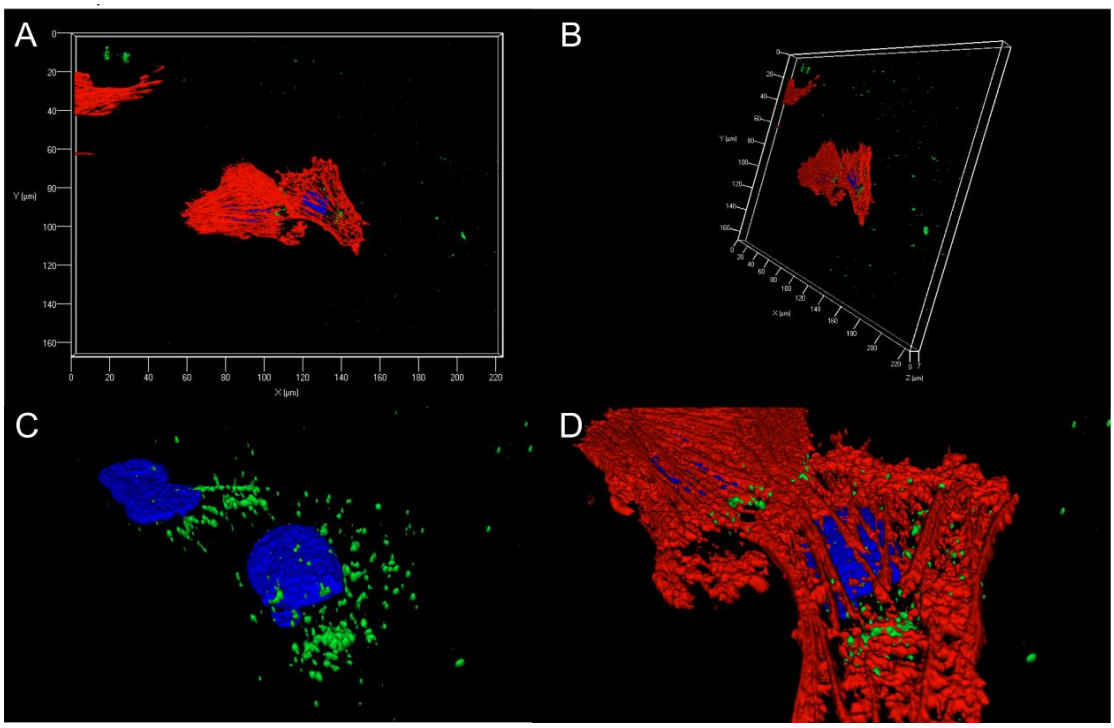


Figure 66. Internalisation and localisation of FITC conjugated LEN assessed by 3D reconstruction. 1 μM FITC labelled LEN (green) that was previously incubated with rat H9c2 cardiomyocytes for 24 hours was taken for further analysis in order to decipher the location of the signal. Combined shows all channels (blue hoechst - nuclei, red phalloidin - F-actin and green – FITC labelled LEN). Scale bar is 20 μm. A and B show 3D reconstruction of three colour z-stacked image of internalisation of FITC conjugated LEN rotated on x and y-axis. C and D show zoomed-in image with (C) and without (D) Phalloidin channel engaged (Red) to demonstrate that the FITC signal is not surface bound. Scale in μm is indicated.

A complete 3D construction that covers the entire depth of the cell (Figure 66) illustrates that the FITC signal is not surface bound. This is most clear in panels C and D where FITC signal can be seen surrounding the cell nuclei.

6.3.6. Detection of internalised FITC-labelled SMA by Immunofluorescence

In a procedure identical to the detection of LEN in cell culture, pure SMA fractions (determined by RP-HPLC Figure 55) conjugated with FITC were incubated with H9c2 cardiomyocytes for a 24-hour period before fixing and visualised using 2D fluorescent microscopy. SMA has a dimer dissociation constant of 40 μM (in comparison to the 10 μM LEN) and so both variable domains are monomeric at the concentrations used in these experiments (1, 5, 10 μM) (Kolmar, Frisch et al. 1994, Qin, Hu et al. 2007, Baden, Owen et al. 2008)

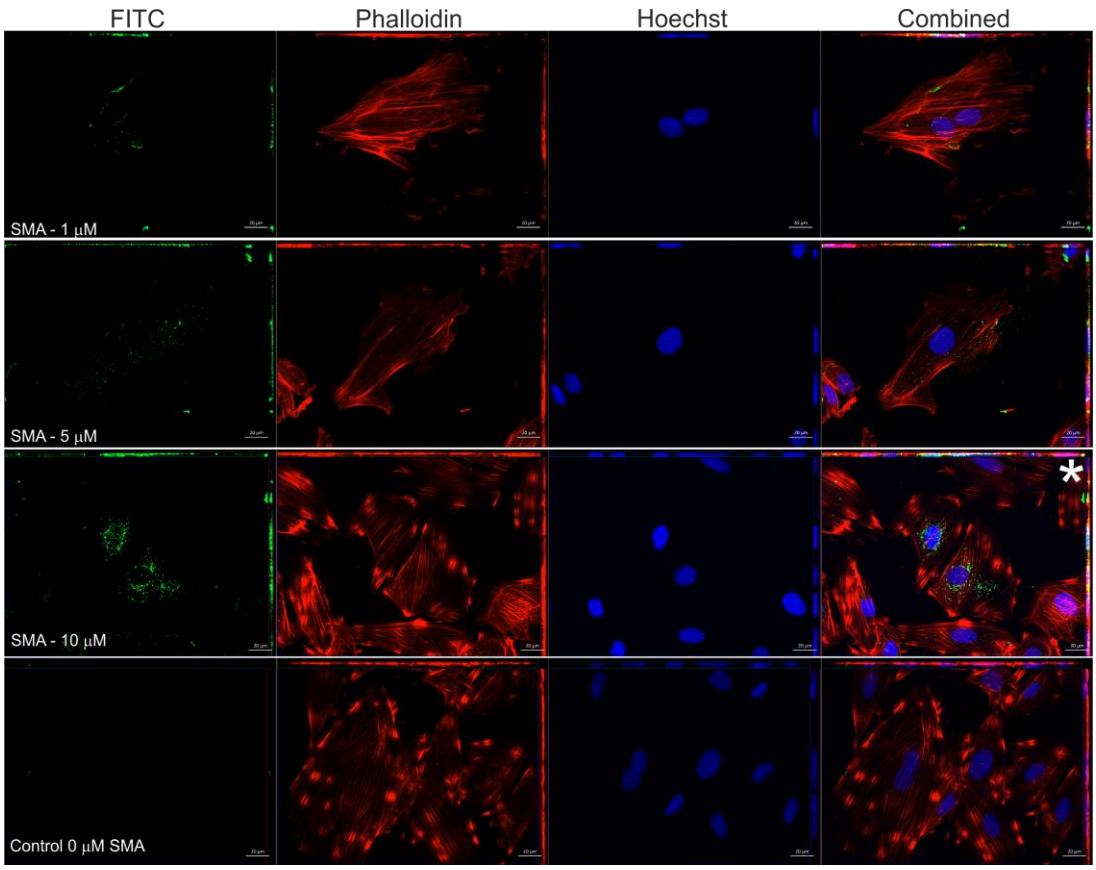


Figure 67 Internalisation of FITC conjugated SMA monitored by immunofluorescent microscopy In multiple independent experiments, FITC conjugated SMA (green) at concentrations of 1, 5 and 10 µM (indicated) were incubated with rat H9c2 cardiomyocytes for 24 hours. Detection of the FITC signal was made by maximum intensity projections and Z-stack analysis. A control experiment (Ctrl) that was free of SMA was conducted. Combined shows all channels (blue hoechst - nuclei, red phalloidin - F-actin and green – FITC labelled LEN). Scale bar is 20 µm. Asterisk (*) in SMA 10 µM indicates data that was taken for further analysis.

As illustrated, FITC labelled SMA is detectable in all experiments (aside from control). However, the level of internalisation for SMA at a concentration of 10 µM is most noteworthy. Despite this level of internalisation, the overall morphology appears unaffected where stained F-actin (Phalloidin, red) retains

the striated pattern in both control and variable domain incubated slides similar to experiments with H9c2 cells incubated with LEN.

6.3.7. Internalised FITC-labelled SMA - Z stack

To show that the origin of the FITC-SMA signal was present within the cell and not on the surface, z-stack images of cardiomyocyte cells incubated with 10 µM SMA (previous Figure 67) were analysed by maximum intensity projection (Figure 68). The z-slice containing the most intense FITC-signal is shown in Z-13 (marked with asterisk, also shown in top right panel as zoomed-in) and clearly indicates the localisation of the fluorescent signal to be within the cell and not surface bound, where individual confocal sections z1 and z20 indicate outside of the cell, and top the of the cell respectively.

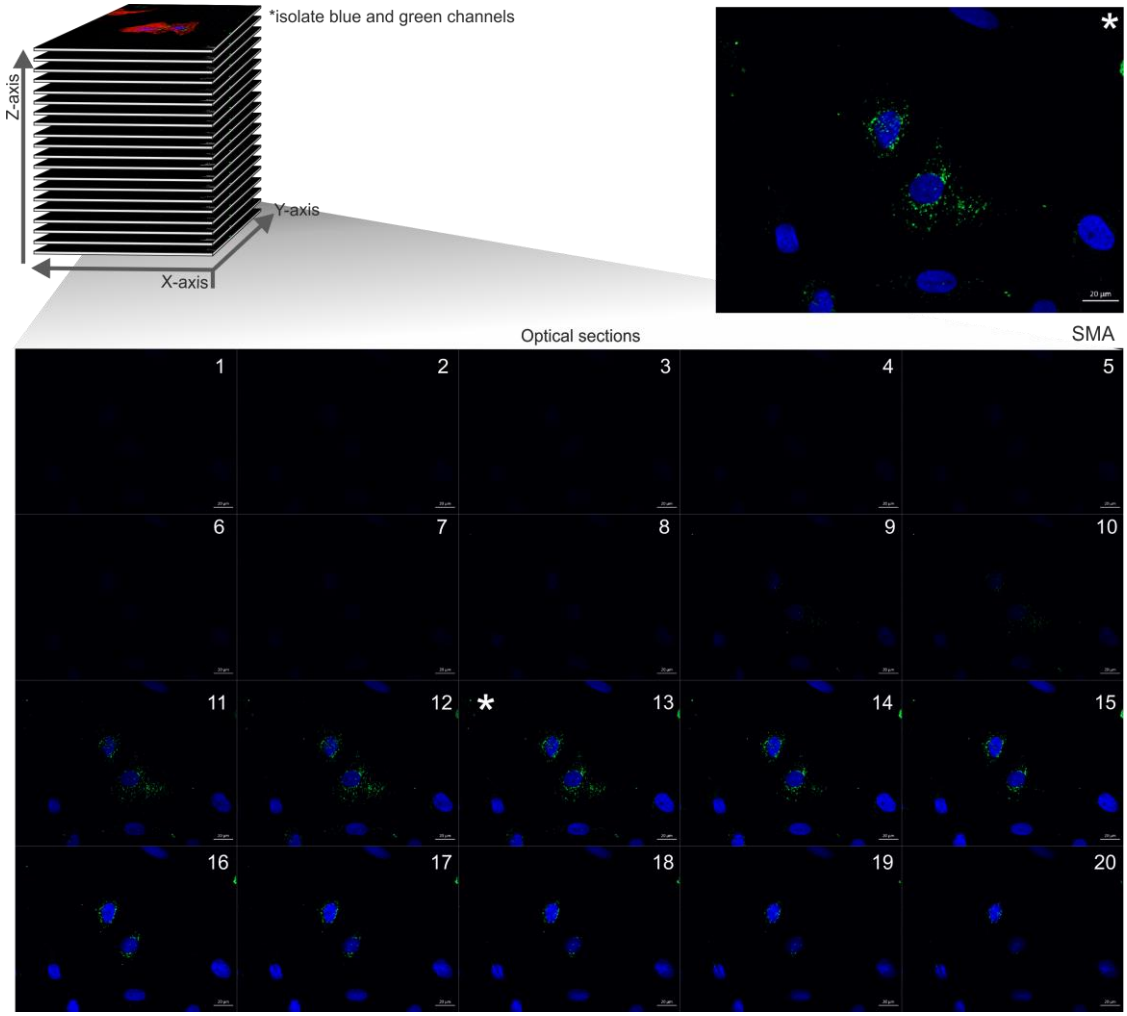


Figure 68 Subcellular localisation of Internalised FITC conjugated SMA assessed by analysis of Z-stack. Optical sectioning of complete Z-stacks reveals FITC-labelled LEN (green) is on the same focal plane as the nuclei (hoescht – blue) indicating the V_L is inside the cells and not surface bound. Top right panel shows enlarged image of z-slice 13 marked asterisks. Scale bar is 20 μm.

For a clearer depiction to the origin of the FITC signal, z-stacked images underwent 3D reconstruction (Figure 69). Orthogonal cross section of the z-projection (panel C) clearly indicates the protein to localise to the perinuclear region of the cell and is not bound to the surface. Interestingly, it appears that some of the FITC conjugated SMA (green) is associated with the nucleus (Figure 69 panel C) where FITC signal can be seen to originate within the nuclear space. While the ability of light chains to localise to perinuclear

compartments have been noted (Marin-Argany, Lin et al. 2016), to the best of our knowledge a demonstration of intranuclear localisation of light chain variable domain has not been documented by such methods presented here previously. Similar staining patterns have however, been documented previously for other fluorescently tagged proteins and small molecules (Lux, Goerlitz et al. 2005, Huang, Mackeyev et al. 2013).

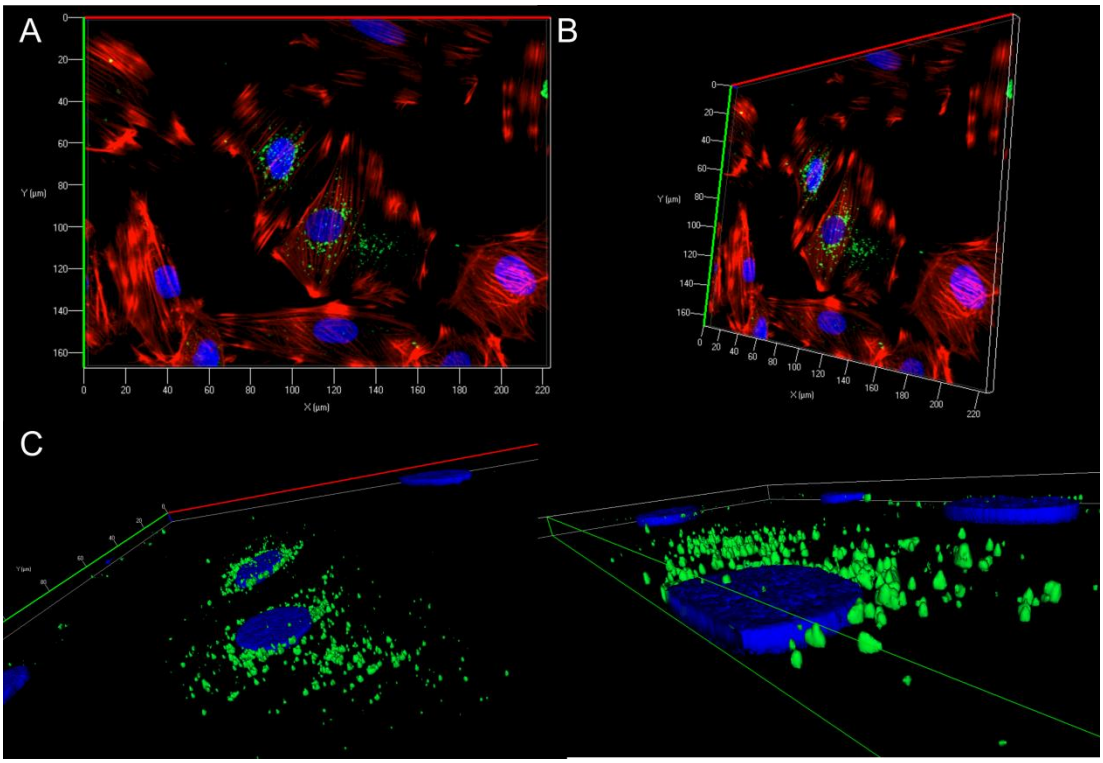


Figure 69 Internalisation and localisation of FITC conjugated SMA assessed by 3D reconstruction. 3D reconstructions of H9c2 rat cardiomyocytes incubated with 5 µM FITC labelled SMA (green) were performed in Zeiss Zen blue software. A and B show 3D reconstruction of three colour z-stacked image of internalisation of FITC conjugated SMA rotated on x and y-axis. Cells are stained with hoechst (nuclei - blue) phalloidin (f-actin red). C (right side) shows cross section of cell nucleus (blue) with FITC on same plane indicating that SMA is not localised to cell surface. Scale in µm is indicated.

In a final set of imaging experiments, we wished to assess the possibility that variable domains are able to enter the cell via endocytosis. FITC conjugated LEN (at concentrations of 0.05 mg/ml L-LEN, 0.5 mg/ml M-LEN and 3 mg/ml H-LEN) were incubated with H9c2 cells for 24 hours. Lysotracker red (endosome stain) was added to the cells 30 minutes before fixation and imaging.

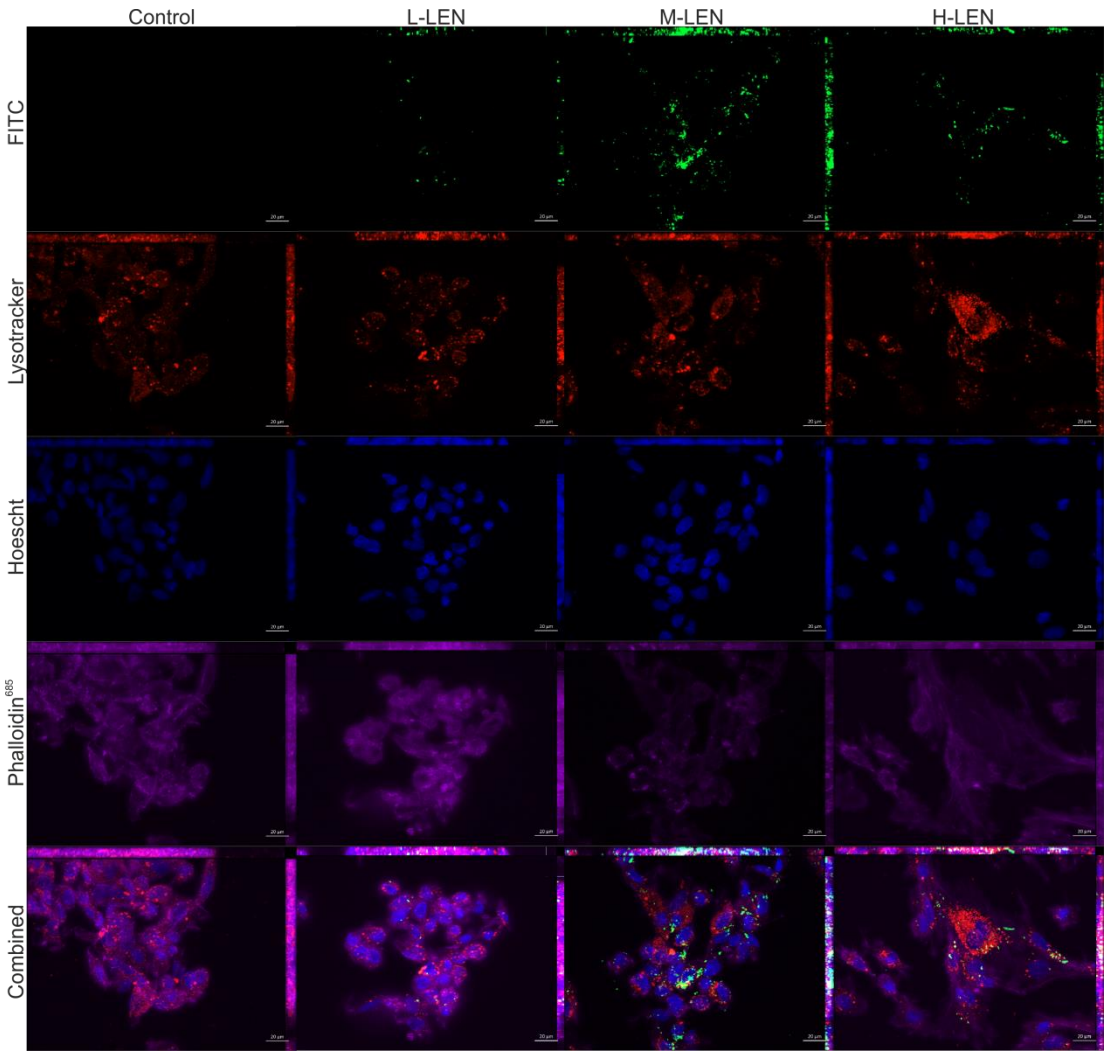


Figure 70 Internalisation of FITC conjugated LEN into lysosomes monitored by immunofluorescent microscopy In multiple independent experiments, FITC conjugated LEN (green) were incubated with rat H9c2 cardiomyocytes for 24 hours. Detection of the FITC signal was made by maximum intensity projections. A control experiment (Ctrl) that was free of SMA was conducted. Combined shows all channels (red lysotracker- endosome and lysosome, blue hoechst - nuclei, purple phalloidin - F-actin and green – FITC labelled LEN). Scale bar is 20 μm.

In all experiments, there was no detection of light chains in lysosomal or endosomal compartments. The lysotracker dye is acidophilic, preferentially staining the acidic compartments of the early and late endosome that have pH of ~ 6.3, 5.5 (respectively) as well as the lysosome (pH 4.7) (Figure 61). If the variable domains were to be trafficked in this pathway, co-localisation would be indicated by a yellow signal, a product of green signals of the FITC conjugated protein and red of the lysotracker being merged. An example of this is documented for light chain variable domain AL-09 (Levinson, Olatoye et al. 2013). Overall, the data presented here suggests V_L enter the cell in a mechanism of trafficking.

6.3.9. Cytotoxicity assay

Having, determined the ability of both SMA and LEN to internalise by 2D fluorescence microscopy, we next sought to determine the cytotoxic potential of each V_L H9c2 rat cardiomyocytes were cultured *in vitro* and incubated with different concentrations of SMA and LEN (1, 5 and 10 μM**)** that correspond to conditions used in fluorescence experiments. The results of a CCK-8 cell toxicity assay where, each variable domain was incubated with cells for 24 hour (plus 2 further hours after CCK 8 dye addition) are shown in Figure 71.

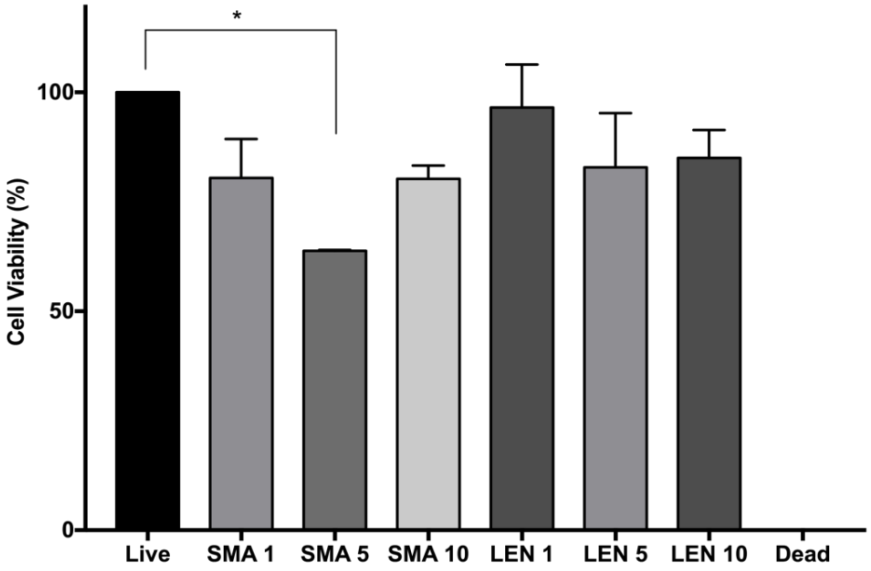


Figure 71 The effect of V_Ls SMA and LEN on rat cardiomyocyte toxicity. Different concentrations of light chain proteins (1, 5, and 10 μM shown as 1,5,10 on the figure) were incubated with H9c2 cell lines for 24 hours before analysis by CCK-8 assay at 450 nm. A one-way analysis of variance (ANOVA) with dunnett's post-hoc analysis was performed. *P<0.05. CCK-8 assays of H9c2 cell toxicity after 24 hour incubation (n=6 for controls and Triton treated cells, n=3 for LC incubated cells, mean ± s.e.m). Dead and live cell wells correspond to 0% and 100% cell viability

As illustrated (Figure 71) there were no statistically significant changes between live cells and cells treated with LEN. However, there was a significant decrease in the percentage viability of cells treated with SMA at concentrations of 5 μM (P<0.05). SMA is notably less thermodynamically stable than LEN (Raffen, Dieckman et al. 1999) which may account for the enhanced toxicity demonstrated. Interestingly however, this was not mirrored in SMA doses of 1 μM or 10 μM. The reason for this unusual dose response is unclear. However, it is worth considering that 12% of the protein is dimeric at 40 μM, it may be therefore possible that at 10 μM there is a slight increase in the number of dimeric species present in solution in comparison to SMA at 5 μM (where there would be more monomers) ((Souillac, Uversky et al. 2002, Souillac, Uversky et

al. 2003). As discussed previously in this thesis dimeric species are suggested to be non amyloidogenic, instead perform a protective role, hence less toxicity is observed at 10 μM compared to 5 μM. Under the same rationale for this observation then, at concentrations of 1 μM, there is simply not enough monomeric V_L present to induce a statistically significant cytotoxic effect.

6.3.10. Pharmacological Inhibition assay

Proteasomal inhibition has emerged as a promising strategy for the treatment of AL amyloidosis. The plasma B clones, found in patients with multiple myeloma and AL amyloidosis have shown they are particular susceptible to the agents bortezomib, carfilzomib, and ixazomib (Driscoll and Girnius 2016, Jelinek, Kryukova et al. 2016). These agents are designed to inhibit proteasome function, thereby preventing the degredation and clearance of intracellular proteins through the ubiquitin-proteasome pathway (Lü and Wang 2013, Kubiczkova, Pour et al. 2014, Driscoll and Girnius 2016). Bortezomib for example, has demonstrated its ability to induce apoptosis through binding to the β5 subunit located within in the 26S proteasome, a component which is responsible for chymotypic-like activity (Jelinek, Kryukova et al. 2016). Several studies have documented other beneficial effects of bortezomib including its ability to inhibit angiogenesis and suppress cellular adhesion molecules, preventing the growth and spread of malignant plasma B cells residing within the bone marrow (Hideshima, Richardson et al. 2011, Kubiczkova, Pour et al. 2014, Jelinek, Kryukova et al. 2016).

For MM, the goal of proteasomal inhibition is to prevent the infiltration of the malignant plasma cell clone into bone marrow (Kristinsson, Minter et al. 2011, Tosi 2013). By comparison, the plasma cell clone in AL is less aggressive (lower tumour burden) and does not pose a significant problem. Rather, these cells are able to produce enormous quantities of light chain fragments that are able to cause widespread organ damage and dysfunction enter the bloodstream and can form amyloid fibrils (Rajkumar, Dispenzieri et al. 2006, Martin and Ramirez-Alvarado 2010). While proteasome inhibitors are routinely used for the treatment of MM, the underlying plasma cell dyscrasia common to both disorders means that such inhibitors can also have a positive effect in treating AL amyloidosis. In AL, these same inhibitors are used to induce apoptosis in the plasma B cell population which leads to a dramatic reduction in circulating toxic light chain fragments chain. In addition to these effects, the proteasome inhibitor MG132 has also demonstrated the ability to perturb the uptake of light chains in rat primary cardiac fibroblasts (Monis, Schultz et al. 2006). Given that the possibility of aggregation to start intracellularly (Walsh, Tseng et al. 2000) the intracellular environment may provide conditions to promote aggregation of light chain variable domains (Monis, Schultz et al. 2006), and preventing their internalisation could act as an additional mechanism to inhibiting the cytotoxic potential of light chain fragments to those described above. To our knowledge, there has been no progression on this study using this particular agent for the study of AL amyloidosis since. This motivated us to investigate the ability of MG132 to arrest the internalisation of light chain variable domain SMA into rat H9c2 cells. As a positive control for MG132 efficiency, the nuclear factor E2 related factor 2 (Nrf2) was used. Nrf2

is a highly conserved transcription factor that regulates a number of antioxidant response elements (Zhou, Sun et al. 2014). Under normal conditions, cytoplasmic Nrf2 is rapidly degraded (half-life of 15 minutes) by the ubiquitin proteasome dependent degradation pathway (Li, Paonessa et al. 2012, Malloy, McIntosh et al. 2013). Failure to establish ubiquitination, promotes Nrf2 retention in the cytoplasmic space where it is subsequently translocated to the nucleus (Cui, Li et al. 2013). In the presence of the proteasomal inhibitor MG132, degradation of Nrf2 is inhibited, and has shown to be significantly elevated in a number of studies (He, Chen et al. 2008, Li, Paonessa et al. 2012). As Nrf2 is no longer degraded its presence should be detectable, therefore giving a positive control for MG132 efficiency. In this experiment, detectable levels of Nrf2 in cells treated with MG132 in comparison to untreated would act as indication that MG132 was having an effect. (Li, Paonessa et al. 2012, Cui, Bai et al. 2013, Zhou, Sun et al. 2014)

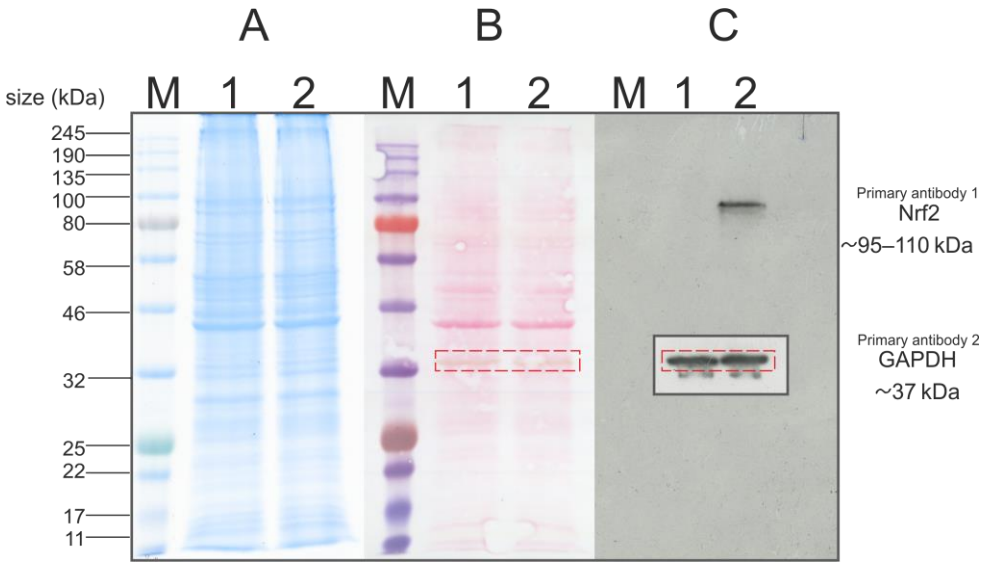


Figure 72 MG132 induced Nrf2 activation detected by western blot analysis. Upregulation of Nrf2 was detected in H9c2 rat cardiomyocytes following 2 hour incubation. Lanes show Marker (M), lane 1 (1) control (whole cell lysate absent of inhibitor) and lane 2 (2) (whole cell lysate incubated with inhibitor). Total protein content was assessed by a BCA assay to ensure equal loading. Image shows same step process of gels resolved by electrophoretic methods and stained with coomassie blue (A) indicating equal load (qualitative) ponceau S of the same gel following destain (B) to show equal transfer to the PVDF membrane (C) X-ray film to show upregulated Nrf2 in comparison to control (1) using antibody 1, where equal loading was confirmed by detection of housekeeping gene Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) (antibody 2). No cross contamination was detected as there were no detectable levels of Nrf2 in the control lane (lane1 X-ray film).

As illustrated in Figure 72C, Nrf2 was completely undetectable in untreated H9c2 cells. A marked increase in detectable levels of Nrf2 is noted following a 2 hour treatment with MG132. This observation is in agreement with the previous studies mentioned above. It is important to note here, that while Nrf2 has a calculated molecular weight of Nrf2 68 kDa, it has been consistency shown to migrate at 95-100 kDa due poly ubiquitination within the cell (Lau,

Tian et al. 2013). Equal loading amounts were confirmed by Immunoblotting GAPDH. Overall these results confirm the efficiency of the inhibitor.

Having determined the efficacy of MG132 (Figure 72), and already determined the ability of SMA to internalise (Figure 69) we next wished to assess if internalisation of SMA could be abolished in the presence of this inhibitor. Here, 10 µM SMA was incubated with H9c2 cells in the presence of MG132 for 24 hours. Unfortunately, internalisation of SMA appears in perturbed by the presence of MG132 at this concentration (Figure 71).

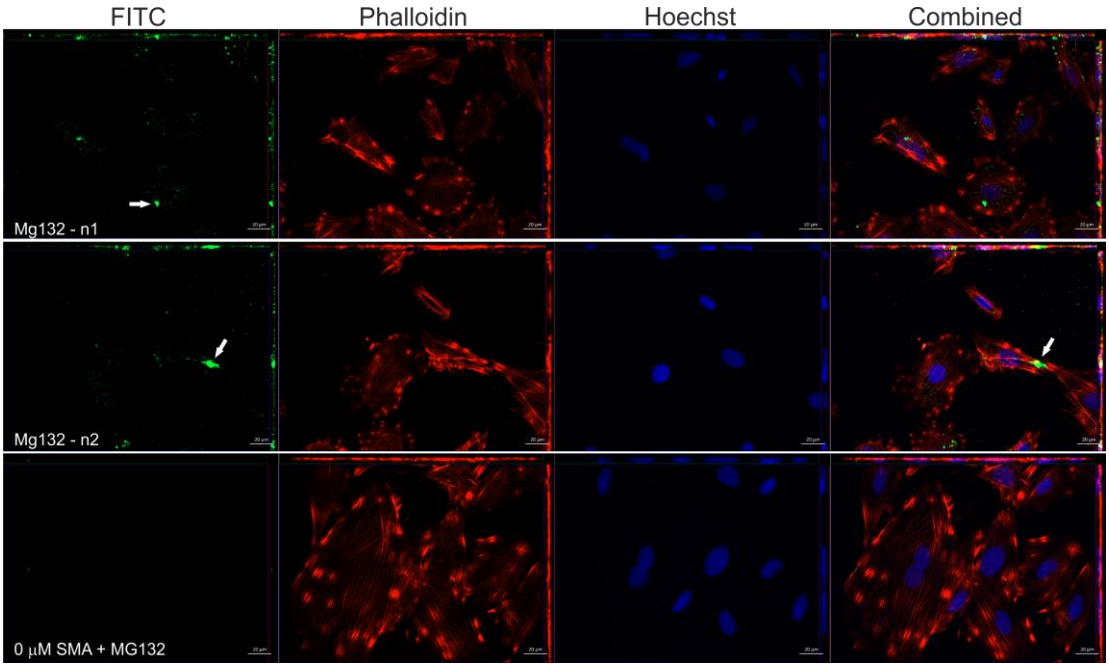


Figure 73 2D confocal microscopy of H9c2 cells incubated with SMA, and MG132. In two independent experiments (indicated as n1 and n2), FITC labelled SMA (green) at 10 µM were incubated with rat H9c2 cardiomyocytes in the presence of 5 µmol/L MG132 for 24 hours. Detection of the FITC signal was made by maximum intensity projections and Z-stack analysis. A control experiment (Ctrl) that was free of MG132 (still in the presence of SMA aimed to assess the ability of MG132 to alter cell morphology) was conducted. Combined shows all channels (blue hoechst - nuclei, red phalloidin - F-actin and green – FITC labelled LEN). Scale bar is 20 µm. Asterisk indicates image that was taken for further processing. Arrows show intense fluorescent pixels.

6.3.11. Main findings and summary

Summary

The mechanisms by which V_Ls are able to internalise remain largely unknown, and to date only a handful of light chains have been demonstrated to internalise into a number of different cell types. In this chapter, we sought to establish if the amyloidogenic SMA and the non-amyloidogenic LEN could internalise into H9C2 rat cardiomyocytes. We demonstrate that the recombinant sources of protein are suitable for fluorescent detection of κIV variable domains by conjugating highly pure fractions of SMA and LEN with FITC and that this cell line internalises both light chains where they appear to localise to the perinuclear area, assessed by z-stack confocal microscopy.

Highlights

At concentrations where each light chain is mostly monomeric, we demonstrate that each is able to internalise after a period of 24 hours where they appear to localise to the perinuclear area, as assessed by z-stack confocal microscopy possibly associating with the endoplasmic reticulum. Cell viability data derived from a CCK8 assay indicates that SMA at a concentration of 5 μM is the most toxic (Figure 71). This is presumably due to the larger population of monomers (which are documented to be more aggregation prone than dimers) in solution at this concentration. We do not detect any differences in the morphology of cells (no observable changes to the F-actin cytoskeleton of the H9C2 cardiomyocytes) which could indicate that these proteins are inducing cell death. However, as the results of the CCK8 assay indicate that SMA is cytotoxic there could be changes to the cellular ultrastructure which

are undetectable at this magnification (Monis, Schultz et al. 2006, Levinson, Olatoye et al. 2013). Aside from one cell that appears to be undergoing apoptosis (Figure 64 LEN- 5 μM) which cannot be associated with the presence of light chain, there were no changes in cellular architecture, where blebbing of the membrane has been previously associated with internalisation (Trinkaus-Randall, Walsh et al. 2005). High magnification techniques such as TEM could provide a more detailed insight into structural changes (Lavatelli, Imperlini et al. 2015) and further studies using metabolomics or proteomic approaches could shed light on changes to the cellular proteome and other biochemical alterations as a consequence of light chain internalization (Imperlini, Gnecchi et al. 2017).

The intracellular location of these proteins is consistent with other light chain internalisation studies (Trinkaus-Randall, Walsh et al. 2005, Monis, Schultz et al. 2006) and there is also evidence to suggest possible lysosomal expansion (Figure 70 high Len) which is a similar observation to those made by *Levinson et al*., ((Levinson, Olatoye et al. 2013)). However, cells did not display any signs of clustering in comparison to control (untreated) which was also observed in this same study (Levinson, Olatoye et al. 2013).

Earlier work has shown that internalisation of amyloidogenic light chains in rat cardiac fibroblasts and human cardiomyocyte proceeds through pinocytosis (Monis, Schultz et al. 2006, Marin-Argany, Lin et al. 2016). The study conducted by *Levinson et al.,* using mouse cardiomyocytes confirms this hypothesis, demonstrating that Oregon green conjugated AL-09 (amyloidogenic) and κI O18/O8 (non-amylodigeonic) were internalised after a 24 hour exposure using live cell imaging. The distribution of punctate signals in

our study is consistent with previous observations (Levinson, Olatoye et al. 2013) however, we were unable to demonstrate any association with endo-lysosomal vesicles, where there was no co-localisation between FITC and lysotracker signals. Any endocytic pathway the light chain may be internalised through (clathrin mediated and clathrin independent, caveolae, transerin, macro pinocytosis) would be expected to proceed through vesicles of low pH and so, we would expect to see co-localisation signals with acidic pH markers – in this case lysotracker red. This however is not the case. There are multiple possibilities for this observation. Since subtle changes to the primary amino acid sequence of a variable domain is able to perturb the rates of internalisation, it is possible that SMA and LEN exhibit slower internalisation kinetics in comparison to AL-09 and κI O18/O8 which did show co-localisation, and so the quantities of internalised protein may be insufficient to facilitate the staining patterns shown in other publications. Alternatively, FITC, the pH sensitive fluorophore used in this study may not be as stable as other fluorophores such as Oregon green, and so (WHAT).

An additional possibility is that the uptake of these light chains may be interfering with normal endocytotic pathways causing the premature release of endo-lysomal contents into the cellular cytosol (Repnik, Česen et al. 2013). Another possible explanation is that following the 24-hour incubation and 2 hour processing time before paraformaldehyde fixation, is that these light chains have already been processed through the endo-lysosomal pathway. Therefore it would be of interest to monitor the internalisation of these light chains in a time-resolved manner.

7. Summary and biological implications

7.1. Disease summary

AL amyloidosis is often viewed as a rare disorder (Sanchorawala 2006). Yet, the incidence (1 in every 100,000) (Pelaez-Aguilar, Rivillas-Acevedo et al. 2015) is not too dissimilar from motor neurone disease which has received far more attention in recent years (McDermott and Shaw 2008). A wide spectrum of clinical manifestations, and general clinical unawareness, means that the average life expectancy of this disorder without treatment can be as little as 6 months (Kastritis and Dimopoulos 2016, Grogan, Dispenzieri et al. 2017). The disease is characterised by an underlying plasma cell clone that moderately infiltrates the bone marrow and synthesises light chain fragments in enormous quantities. Some of these light chain fragments are able to misfold, and deposit in multiple organs leading to dysfunction and death. Not all light chains are however amyloidogenic. Comparisons made between amyloidogenic and non-amylogideonic V_Ls have shown that somatic mutations occurring in specific structural locations are a crucial factor in determining amyloidogenicity. Cataloguing the thermodynamic burden of a mutation, and important structural changes that are likely contributors to an increased aggregation potential is of particular importance in a disease that has such enormous sequence and mutational diversity (Randles, Thompson et al. 2009). Yet, very few studies provide the much-needed structural link to how a mutation confers enhanced aggregation propensity. As a consequence, the exact molecular mechanisms that render a V_L amyloidogenic are remain poorly understood (Hernández-Santoyo, del Pozo Yauner et al. 2010).

On a separate, but equally as important issue, is the accumulating evidence that suggest that variable domains can internalise and cause cytomorphological and biochemical changes within cells. In this thesis, we focus on both of these topics and adopt a multidisciplinary approach, to work towards a better understanding of AL amyloidosis using the three κIV V_L domains SMA, REC, and LEN as model proteins.

7.2. Major findings

To date, no crystallographic information of SMA is available, and so the orientation of the dimer (canonical or non-canonical example in Figure 10) is currently unknown. The first experimental chapter (Chapter 2) takes a computational approach to decipher the orientation of SMA, by generating dimer models based on the crystal structure of known homologs. Interfacial analysis of generated structures indicates that SMA is more favourable as a non-canonical dimer. As preventing dissociation of the dimer is of key therapeutic interest, the druggability of the newly generated SMA homodimer was calculated by DogSiteScorer. A shallow cavity present in both SMA and REC that scored well may accommodate small molecule stabilisers as future therapeutic avenues.

To pinpoint which SMA or REC-like mutation(s) are capable of dictating the orientation of the dimer, each were introduced into the native structure of LEN by *in silico* mutagenesis and the changes in the energetic and geometrical properties were compared to the native structure of LEN. We found that out of all mutations, only residues K30T (REC-like), Q89H and Y96Q (SMA-like) were able to dictate the orientation of the dimer.

Several lines of evidence suggest that AL proteins are less stable than their non-amyloidogenic counterpart's due to the accumulation of somatic mutations in CDR and FR regions of the V_L. The way in which a somatic mutation can destabilise the native V_L can proceed through multiple independent mechanisms that include a) destabilise the dimer which may lead to an increased population of aggregation prone monomer in solution b) globally destabilise the thermodynamic stability of the native V_L monomer which may allow the protein to sample partially unfolded states which are suggested to be critical for aggregation c) inducing structural changes that are localised to segments of the protein which can enhance aggregation propensity independent of thermodynamic stability. The latter point is slightly more challenging, and requires a more sophisticated approach than merely calculating the change in the Gibbs free energy of binding and folding. As the outcome of each SMA and REC-like mutation remain largely unknown, chapters 2, 3 and 4 describe the use of numerous digital platforms aimed to assess the nature of a mutation where, we find that the CDR3, a loop located at the dimer interface and harbouring residues Q89 and Y96 is key to modulating the aggregation potential of the protein, the results of these chapters have been published (Mukherjee, Pondaven et al. 2017).

Another goal (documented in Chapter 5) was to build upon the existing work of Immunoglobulin LC by *E.coli* synthesis, and provide a refined and a reproducible protocol of SMA and LEN expression that provides high yields suitable for experiments that are demanding of high protein concentrations. Recent publications and correspondence with collaborators have discussed the sub-optimal yields and low levels of purity in expression trials which did not

match the original study. Using a limited step chromatographic technique, we are able to obtain V_L's of high purity, and include data on secondary structure and fold assessment. This process is thoroughly documented, and we believe may be applicable to other difficult to expression light chain variable domains.

Recently, the ability of V_L from the λ family to internalise and induce toxic effects has been confirmed. The purpose of chapter 6 is to adopt the process of these studies to assess the ability of V_Lκ4 domains to internalise which is currently unknown. In this chapter (6) we demonstrate using z-stack confocal microscopy that Fluorescein conjugated V_Ls SMA and LEN are internalized in H9C2 rat cardiomyocytes within 24 hours. Internalized protein exhibits a punctate reminiscent of pinocytosis and appears to be located in the peri-nuclear region of the cell. Out of these two proteins, only SMA exhibited cytotoxicity. The results of this chapter in combination with chapter 5 are forming an additional publication.

The following sections further detail interesting aspects of this work.

7.1.1. Stabilising the quaternary arrangement of V_L-V_L homodimers: therapeutic promise

The majority of V_L-V_L homodimers have crystallised in a structural arrangement similar to the light and heavy variable domains (V_L and V_H) of the antigen-binding fragment (Figure 8 and Figure 10A) (Padlan 1994, Stevens and Schiffer 1995, Peterson, Baden et al. 2010). From the V_Ls in this study, only REC and LEN have been crystallised, forming non-canonical and

canonical dimers respectively. By using these structures as templates for *in silico mutagenesis*, each were converted so they contained the primary sequence of SMA, subject to geometry optimisation to allow for the best fit of the newly introduced side chain rotamer, and the energetic features of dimer interface calculated. From this analysis, SMA (which has not been solved by high resolution techniques such as X-ray crystallography or NMR) preferentially adopts a twisted 180 ^{o} altered dimer interface that is similar to REC.

This procedure has several distinct advantages over the routinely used methods (NMR and X-ray crystallography) of acquiring quaternary structural arrangements of V_L-V_L homodimers. Firstly, it bypasses the main hurdle associated with high-resolution techniques, which is the demand for high quantities of pure target protein. This is particular challenging in AL amyloidosis, where each patient possesses a V_L with a unique amino acid sequence. When a protein source is no longer available directly from the patient, recombinant procedures are used. A number of publications have demonstrated that the V_L is often found in inclusion bodies and methods of refolding often provide suboptimal yields. Secondly, there are examples where the harsh conditions present in X-ray crystal screens have been able to falsely dictate the dimer orientation (Peterson, Baden et al. 2010).

Aside from bypassing some of the commonly encountered hurdles associated with these experimental methods described above, from a bioinformatics point of view, the procedure is not computationally demanding, can be performed locally, and requires only freely available open source software. Secondly, it does not require a large library of information with

known outcomes. Rather, this method takes a structural approach and simply identifies the energetic favourability of a particular pose.

Current AL therapies focus only on supressing the plasma cell dyscrasia (Pelaez-Aguilar, Rivillas-Acevedo et al. 2015) and the majority of emerging strategies aim to remove resident amyloid (Grogan, Dispenzieri et al. 2017). Yet it is important to note that removing amyloid from the affected organ results in the formation of non-functional scar tissue. We mention here again, that stabilisation of the V_L homodimer presents arguably the most promising alternative to the current strategies of treatment in AL amyloidosis, which often have high toxicity and an associated poor quality of life.

The process documented within this thesis of generating an SMA model, establishing its orientation and identifying druggable pockets within the structure that may be able to accommodate small molecule stabilisers in an entirely computational approach raises the interesting and exciting possibility such a scheme may be applicable to other uncharacterised V_L homodimers where only the amino acid sequence is available. Given the ability of mass spectrometry to rapidly identify the amino acid sequence of a patient derived V_L from serum or fat aspirate samples, there is also the prospect that such a strategy, may allow for the future development of personalised medicines. At the very least, it would allow for some of the hurdles of acquiring experimental data to be circumvented and significantly narrow down screening trials to more relevant small molecules.

The focus should now be, assessing other non-canonical V_L-homodimers to see if druggable pockets with high druggability scores (assessed by DogSiteScorer) such as the ones highlighted in this thesis are

conserved. Further data sets are needed to confirm the efficacy of the process, specifically; performing blind tests to see if this method can confirm the orientation of the dimer using patient derived sequences where the structure has been solved using a high-resolution technique would be particularly useful and add confidence to these findings. Once a significant data set has been reached, then the use of docking programs would be the next logical step in the progression of this work.

Recent data suggests that V_L homodimers can interconvert between different quaternary states that lay within a global energy minima. It is yet unclear if this is common amongst all homodimers, if the local environmental conditions can dictate orientation (Novotny and Haber 1985, Baden, Owen et al. 2008), and what proportion of V_L present in solution occupies a particular state at any given time. To the best of our knowledge, It seems that no study other has addressed this dynamic nature of V_H and V_L domains. An additional advantage of using a structural bioinformatics approach for this scenario is the ability to rapidly assess the energetic features of each quaternary state, determine the most preferential, and choose to identify pockets that can accommodate small molecules that aimed to stabilise this energetically favourable form, rather than attempting to stabilise an energetically less stable state.

In further studies, we aim to pursue identifying the orientation of SMA by further crystallographic screening trials although, with this information in mind an alternative method such as small-angle X-ray scattering (SAXS) may allow the orientation of the SMA dimer (amongst other V_L-V_L dimers) to be resolved in a more physiologically relevant setting. The results of this study will

either confirm or refute our analysis, but fundamentally assesses the accuracy of this computational strategy, and the predictive power of tools available to date.

7.1.2. Single mutations can dictate the quaternary state of the V_L homodimer

The present *in silico* strategy is able to predict which of the SMA or REC-like mutants are able to dictate the arrangement of the LEN dimer. By using the experimentally resolved structure of LEN in two different quaternary states (canonical and 180 ^{o} twisted) as a scaffold, single SMA or REC like mutations were incorporated, and the consequence of each mutation to the geometric and energetic features of the mutation calculated. While many of the mutations were incapable of significantly destabilising the canonical LEN interface, PISA was able to identify those interfacial interactions and energetic values were markedly lower in structures harbouring single mutants K30T (REC-like), Q89H and Y96Q (SMA-like) than that in the homodimeric complex of LEN. In the non-canonical structure of LEN (which is said to be non-physiologically relevant) K30T and Q89H were able to stabilise this form suggesting these are key residues that can dictate the dimer orientation. Y96Q however, showed to impair the non-canonical structure as well as destabilising the canonical form. By comparing $\Delta G_{unfolding}$ (acquired from previous experimental values) and $\Delta G_{dissociation}$ values (acquired from PISA) as well as a structural assessment for Y96Q (both by Rosetta and NMR based) this mutation may dictate dissociation of the dimer into the monomeric form (Fred J. Stevens 1980, Blancas-Mejia, Misra et al. 2017). In this state, the mutation may result in a poorly packed

monomer that has enhanced conformational dynamics and is more liable to unfolding and subsequent aggregation.

Despite the inconsistency of the 5 programs (*FoldX*, *mCSM PPI*, *ELASPIC*, *MutaBind* and *BeAtMusic*) used to detect changes in binding energies as a consequence of the mutations, we report that mutations occurring at position Y96 were consistently shown to be destabilising. This provides rationale for using PISA as a predictive tool for assessing the impact of a mutation on the energetic features of the dimer interface. The ability of PISA to identify the destabilising nature of Q89H and Y96Q was strongly supported by its ability to also detect the destabilising nature of K30T, which has shown to dictate the orientation of the dimer by experimental methods (Pokkuluri, Cai et al. 2000). As dimerisation confers resistance to the more amyloid prone monomer form, the procedure documented here could be used a starting point for rapidly assessing the outcome of a single mutation, not only on the structural integrity and thermodynamic stability, but also its ability to dictate the orientation of dimer with a focus on stabilising this state.

7.1.2. The thermodynamic analysis of SMA and REC-like mutations

Changes to the folding free energies of the variable domain LEN for each SMA-like and REC-like mutation has not been completely determined by experimental methods and is why it is a focus of this thesis. During the assessment of changes to the folding free energies as a consequence of mutation, we encountered several residues that had ambiguous calculations from the digital platforms used. To get a reliability estimate, calculations were compared to existing experimental data, where discrepancies were also

apparent. We conclude here, that unfortunately, these platforms are unable to accurately assess the outcome of a particular mutation.

Instead, assessing the outcome of a mutation requires a more sophisticated approach. Compiling existing experimental data (for the V_L domains at the basis of this study amongst other V_L domains), and using a combined application of *in silico* methods and ^{1}H-^{15}N NMR spectroscopy to acquire R_{ex} values of SMA-like mutations presented within this study has provided a thorough and complete analysis of the thermodynamic stability, conformational flexibility and the ability to identify structural changes at atomic level detail. This work has not been published (see (Mukherjee, Pondaven et al. 2017)

We hypothesised that some of the changes in thermodynamic stability may originate from the disruption of one or more intra, and inter molecular hydrogen bonds. Indeed, we noted that removing Q89 and the highly conserved Y96 of the (FR3 and CDR3 respectively) lead to disruptions in conserved hydrogen bonds across the interface as well as conformational changes in the structure near to the mutation and additional sites in key regions distal to the mutation.

7.1.3. Expression of V_Ls SMA and LEN and light chain internalisation

Herein, a fully documented and reproducible protocol for the expression and purification SMA and LEN is made, overcoming previously documented difficulties that included low yields and low purity. Protein produced using these methods will be continued to be used as test for anti-fibrillation molecules by Dr Wolfgang Hoyer as well as a model set of V_Ls for assess the

ability of catalytic antibodies to reduce amyloid deposition in a recent partnership with UCB pharma.

From this, we investigated another topic much in the spotlight of AL amyloidosis, which is the ability of light chain variable domain to internalise and induce cytomorpholoical and biochemical changes that cause toxicity. There are a limited number of studies that assess this possibility, and so the goal of this chapter was to evaluate the internalisation ability of two light chain variable domains of the kappa isotype which have not been explored previously. Using, FITC conjugated SMA and LEN and the popular rat cardiac cell line, the main discovery was that internalisation was demonstrated for both proteins. Now established, we aim to move to live cell imaging methods using, human cardiac cells, and finally human 3D culture models, providing higher degrees of physiological relevance to our findings, where we will continue to assess preventing internalisation using pharmacological inhibition assays.

7.2. Limitations to the study

Less stable proteins (by measurement of their thermodynamic stability) are reputed to be more susceptible to fibril formation (Hurle, Helms et al. 1994, Ramirez-Alvarado 2007, Ramirez-Alvarado, De Stigter et al. 2007). Mutations that reduce the thermodynamic stability of a protein may compromise its fold and allow it to sample partially unfolded states that are key in amyloid formation (Rochet and Lansbury 2000, Ramirez-Alvarado, De Stigter et al. 2007, Del Pozo-Yauner, Becerril et al. 2015). In nature, destabilisation of the native structure is a likely consequence of all mutations (DePristo, Weinreich et al. 2005, Tokuriki, Stricher et al. 2008, Bromberg and Rost 2009, Soskine

and Tawfik 2010, Araya, Fowler et al. 2012, Studer, Christin et al. 2014). Immunoglobulins are commonly subject to mutations during maturation, specifically complementarity determining regions, sections of the variable domain that undergoes hyper mutation in order to enhance their antigen affinity (Kastritis and Dimopoulos 2016). Mutations in these CDR regions are often negligible to antibody stability, or acquire stabilising mutations to compensate for the destabilising nature of another (Jolly, Wagner et al. 1996, Julian, Li et al. 2017). However, In AL amyloidosis, the light chain variable domain is synthesised independent from any other Immunoglobulin subunit. Thus, the effect of a single mutation is likely to be amplified on this 12.8 kDa molecule in comparison to the full antibody that has a molecular weight of ~150 kDa. It is however, becoming apparent that a decrease in thermodynamic stability does not necessarily indicate enhanced aggregation potential. For example, *Marta Marin-Argany et al.,* (previously mentioned in this thesis) revealed that H32Y was less stable than one of the most destabilising mutations in this set, yet this mutation was unable to form amyloid fibrils instead forming nontoxic amorphous aggregates (Marin-Argany, Guell-Bosch et al. 2015). Interestingly, other variable domains that have been relatively insensitive to mutations (leading to only small changes to unfolding free energies) have demonstrated the ability to form amyloid, suggesting that local structural changes may increase aggregation propensity independent of large global destabilising effects (Raffen, Dieckman et al. 1999). These factors make it incredibly challenging to apply a thermodynamic stability threshold or cut-off to predict whether a mutation can lead to amyloid formation or not.

We use a number of predictors in this study to infer which mutation may be able to increase/ decrease aggregation potential by predicting changes to binding or folding free energies. Unfortunately, we were unable to unambiguously infer the stability effects of these mutations, which we acknowledge as a limitation of this study. Furthermore, the algorithms used for these predictions themselves, also have limitations that have been described previously in chapters and inaccuracies in predictors previously described in the literature (Guerois, Nielsen et al. 2002, Kumar, Rahman et al. 2017) To evaluate the predictive accuracy of each and every program used in this thesis with the aim of highlight the most accuracy (which this study does not focus on, rather we wish to assess which mutations likely contribute to increased aggregation potential) a far more extensive set than the mutations presented here, an expanded number of different programs that used vastly different datasets, and a number of different 3 dimensional structures of different resolutions that had been acquired in different conditions (evaluate structural sensitivity) would be needed to more accurate assumptions, similar to the following publication (Kumar, Rahman et al. 2017). Alternatively, a molecular dynamics approach could be used to how single substitutions effect dimer and monomer stability.

In the attempt to overcome these limitations and avoid placing assumptions on how changes to the net stability influences protein aggregation we proceeded to dissect each SMA and REC-like mutation in the attempt to find key “gatekeeper residues”, those that are critical for maintaining not only the stability of the monomer, but also of the V_L homodimer. To one extent, this has been successful. By calculating changes to the 3d-dimensional

structure and combining this work with previously unpublished NMR data, we have indicated that the CDR3 is a key region within the variable domains in thus study which mediate their aggregation potential (Mukherjee, Pondaven et al. 2017).

Limitations in a technical aspect include the work described in chapter 6. This chapter wishes to assess the internalisation of SMA and LEN in to H9C2 rat cardiomyocytes. Internalisation was noted however one possible caveat to this study is the possibility that FITC (the fluorophore that enabled us to detected internalised V_L) may perturb the kinetics of cellular uptake, and even dictates the subcellular localisation of internalised material (Mulcahy, Pink et al. 2014). Rather than express each V_L to a fusion protein such as green fluorescent protein which may have impeded dimerization, stability, and native properties of each V_L due to its large size (26.9 kDa) we chose to label each variable domain post-translationally. FITC and the analog Oregon green 488 have been extensively used in internalisation studies (Monis, Schultz et al. 2006, Wang, Li et al. 2006, Perera, Zoncu et al. 2007, Morris, Craig et al. 2009, Levinson, Olatoye et al. 2013, Marin-Argany, Lin et al. 2016) yet there have been evidence showing that some Fluoresceins are able to alter the subcellular localisation of some molecules upon uptake (Puckett and Barton 2009). Given the ability of minor modifications to the primary structure has shown to perturb internalisation rates of variable domains (Levinson, Olatoye et al. 2013), FITC labelling could also have appreciable changes to the rate of internalisation within this study. Further experiments using fluorescein alternatives would elucidate the influence of FITC on internalisation and location. Having demonstrated these cells to internalise into a monoculture of

rat H9C2 cardiomyocytes, the use of human and 3-dimensional cultures would also provide greater physiological relevance to this study.

7.3. Concluding statement

Currently no drug has received regulatory approval for AL amyloidosis. In taking a multi-dimensional approach and complementing experimental methods with computational work, we address some of the gaps in current literature. Specifically, we sought to enhance our understanding on the underlying mechanisms of LC V_L amyloid formation that include internalisation, and the contributions of somatic mutations to pathogenesis. By focussing on two emerging hot topics of potential therapeutic targets, (stabilisation of the dimer and inhibition of internalisation), we hope the findings of this work detailed in this thesis will be built upon and present a wealth of knowledge in order to make the much-needed therapeutic advances for devastating and currently incurable disease.

I thank Dr Marina Alvarado Ramirez, Dr Diana Penha and Dr Daniel Rigden for stimulating discussion around the points addressed in this summary section.

References

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

Abraham, R. S., S. M. Geyer, T. L. Price-Troska, C. Allmer, R. A. Kyle, M. A. Gertz and R. Fonseca (2003). "Immunoglobulin light chain variable (V) region genes influence clinical presentation and outcome in light chain-associated amyloidosis (AL)." Blood **101**(10): 3801-3808.

Adamcik, J., J.-M. Jung, J. Flakowski, P. De Los Rios, G. Dietler and R. Mezzenga (2010). "Understanding amyloid aggregation by statistical analysis of atomic force microscopy images." Nat Nano **5**(6): 423-428.

Addgene. (2017). "Bacterial transformation." Retrieved 15/08, 2017, from https://www.addgene.org/protocols/bacterial-transformation/.

Aeschimann, W., S. Staats, S. Kammer, N. Olieric, J. M. Jeckelmann, D. Fotiadis, T. Netscher, G. Rimbach, M. Cascella and A. Stocker (2017). "Self-assembled alpha-Tocopherol Transfer Protein Nanoparticles Promote Vitamin E Delivery Across an Endothelial Barrier." Sci Rep **7**(1): 4970.

Ahlstrom, L. S., I. I. Vorontsov, J. Shi and O. Miyashita (2017). "Effect of the Crystal Environment on Side-Chain Conformational Dynamics in Cyanovirin-N Investigated through Crystal and Solution Molecular Dynamics Simulations." PLoS ONE **12**(1): e0170337.

Anfinsen, C. B. (1973). "Principles that govern the folding of protein chains." Science **181**(4096): 223-230.

Anfinsen, C. B. and E. Haber (1961). "Studies on the reduction and re-formation of protein disulfide bonds." J Biol Chem **236**: 1361-1363.

Ara Celi DiCostanzo, M. R.-A. (2011). Current and New Perspectives on the Molecular and Cellular Mechanisms of Amyloid Formation and Toxicity in Light Chain Amyloidosis, Amyloidosis - Mechanisms and Prospects for Therapy, InTech.

Ara Celi DiCostanzo, M. R.-A. (2011). "Current and New Perspectives on the Molecular and Cellular Mechanisms of Amyloid Formation and Toxicity in Light Chain Amyloidosis, Amyloidosis - Mechanisms and Prospects for Therapy." InTech.

Araya, C. L., D. M. Fowler, W. Chen, I. Muniez, J. W. Kelly and S. Fields (2012). "A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function." Proc Natl Acad Sci U S A **109**(42): 16858-16863.

Arosio, P., T. P. J. Knowles and S. Linse (2015). "On the lag phase in amyloid fibril formation." Physical Chemistry Chemical Physics **17**(12): 7606-7618.

Arosio, P., M. Owczarz, T. Muller-Spath, P. Rognoni, M. Beeg, H. Wu, M. Salmona and M. Morbidelli (2012). "In vitro aggregation behavior of a non-amyloidogenic lambda light chain dimer deriving from U266 multiple myeloma cells." PLoS One **7**(3): e33372.

Astbury, W. T., S. Dickinson and K. Bailey (1935). "The X-ray interpretation of denaturation and the structure of the seed globulins." Biochem J **29**(10): 2351-2360.2351.

Baden, E. M., B. A. Owen, F. C. Peterson, B. F. Volkman, M. Ramirez-Alvarado and J. R. Thompson (2008). "Altered dimer interface decreases stability in an amyloidogenic protein." J Biol Chem **283**(23): 15853-15860.

Baden, E. M., E. G. Randles, A. K. Aboagye, J. R. Thompson and M. Ramirez-Alvarado (2008). "Structural Insights into the Role of Mutations in Amyloidogenesis." The Journal of Biological Chemistry **283**(45): 30950-30956.

Baden, E. M., L. A. Sikkink and M. Ramirez-Alvarado (2009). "Light chain amyloidosis - current findings and future prospects." Curr Protein Pept Sci **10**(5): 500-508.

Baker, K. R. and L. Rice (2012). "The Amyloidoses: Clinical Features, Diagnosis and Treatment." Methodist DeBakey Cardiovascular Journal **8**(3): 3-7.

Ban, T. and Y. Goto (2006). "Direct observation of amyloid growth monitored by total internal reflection fluorescence microscopy." Methods Enzymol **413**: 91-102.

Barrett, P. J. and J. Timothy Greenamyre (2015). "Post-translational modification of alpha-synuclein in Parkinson's disease." Brain Res **1628**(Pt B): 247-253.

Basnayake, K., S. J. Stringer, C. A. Hutchison and P. Cockwell (2011). "The biology of immunoglobulin free light chains and kidney injury." Kidney International **79**(12): 1289-1301.

Bayliss, M., K. L. McCausland, S. D. Guthrie and M. K. White (2017). "The burden of amyloid light chain amyloidosis on health-related quality of life." Orphanet Journal of Rare Diseases **12**(1): 15.

Bell, C. E. and M. Lewis (2000). "A closer view of the conformation of the Lac repressor bound to operator." Nat Struct Biol **7**(3): 209-214.

Bendor, J. T., T. P. Logan and R. H. Edwards (2013). "The function of alpha-synuclein." Neuron **79**(6): 1044-1066.

Berkmen, M. (2012). "Production of disulfide-bonded proteins in Escherichia coli." Protein Expression and Purification **82**(1): 240-251.

Berliner, N., J. Teyra, R. Colak, S. Garcia Lopez and P. M. Kim (2014). "Combining structural modeling with ensemble machine learning to accurately

predict protein fold stability and binding affinity effects upon mutation." PLoS One **9**(9): e107353.

Beyer, K. and A. Ariza (2013). "alpha-Synuclein posttranslational modification and alternative splicing as a trigger for neurodegeneration." Mol Neurobiol **47**(2): 509-524.

Bhavaraju, M. and U. H. Hansmann (2015). "Effect of single point mutations in a form of systemic amyloidosis." Protein Sci **24**(9): 1451-1462.

Biancalana, M. and S. Koide (2010). "Molecular mechanism of Thioflavin-T binding to amyloid fibrils." Biochim Biophys Acta **1804**(7): 1405-1412.

Bieschke, J., J. Russ, R. P. Friedrich, D. E. Ehrnhoefer, H. Wobst, K. Neugebauer and E. E. Wanker (2010). "EGCG remodels mature alpha-synuclein and amyloid-beta fibrils and reduces cellular toxicity." Proc Natl Acad Sci U S A **107**(17): 7710-7715.

Blancas-Mejia, L. M., J. Hammernik, M. Marin-Argany and M. Ramirez-Alvarado (2015). "Differential effects on light chain amyloid formation depend on mutations and type of glycosaminoglycans." J Biol Chem **290**(8): 4953-4965.

Blancas-Mejía, L. M., T. J. Horn, M. Marin-Argany, M. Auton, A. Tischer and M. Ramirez-Alvarado (2015). "Thermodynamic and Fibril Formation Studies of Full Length Immunoglobulin Light Chain AL-09 and its germline protein using Scan Rate Dependent Thermal Unfolding." Biophysical chemistry **207**: 13-20.

Blancas-Mejia, L. M., P. Misra and M. Ramirez-Alvarado (2017). "Differences in Protein Concentration Dependence for Nucleation and Elongation in Light Chain Amyloid Formation." Biochemistry **56**(5): 757-766.

Blancas-Mejia, L. M. and M. Ramirez-Alvarado (2016). "Recruitment of Light Chains by Homologous and Heterologous Fibrils Shows Distinctive Kinetic and Conformational Specificity." Biochemistry **55**(21): 2967-2978.

Blancas-Mejía, L. M. and M. Ramirez-Alvarado (2013). "Systemic Amyloidoses." Annual review of biochemistry **82**: 745-774.

Blancas-Mejia, L. M., L. A. Tellez, L. del Pozo-Yauner, B. Becerril, J. M. Sanchez-Ruiz and D. A. Fernandez-Velasco (2009). "Thermodynamic and Kinetic Characterization of a Germ Line Human λ6 Light-Chain Protein: The Relation between Unfolding and Fibrillogenesis." Journal of Molecular Biology **386**(4): 1153-1166.

Blancas-Mejia, L. M., A. Tischer, J. R. Thompson, J. Tai, L. Wang, M. Auton and M. Ramirez-Alvarado (2014). "Kinetic control in protein folding for light chain amyloidosis and the differential effects of somatic mutations." J Mol Biol **426**(2): 347-361.

Boccadoro, M. and A. Pileri (1995). "Plasma cell dyscrasias: classification, clinical and laboratory characteristics, and differential diagnosis." Baillieres Clin Haematol **8**(4): 705-719.

Bosco, D. A., D. M. Fowler, Q. Zhang, J. Nieva, E. T. Powers, P. Wentworth, Jr., R. A. Lerner and J. W. Kelly (2006). "Elevated levels of oxidized cholesterol metabolites in Lewy body disease brains accelerate alpha-synuclein fibrilization." Nat Chem Biol **2**(5): 249-253.

Boyko, K. M., T. V. Rakitina, D. A. Korzhenevskiy, A. V. Vlaskina, Y. K. Agapova, D. E. Kamashev, S. Y. Kleymenov and V. O. Popov (2016). "Structural basis of the high thermal stability of the histone-like HU protein from the mollicute Spiroplasma melliferum KC3." Scientific Reports **6**: 36366.

Bradley, P., K. M. Misura and D. Baker (2005). "Toward high-resolution de novo structure prediction for small proteins." Science **309**(5742): 1868-1871.

Brenner, D. A., M. Jain, D. R. Pimentel, B. Wang, L. H. Connors, M. Skinner, C. S. Apstein and R. Liao (2004). "Human amyloidogenic light chains directly impair cardiomyocyte function through an increase in cellular oxidant stress." Circ Res **94**(8): 1008-1010.

Bromberg, Y. and B. Rost (2009). "Correlating protein function and stability through the analysis of single amino acid substitutions." BMC Bioinformatics **10**(Suppl 8): S8-S8.

Broom, H. R., J. A. O. Rumfeldt, K. A. Vassall and E. M. Meiering (2015). "Destabilization of the dimer interface is a common consequence of diverse ALS-associated mutations in metal free SOD1." Protein Science : A Publication of the Protein Society **24**(12): 2081-2089.

Bruce, A., J. Alexander, L. Julian, R. Martin, R. Keith and W. Peter (2002). Molecular Biology of the Cell, Garland Science.

Brummitt, R. K., J. M. Andrews, J. L. Jordan, E. J. Fernandez and C. J. Roberts (2012). "Thermodynamics of amyloid dissociation provide insights into aggregate stability regimes." Biophysical Chemistry **168**: 10-18.

Brumshtein, B. and S. R. Esswein (2015). "Inhibition by small-molecule ligands of formation of amyloid fibrils of an immunoglobulin light chain variable domain." **4**: e10935.

Brumshtein, B., S. R. Esswein, M. Landau, C. M. Ryan, J. P. Whitelegge, M. L. Phillips, D. Cascio, M. R. Sawaya and D. S. Eisenberg (2014). "Formation of amyloid fibers by monomeric light chain variable domains." J Biol Chem **289**(40): 27513-27525.

Brumshtein, B., S. R. Esswein, L. Salwinski, M. L. Phillips, A. T. Ly, D. Cascio, M. R. Sawaya and D. S. Eisenberg (2015). "Inhibition by small-molecule ligands of formation of amyloid fibrils of an immunoglobulin light chain variable domain." eLife **4**: e10935.

Bryngelson, J. D., J. N. Onuchic, N. D. Socci and P. G. Wolynes (1995). "Funnels, pathways, and the energy landscape of protein folding: A synthesis." Proteins: Structure, Function, and Bioinformatics **21**(3): 167-195.

Burre, J. (2015). "The Synaptic Function of alpha-Synuclein." J Parkinsons Dis **5**(4): 699-713.

Cabrita, L. D., S.-T. D. Hsu, H. Launay, C. M. Dobson and J. Christodoulou (2009). "Probing ribosome-nascent chain complexes produced in vivo by NMR spectroscopy." Proceedings of the National Academy of Sciences **106**(52): 22239-22244.

Campioni, S., E. Monsellier and F. Chiti (2010). Why Proteins Misfold. Protein Misfolding Diseases, John Wiley & Sons, Inc.**:** 1-20.

Capitani, G., J. M. Duarte, K. Baskaran, S. Bliven and J. C. Somody (2016). "Understanding the fabric of protein crystals: computational classification of biological interfaces and crystal contacts." Bioinformatics **32**(4): 481-489.

Capriotti, E., P. Fariselli, I. Rossi and R. Casadio (2008). "A three-state prediction of single point mutations on protein stability changes." BMC Bioinformatics **9**(Suppl 2): S6-S6.

Care, S. H. (2017). "Amyloidosis." Retrieved 02/08/2017, 2017, from https://stanfordhealthcare.org/medical-conditions/blood-heart-circulation/amyloidosis.html.

Carter, P. J. (2006). "Potent antibody therapeutics by design." Nat Rev Immunol **6**(5): 343-357.

Cassaignau, A. M. E., H. M. M. Launay, M.-E. Karyadi, X. Wang, C. A. Waudby, A. Deckert, A. L. Robertson, J. Christodoulou and L. D. Cabrita (2016). "A strategy for co-translational folding studies of ribosome-bound nascent chain complexes using NMR spectroscopy." Nat. Protocols **11**(8): 1492-1507.

Chakraborty, K., M. Chatila, J. Sinha, Q. Shi, B. C. Poschner, M. Sikor, G. Jiang, D. C. Lamb, F. U. Hartl and M. Hayer-Hartl (2010). "Chaperonin-

catalyzed rescue of kinetically trapped states in protein folding." Cell **142**(1): 112-122.

Chapman, Brynmor K., O. Davulcu, Jack J. Skalicky, Rafael P. Brüschweiler and Michael S. Chapman (2015). "Parsimony in Protein Conformational Change." Structure **23**(7): 1190-1198.

Chaulagain, C. P. and R. L. Comenzo (2015). "How we treat systemic light-chain amyloidosis." Clin Adv Hematol Oncol **13**(5): 315-324.

Chen, E. H. L., T. T. Y. Lu, J. C. C. Hsu, Y. J. Tseng, T. S. Lim and R. P. Y. Chen (2017). "Directly monitor protein rearrangement on a nanosecond-to-millisecond time-scale." Scientific Reports **7**(1): 8691.

Chikenji, G., Y. Fujitsuka and S. Takada (2006). "Shaping up the protein folding funnel by local interaction: lesson from a structure prediction study." Proc Natl Acad Sci U S A **103**(9): 3141-3146.

Chiu, J., P. E. March, R. Lee and D. Tillett (2004). "Site-directed, Ligase-Independent Mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h." Nucleic Acids Research **32**(21): e174-e174.

Cibeira, M. T., V. Sanchorawala, D. C. Seldin, K. Quillen, J. L. Berk, L. M. Dember, A. Segal, F. Ruberg, H. Meier-Ewert, N. T. Andrea, J. M. Sloan, K. T. Finn, G. Doros, J. Blade and M. Skinner (2011). "Outcome of AL amyloidosis after high-dose melphalan and autologous stem cell transplantation: long-term results in a series of 421 patients." Blood **118**(16): 4346-4352.

Clark, P. L. (2004). "Protein folding in the cell: reshaping the folding funnel." Trends in Biochemical Sciences **29**(10): 527-534.

Cohen, A. D. and R. L. Comenzo (2010). "Systemic light-chain amyloidosis: advances in diagnosis, prognosis, and therapy." Hematology Am Soc Hematol Educ Program **2010**: 287-294.

Cohen, A. S. (1986). General Introduction and a Brief History of Amyloidosis. Amyloidosis. J. Marrink and M. H. Van Rijswijk. Dordrecht, Springer Netherlands**:** 3-19.

Comenzo, R. L., J. Wally, G. Kica, J. Murray, T. Ericsson, M. Skinner and Y. Zhang (1999). "Clonal immunoglobulin light chain variable region germline gene use in AL amyloidosis: association with dominant amyloid-related organ involvement and survival after stem cell transplantation." Br J Haematol **106**(3): 744-751.

Comenzo, R. L., Y. Zhang, C. Martinez, K. Osman and G. A. Herrera (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**(3): 714-720.

Cooley, C. B., L. M. Ryno, L. Plate, G. J. Morgan, J. D. Hulleman, J. W. Kelly and R. L. Wiseman (2014). "Unfolded protein response activation reduces secretion and extracellular aggregation of amyloidogenic immunoglobulin light chain." Proceedings of the National Academy of Sciences of the United States of America **111**(36): 13046-13051.

Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman (1995). "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules." Journal of the American Chemical Society **117**(19): 5179-5197.

Cui, W., Y. Bai, P. Luo, L. Miao and L. Cai (2013). "Preventive and therapeutic effects of MG132 by activating Nrf2-ARE signaling pathway on oxidative stress-induced cardiovascular and renal injury." Oxid Med Cell Longev **2013**: 306073.

Cui, W., B. Li, Y. Bai, X. Miao, Q. Chen, W. Sun, Y. Tan, P. Luo, C. Zhang, S. Zheng, P. N. Epstein, L. Miao and L. Cai (2013). "Potential role for Nrf2

activation in the therapeutic effect of MG132 on diabetic nephropathy in OVE26 diabetic mice." Am J Physiol Endocrinol Metab **304**(1): E87-99.

Cvicek, V., W. A. Goddard, III and R. Abrol (2016). "Structure-Based Sequence Alignment of the Transmembrane Domains of All Human GPCRs: Phylogenetic, Structural and Functional Implications." PLOS Computational Biology **12**(3): e1004805.

Dannenberg, J. J. (1998). "An Introduction to Hydrogen Bonding By George A. Jeffrey (University of Pittsburgh). Oxford University Press: New York and Oxford. 1997. ix + 303 pp. $60.00. ISBN 0-19-509549-9." Journal of the American Chemical Society **120**(22): 5604-5604.

Dartigalongue, C., H. Nikaido and S. Raina (2000). "Protein folding in the periplasm in the absence of primary oxidant DsbA: modulation of redox potential in periplasmic space via OmpL porin." Embo j **19**(22): 5980-5988.

Davis, D. P., R. Raffen, J. L. Dul, S. M. Vogen, E. K. Williamson, F. J. Stevens and Y. Argon (2000). "Inhibition of Amyloid Fiber Assembly by Both BiP and Its Target Peptide." Immunity **13**(4): 433-442.

de Marco, A. (2009). "Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli." Microbial Cell Factories **8**(1): 26.

Dehouck, Y., J. M. Kwasigroch, M. Rooman and D. Gilis (2013). "BeAtMuSiC: Prediction of changes in protein-protein binding affinity on mutations." Nucleic Acids Res **41**(Web Server issue): W333-339.

Dehouck, Y., J. M. Kwasigroch, M. Rooman and D. Gilis (2013). "BeAtMuSiC: prediction of changes in protein–protein binding affinity on mutations." Nucleic Acids Research **41**(Web Server issue): W333-W339.

Del Pozo-Yauner, L., B. Becerril, A. Ochoa-Leyva, S. L. Rodríguez-Ambriz, J. I. P. Carrión, G. Zavala-Padilla, R. Sánchez-López and D. A. F. Velasco

(2015). The Structural Determinants of the Immunoglobulin Light Chain Amyloid Aggregation. Physical Biology of Proteins and Peptides: Theory, Experiment, and Simulation. L. Olivares-Quiroz, O. Guzmán-López and H. E. Jardón-Valadez. Cham, Springer International Publishing**:** 1-28.

Del Pozo Yauner, L., E. Ortiz, R. Sanchez, R. Sanchez-Lopez, L. Guereca, C. L. Murphy, A. Allen, J. S. Wall, D. A. Fernandez-Velasco, A. Solomon and B. Becerril (2008). "Influence of the germline sequence on the thermodynamic stability and fibrillogenicity of human lambda 6 light chains." Proteins **72**(2): 684-692.

Dember, L. M. (2006). "Amyloidosis-associated kidney disease." J Am Soc Nephrol **17**(12): 3458-3471.

DePristo, M. A., D. M. Weinreich and D. L. Hartl (2005). "Missense meanderings in sequence space: a biophysical view of protein evolution." Nat Rev Genet **6**(9): 678-687.

Di Scala, C., N. Yahi, S. Boutemeur, A. Flores, L. Rodriguez, H. Chahinian and J. Fantini (2016). "Common molecular mechanism of amyloid pore formation by Alzheimer's β-amyloid peptide and α-synuclein." Scientific Reports **6**: 28781.

DiCostanzo, A. C., J. R. Thompson, F. C. Peterson, B. F. Volkman and M. Ramirez-Alvarado (2012). "Tyrosine Residues Mediate Fibril Formation in a Dynamic Light Chain Dimer Interface." Journal of Biological Chemistry **287**(33): 27997-28006.

Dill, K. A. and H. S. Chan (1997). "From Levinthal to pathways to funnels." Nat Struct Biol **4**(1): 10-19.

Dill, K. A. and J. L. MacCallum (2012). "The protein-folding problem, 50 years on." Science **338**(6110): 1042-1046.

Dill, K. A., S. B. Ozkan, M. S. Shell and T. R. Weikl (2008). "The Protein Folding Problem." Annual review of biophysics **37**: 289-316.

DiMaio, F., A. Leaver-Fay, P. Bradley, D. Baker and I. Andre (2011). "Modeling symmetric macromolecular structures in Rosetta3." PLoS One **6**(6): e20450.

Dinner, S., W. Witteles, R. Witteles, A. Lam, S. Arai, R. Lafayette, T. I. George, S. L. Schrier and M. Liedtke (2013). "The prognostic value of diagnosing concurrent multiple myeloma in immunoglobulin light chain amyloidosis." Br J Haematol **161**(3): 367-372.

Diomede, L., P. Rognoni, F. Lavatelli, M. Romeo, E. del Favero, L. Cantù, E. Ghibaudi, A. di Fonzo, A. Corbelli, F. Fiordaliso, G. Palladini, V. Valentini, V. Perfetti, M. Salmona and G. Merlini (2014). "A Caenorhabditis elegans–based assay recognizes immunoglobulin light chains causing heart amyloidosis." Blood **123**(23): 3543-3552.

Diomede, L., P. Rognoni, F. Lavatelli, M. Romeo, A. di Fonzo, C. Foray, F. Fiordaliso, G. Palladini, V. Valentini, V. Perfetti, M. Salmona and G. Merlini (2014). "Investigating heart-specific toxicity of amyloidogenic immunoglobulin light chains: A lesson from C. elegans." Worm **3**(3): e965590.

Dispenzieri, A., F. Buadi, S. K. Kumar, C. B. Reeder, T. Sher, M. Q. Lacy, R. A. Kyle, J. R. Mikhael, V. Roy, N. Leung, M. Grogan, P. Kapoor, J. A. Lust, D. Dingli, R. S. Go, Y. L. Hwa, S. R. Hayman, R. Fonseca, S. Ailawadhi, P. L. Bergsagel, A. Chanan-Khan, S. V. Rajkumar, S. J. Russell, K. Stewart, S. R. Zeldenrust and M. A. Gertz (2015). "Treatment of Immunoglobulin Light Chain Amyloidosis: Mayo Stratification of Myeloma and Risk-Adapted Therapy (mSMART) Consensus Statement." Mayo Clin Proc **90**(8): 1054-1081.

Dispenzieri, A. and G. Merlini (2016). Immunoglobulin Light Chain Systemic Amyloidosis. Plasma Cell Dyscrasias. A. M. Roccaro and I. M. Ghobrial. Cham, Springer International Publishing**:** 273-318.

Dispenzieri, A., K. Seenithamby, M. Q. Lacy, S. K. Kumar, F. K. Buadi, S. R. Hayman, D. Dingli, M. R. Litzow, D. A. Gastineau, D. J. Inwards, I. N. Micallef, S. M. Ansell, P. B. Johnston, L. F. Porrata, M. M. Patnaik, W. J. Hogan and M. A. A. Gertz (2013). "Patients with immunoglobulin light chain amyloidosis undergoing autologous stem cell transplantation have superior outcomes compared with patients with multiple myeloma: a retrospective review from a tertiary referral center." Bone Marrow Transplant **48**(10): 1302-1307.

Dobson, C. M., A. Šali and M. Karplus (1998). "Protein Folding: A Perspective from Theory and Experiment." Angewandte Chemie International Edition: 868–893.

Douglas, P. M., D. W. Summers and D. M. Cyr (2009). "Molecular chaperones antagonize proteotoxicity by differentially modulating protein aggregation pathways." Prion **3**(2): 51-58.

Dourado, D. F. A. R. and S. C. Flores (2016). "Modeling and fitting protein-protein complexes to predict change of binding energy." **6**: 25406.

Driscoll, J. J. and S. Girnius (2016). Proteasome Inhibitors to Treat AL Amyloidosis. Exploring New Findings on Amyloidosis. A.-M. Fernandez-Escamilla. Rijeka, InTech**:** Ch. 11.

Dunbrack, R. L., Jr. and F. E. Cohen (1997). "Bayesian statistical analysis of protein side-chain rotamer preferences." Protein Sci **6**(8): 1661-1681.

Eanes, E. D. and G. G. Glenner (1968). "X-ray diffraction studies on amyloid filaments." J Histochem Cytochem **16**(11): 673-677.

Eberth, C. J. (1881). "Zur Amyloidfrage." Archiv für pathologische Anatomie und Physiologie und für klinische Medicin **84**(1): 111-118.

Eder , L. and H. Bitterman (2007). "Amyloid Purpura." New England Journal of Medicine **356**(23): 2406-2406.

Eichner, T. and Sheena E. Radford (2011). "A Diversity of Assembly Mechanisms of a Generic Amyloid Fold." Molecular Cell **43**(1): 8-18.

Eisenberg, D. and M. Jucker "The Amyloid State of Proteins in Human Diseases." Cell **148**(6): 1188-1203.

Eisenberg, D. and M. Jucker (2012). "The amyloid state of proteins in human diseases." Cell **148**(6): 1188-1203.

Emamzadeh, F. N. (2016). "Alpha-synuclein structure, functions, and interactions." J Res Med Sci **21**: 29.

Esteras-Chopo, A., L. Serrano and M. L. de la Paz (2005). "The amyloid stretch hypothesis: Recruiting proteins toward the dark side." Proceedings of the National Academy of Sciences of the United States of America **102**(46): 16672-16677.

Falk, R. H. (2005). "Diagnosis and management of the cardiac amyloidoses." Circulation **112**(13): 2047-2060.

Falk, R. H. (2014). "AL amyloidosis or multiple myeloma? An important distinction." British Journal of Haematology **164**(5): 748-749.

Falk, R. H. and S. W. Dubrey (2010). "Amyloid heart disease." Prog Cardiovasc Dis **52**(4): 347-361.

Fandrich, M., M. Schmidt and N. Grigorieff (2011). "Recent progress in understanding Alzheimer's beta-amyloid structures." Trends Biochem Sci **36**(6): 338-345.

Fedorov, A. N. and T. O. Baldwin (1997). "Cotranslational Protein Folding." Journal of Biological Chemistry **272**(52): 32715-32718.

Fehlhammer, H., M. Schiffer, O. Epp, P. M. Colman, E. E. Lattman, P. Schwager, W. Steigemann and H. J. Schramm (1975). "The structure

determination of the variable portion of the Bence-Jones protein Au." Biophysics of structure and mechanism **1**(2): 139-146.

Fikrle, M., T. Paleček, P. Kuchynka, E. Němeček, L. Bauerová, J. Straub and R. Ryšavá (2013). "Cardiac amyloidosis: A comprehensive review." Cor et Vasa **55**(1): e60-e75.

Fitzpatrick, A. W. P., G. T. Debelouchina, M. J. Bayro, D. K. Clare, M. A. Caporini, V. S. Bajaj, C. P. Jaroniec, L. Wang, V. Ladizhansky, S. A. Müller, C. E. MacPhee, C. A. Waudby, H. R. Mott, A. De Simone, T. P. J. Knowles, H. R. Saibil, M. Vendruscolo, E. V. Orlova, R. G. Griffin and C. M. Dobson (2013). "Atomic structure and hierarchical assembly of a cross-β amyloid fibril." Proceedings of the National Academy of Sciences **110**(14): 5468-5473.

Fitzpatrick, A. W. P., S. T. Park and A. H. Zewail (2013). "Exceptional rigidity and biomechanics of amyloid revealed by 4D electron microscopy." Proceedings of the National Academy of Sciences of the United States of America **110**(27): 10976-10981.

Fred J. Stevens, F. A. W., Alan Soloman, Marianne Schiffer (1980). "Self-association of human immunoglobulin KI light chains: Role of the third hypervariable region." PNAS **77**(2): 1144-1148.

Friedrich, R. P., K. Tepper, R. Rönicke, M. Soom, M. Westermann, K. Reymann, C. Kaether and M. Fändrich (2010). "Mechanism of amyloid plaque formation suggests an intracellular basis of Aβ pathogenicity." Proceedings of the National Academy of Sciences **107**(5): 1942-1947.

Galvagnion, C., A. K. Buell, G. Meisl, T. C. Michaels, M. Vendruscolo, T. P. Knowles and C. M. Dobson (2015). "Lipid vesicles trigger alpha-synuclein aggregation by stimulating primary nucleation." Nat Chem Biol **11**(3): 229-234.

Gandini, R., T. Reichenbach, T. C. Tan and C. Divne (2017). "Structural basis for dolichylphosphate mannose biosynthesis." Nat Commun **8**(1): 120.

Garland, S. L. (2013). "Are GPCRs still a source of new targets?" J Biomol Screen **18**(9): 947-966.

Gasior, P. and M. Kotulska (2014). "FISH Amyloid - a new method for finding amyloidogenic segments in proteins based on site specific co-occurrence of aminoacids." BMC Bioinformatics **15**: 54.

Geddes, A. J., K. D. Parker, E. D. Atkins and E. Beighton (1968). ""Cross-beta" conformation in proteins." J Mol Biol **32**(2): 343-358.

Gertz, M. A., R. Comenzo, R. H. Falk, J. P. Fermand, B. P. Hazenberg, P. N. Hawkins, G. Merlini, P. Moreau, P. Ronco, V. Sanchorawala, O. Sezer, A. Solomon and G. Grateau (2005). "Definition of organ involvement and treatment response in immunoglobulin light chain amyloidosis (AL): a consensus opinion from the 10th International Symposium on Amyloid and Amyloidosis, Tours, France, 18-22 April 2004." Am J Hematol **79**(4): 319-328.

Gertz, M. A., M. Q. Lacy and A. Dispenzieri (2002). "Immunoglobulin light chain amyloidosis and the kidney." Kidney International **61**(1): 1-9.

Gertz, M. A., H. Landau, R. L. Comenzo, D. Seldin, B. Weiss, J. Zonder, G. Merlini, S. Schonland, J. Walling, G. G. Kinney, M. Koller, D. B. Schenk, S. D. Guthrie and M. Liedtke (2016). "First-in-Human Phase I/II Study of NEOD001 in Patients With Light Chain Amyloidosis and Persistent Organ Dysfunction." J Clin Oncol **34**(10): 1097-1103.

Ghosh, P., A. Vaidya, A. Kumar and V. Rangachari (2016). "Determination of critical nucleation number for a single nucleation amyloid-β aggregation model." Mathematical biosciences **273**: 70-79.

Gilbert, H. F. (1997). "Protein Disulfide Isomerase and Assisted Protein Folding." Journal of Biological Chemistry **272**(47): 29399-29402.

Gillmore, J. D. and P. N. Hawkins (2013). "Pathophysiology and treatment of systemic amyloidosis." Nat Rev Nephrol **9**(10): 574-586.

Glabe, C. (2001). "Intracellular mechanisms of amyloid accumulation and pathogenesis in Alzheimer's disease." J Mol Neurosci **17**(2): 137-145.

Glenner, G. G., P. Cuatrecasas, C. Isersky, H. A. Bladen and E. D. Eanes (1969). "Physical and chemical properties of amyloid fibers. II. Isolation of a unique protein constituting the major component from human splenic amyloid fibril concentrates." J Histochem Cytochem **17**(12): 769-780.

Goemans, C., K. Denoncin and J.-F. Collet (2014). "Folding mechanisms of periplasmic proteins." Biochimica et Biophysica Acta (BBA) - Molecular Cell Research **1843**(8): 1517-1528.

Goldschmidt, L., P. K. Teng, R. Riek and D. Eisenberg (2010). "Identifying the amylome, proteins capable of forming amyloid-like fibrils." Proceedings of the National Academy of Sciences **107**(8): 3487-3492.

Goni, F. and B. Frangione (1983). "Amino acid sequence of the Fv region of a human monoclonal IgM (protein WEA) with antibody activity against 3,4-pyruvylated galactose in Klebsiella polysaccharides K30 and K33." Proc Natl Acad Sci U S A **80**(15): 4837-4841.

Gopavajhula, V. R., K. V. Chaitanya, P. Akbar Ali Khan, J. P. Shaik, P. N. Reddy and M. Alanazi (2013). "Modeling and analysis of soybean (Glycine max. L) Cu/Zn, Mn and Fe superoxide dismutases." Genet Mol Biol **36**(2): 225-236.

Graña-Montes, R., J. Pujols-Pujol, C. Gómez-Picanyol and S. Ventura (2017). Prediction of Protein Aggregation and Amyloid Formation. From Protein Structure to Function with Bioinformatics. D. J. Rigden. Dordrecht, Springer Netherlands**:** 205-263.

Gras, S., L. Waddington and K. N Goldie (2011). Transmission Electron Microscopy of Amyloid Fibrils.

Gray, V. E., R. J. Hause and D. M. Fowler (2017). "Analysis of Large-Scale Mutagenesis Data To Assess the Impact of Single Amino Acid Substitutions." Genetics **207**(1): 53-61.

Grogan, M., A. Dispenzieri and M. A. Gertz (2017). "Light-chain cardiac amyloidosis: strategies to promote early diagnosis and cardiac response." Heart.

Guan, J., S. Mishra, Y. Qiu, J. Shi, K. Trudeau, G. Las, M. Liesa, O. S. Shirihai, L. H. Connors, D. C. Seldin, R. H. Falk, C. A. MacRae and R. Liao (2014). "Lysosomal dysfunction and impaired autophagy underlie the pathogenesis of amyloidogenic light chain-mediated cardiotoxicity." EMBO Mol Med **6**(11): 1493-1507.

Guerois, R., J. E. Nielsen and L. Serrano (2002). "Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations." J Mol Biol **320**(2): 369-387.

Hajra, A. and D. Bandyopadhyay (2016). "An interesting case of renal amyloidosis." Indian Journal of Nephrology **26**(6): 467-469.

Hall, T. A. (1999). "BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT." Nucleic Acids Symposium Series **41**: 95-98.

Hamelryck, T., J. T. Kent and A. Krogh (2006). "Sampling Realistic Protein Conformations Using Local Structural Bias." PLOS Computational Biology **2**(9): e131.

Hane, F. and Z. Leonenko (2014). "Effect of metals on kinetic pathways of amyloid-beta aggregation." Biomolecules **4**(1): 101-116.

Hartl, F. U., A. Bracher and M. Hayer-Hartl (2011). "Molecular chaperones in protein folding and proteostasis." Nature **475**(7356): 324-332.

Hashimoto, K. and A. R. Panchenko (2010). "Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states." Proc Natl Acad Sci U S A **107**(47): 20352-20357.

Hassan, W., H. Al-Sergani, W. Mourad and R. Tabbaa (2005). "Amyloid Heart Disease: New Frontiers and Insights in Pathophysiology, Diagnosis, and Management." Texas Heart Institute Journal **32**(2): 178-184.

He, X., M. G. Chen and Q. Ma (2008). "Activation of Nrf2 in defense against cadmium-induced oxidative stress." Chem Res Toxicol **21**(7): 1375-1383.

Hemmingsen, J. M., K. M. Gernert, J. S. Richardson and D. C. Richardson (1994). "The tyrosine corner: a feature of most Greek key beta-barrel proteins." Protein Science : A Publication of the Protein Society **3**(11): 1927-1937.

Hermanson, G. T. (2013). Chapter 1 - Introduction to Bioconjugation. Bioconjugate Techniques (Third edition). Boston, Academic Press**:** 1-125.

Hernández-Santoyo, A., L. del Pozo Yauner, D. Fuentes-Silva, E. Ortiz, E. Rudiño-Piñera, R. Sánchez-López, E. Horjales, B. Becerril and A. Rodríguez-Romero (2010). "A Single Mutation at the Sheet Switch Region Results in Conformational Changes Favoring λ6 Light-Chain Fibrillogenesis." Journal of Molecular Biology **396**(2): 280-292.

Herranz-Trillo, F., M. Groenning, A. van Maarschalkerweerd, R. Tauler, B. Vestergaard and P. Bernado (2017). "Structural Analysis of Multi-component Amyloid Systems by Chemometric SAXS Data Decomposition." Structure **25**(1): 5-15.

Hideshima, T., P. G. Richardson and K. C. Anderson (2011). "Mechanism of Action of Proteasome Inhibitors and Deacetylase Inhibitors and the Biological Basis of Synergy in Multiple Myeloma." Molecular cancer therapeutics **10**(11): 2034-2042.

Hinault, M.-P., A. Ben-Zvi and P. Goloubinoff (2006). "Chaperones and proteases." Journal of Molecular Neuroscience **30**(3): 249-265.

Holtkamp, W., G. Kokic, M. Jäger, J. Mittelstaet, A. A. Komar and M. V. Rodnina (2015). "Cotranslational protein folding on the ribosome monitored in real time." Science **350**(6264): 1104-1107.

Hooper, D. H. a. N. (2000). BIOS Instant Notes in Biochemistry, Taylor & Francis.

Hornak, V., R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling (2006). "Comparison of multiple Amber force fields and development of improved protein backbone parameters." Proteins **65**(3): 712-725.

Horwich, A. (2002). "Protein aggregation in disease: a role for folding intermediates forming specific multimeric interactions." The Journal of Clinical Investigation **110**(9): 1221-1232.

Hu, Y.-B., E. B. Dammer, R.-J. Ren and G. Wang (2015). "The endosomal-lysosomal system: from acidification and cargo sorting to neurodegeneration." Translational Neurodegeneration **4**: 18.

Huang, D.-B., C.-H. Chang, C. Ainsworth, G. Johnson, A. Solomon, F. J. Stevens and M. Schiffer (1997). "Variable domain structure of κIV human light chain Len: High homology to the murine light chain McPC603." Molecular Immunology **34**(18): 1291-1301.

Huang, F., Y. Mackeyev, E. Watson, A. Cheney Matthew, J. Wilson Lon and J. Suh (2013). Evidence for nuclear internalisation of biocompatible [60]fullerene1). European Journal of Nanomedicine. **5:** 51.

Huang, X., Q. Wang, W. Chen, G. Ren and Z. Liu (2016). "Bortezomib with dexamethasone as first-line treatment for AL amyloidosis with renal involvement." Amyloid **23**(1): 51-57.

Hubin, E., N. A. van Nuland, K. Broersen and K. Pauwels (2014). "Transient dynamics of Abeta contribute to toxicity in Alzheimer's disease." Cell Mol Life Sci **71**(18): 3507-3521.

Hurle, M. R., L. R. Helms, L. Li, W. Chan and R. Wetzel (1994). "A role for destabilizing amino acid replacements in light-chain amyloidosis." Proceedings of the National Academy of Sciences of the United States of America **91**(12): 5446-5450.

Hutchinson, E. G. and J. M. Thornton (1994). "A revised set of potentials for beta-turn formation in proteins." Protein Science : A Publication of the Protein Society **3**(12): 2207-2216.

Imamura, T., M. Ogata, K. Kohno, T. Tomo, E. Ohtsuka, H. Kikuchi and J. Kadota (2006). "Successful reduced intensity allogeneic stem cell transplantation for systemic AL amyloidosis." Am J Hematol **81**(4): 281-283.

Imperlini, E., M. Gnecchi, P. Rognoni, E. Sabido, M. C. Ciuffreda, G. Palladini, G. Espadas, F. M. Mancuso, M. Bozzola, G. Malpasso, V. Valentini, G. Palladini, S. Orru, G. Ferraro, P. Milani, S. Perlini, F. Salvatore, G. Merlini and F. Lavatelli (2017). "Proteotoxicity in cardiac amyloidosis: amyloidogenic light chains affect the levels of intracellular proteins in human heart cells." Sci Rep **7**(1): 15661.

Jahn, T. R. and S. E. Radford (2008). "Folding versus aggregation: Polypeptide conformations on competing pathways." Archives of Biochemistry and Biophysics **469**(1): 100-117.

Jamroz, M., A. Kolinski and S. Kmiecik (2013). "CABS-flex: Server for fast simulation of protein structure fluctuations." Nucleic Acids Res **41**(Web Server issue): W427-431.

Jelinek, T., E. Kryukova, Z. Kufova, F. Kryukov and R. Hajek (2016). "Proteasome inhibitors in AL amyloidosis: focus on mechanism of action and clinical activity." Hematol Oncol.

Jenner, E. (2014). "Serum free light chains in clinical laboratory diagnostics." Clinica Chimica Acta **427**: 15-20.

Jerson, L. S. and F. Debora (2009). "Hydration, cavities and volume in protein folding, aggregation and amyloid assembly." Physical Biology **6**(1): 015002.

Jimenez, J. L., E. J. Nettleton, M. Bouchard, C. V. Robinson, C. M. Dobson and H. R. Saibil (2002). "The protofilament structure of insulin amyloid fibrils." Proc Natl Acad Sci U S A **99**(14): 9196-9201.

Jiménez, J. L., E. J. Nettleton, M. Bouchard, C. V. Robinson, C. M. Dobson and H. R. Saibil (2002). "The protofilament structure of insulin amyloid fibrils." Proceedings of the National Academy of Sciences **99**(14): 9196-9201.

Jobbagy, A. and K. Kiraly (1966). "Chemical characterization of fluorescein isothiocyanate-protein conjugates." Biochim Biophys Acta **124**(1): 166-175.

Johnson, S. M., S. Connelly, C. Fearns, E. T. Powers and J. W. Kelly (2012). "The Transthyretin Amyloidoses: From Delineating the Molecular Mechanism of Aggregation Linked to Pathology to a Regulatory Agency Approved Drug." Journal of Molecular Biology **421**(2-3): 185-203.

Jolly, C. J., S. D. Wagner, C. Rada, N. Klix, C. Milstein and M. S. Neuberger (1996). "The targeting of somatic hypermutation." Semin Immunol **8**(3): 159-168.

Joob, B. and V. Wiwanitkit (2014). "Identification of active pocket and protein druggability within envelope glycoprotein GP2 from Ebola virus." Asian Pacific Journal of Tropical Biomedicine **4**(12): 939-940.

Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein (1983). "Comparison of simple potential functions for simulating liquid water." The Journal of Chemical Physics **79**(2): 926-935.

Jucker, M. and L. C. Walker (2013). "Self-propagation of pathogenic protein aggregates in neurodegenerative diseases." Nature **501**(7465): 45-51.

Julian, M. C., L. Li, S. Garde, R. Wilen and P. M. Tessier (2017). "Efficient affinity maturation of antibody variable domains requires co-selection of compensatory mutations to maintain thermodynamic stability." Sci Rep **7**: 45259.

Kabat, E. A., T. Te Wu, C. Foeller, H. M. Perry and K. S. Gottesman (1992). Sequences of proteins of immunological interest, DIANE publishing.

Kaplan, B., A. Livneh and B. A. Sela (2011). "Immunoglobulin free light chain dimers in human diseases." ScientificWorldJournal **11**: 726-735.

Karamanos, T. K., A. P. Kalverda, G. S. Thompson and S. E. Radford (2015). "Mechanisms of amyloid formation revealed by solution NMR." Prog Nucl Magn Reson Spectrosc **88-89**: 86-104.

Karamanos, T. K., C. L. Pashley, A. P. Kalverda, G. S. Thompson, M. Mayzel, V. Y. Orekhov and S. E. Radford (2016). "A Population Shift between Sparsely Populated Folding Intermediates Determines Amyloidogenicity." J Am Chem Soc **138**(19): 6271-6280.

Kastritis, E. and M. A. Dimopoulos (2016). "Recent advances in the management of AL Amyloidosis." British Journal of Haematology **172**(2): 170-186.

Khurana, R., C. Coleman, C. Ionescu-Zanetti, S. A. Carter, V. Krishna, R. K. Grover, R. Roy and S. Singh (2005). "Mechanism of thioflavin T binding to amyloid fibrils." J Struct Biol **151**(3): 229-238.

Khurana, R., P. O. Souillac, A. C. Coats, L. Minert, C. Ionescu-Zanetti, S. A. Carter, A. Solomon and A. L. Fink (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**(2): 97-109.

Kim, Y.-S., S. P. Cape, E. Chi, R. Raffen, P. Wilkins-Stevens, F. J. Stevens, M. C. Manning, T. W. Randolph, A. Solomon and J. F. Carpenter (2001). "Counteracting Effects of Renal Solutes on Amyloid Fibril Formation by Immunoglobulin Light Chains." Journal of Biological Chemistry **276**(2): 1626-1633.

Kimes, B. W. and B. L. Brandt (1976). "Properties of a clonal muscle cell line from rat heart." Exp Cell Res **98**(2): 367-381.

Kisilevsky, R., S. Raimondi and V. Bellotti (2016). "Historical and Current Concepts of Fibrillogenesis and In vivo Amyloidogenesis: Implications of Amyloid Tissue Targeting." Frontiers in Molecular Biosciences **3**: 17.

Knowles, T. P., A. W. Fitzpatrick, S. Meehan, H. R. Mott, M. Vendruscolo, C. M. Dobson and M. E. Welland (2007). "Role of Intermolecular Forces in Defining Material Properties of Protein Nanofibrils." Science **318**(5858): 1900-1903.

Knowles, T. P., M. Vendruscolo and C. M. Dobson (2014). "The amyloid state and its association with protein misfolding diseases." Nat Rev Mol Cell Biol **15**(6): 384-396.

Knowles, T. P. J., M. Vendruscolo and C. M. Dobson (2014). "The amyloid state and its association with protein misfolding diseases." Nat Rev Mol Cell Biol **15**(6): 384-396.

Koenig, P., C. V. Lee, B. T. Walters, V. Janakiraman, J. Stinson, T. W. Patapoff and G. Fuh (2017). "Mutational landscape of antibody variable domains reveals a switch modulating the interdomain conformational dynamics and antigen binding." Proceedings of the National Academy of Sciences **114**(4): E486-E495.

Kolmar, H., C. Frisch, G. Kleemann, K. Gotze, F. J. Stevens and H. J. Fritz (1994). "Dimerization of Bence Jones proteins: linking the rate of transcription

from an Escherichia coli promoter to the association constant of REIV." Biol Chem Hoppe Seyler **375**(1): 61-70.

Konagurthu, A. S., J. C. Whisstock, P. J. Stuckey and A. M. Lesk (2006). "MUSTANG: a multiple structural alignment algorithm." Proteins **64**(3): 559-574.

Koniev, O. and A. Wagner (2015). "Developments and recent advancements in the field of endogenous amino acid selective bond forming reactions for bioconjugation." Chem Soc Rev **44**(15): 5495-5551.

Korbet, S. M. and M. M. Schwartz (2006). "Multiple Myeloma." Journal of the American Society of Nephrology **17**(9): 2533-2545.

Krissinel, E. (2011). "Macromolecular complexes in crystals and solutions." Acta Crystallographica Section D: Biological Crystallography **67**(Pt 4): 376-385.

Krissinel, E. and K. Henrick (2004). "Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2256-2268.

Krissinel, E. and K. Henrick (2005). Detection of Protein Assemblies in Crystals. Computational Life Sciences: First International Symposium, CompLife 2005, Konstanz, Germany, September 25-27, 2005. Proceedings. M. R. Berthold, R. C. Glen, K. Diederichs, O. Kohlbacher and I. Fischer. Berlin, Heidelberg, Springer Berlin Heidelberg**:** 163-174.

Krissinel, E. and K. Henrick (2007). "Inference of macromolecular assemblies from crystalline state." J Mol Biol **372**(3): 774-797.

Kristinsson, S. Y., A. R. Minter, N. Korde, E. Tan and O. Landgren (2011). "Bone disease in multiple myeloma and precursor disease: novel diagnostic approaches and implications on clinical management." Expert review of molecular diagnostics **11**(6): 593-603.

Kubiczkova, L., L. Pour, L. Sedlarikova, R. Hajek and S. Sevcikova (2014). "Proteasome inhibitors – molecular basis and current perspectives in multiple myeloma." Journal of Cellular and Molecular Medicine **18**(6): 947-961.

Kumar, J., R. Namsechi and V. L. Sim (2015). "Structure-Based Peptide Design to Modulate Amyloid Beta Aggregation and Reduce Cytotoxicity." PLOS ONE **10**(6): e0129087.

Kumar, S., A. Dispenzieri, J. A. Katzmann, D. R. Larson, C. L. Colby, M. Q. Lacy, S. R. Hayman, F. K. Buadi, N. Leung, S. R. Zeldenrust, M. Ramirez-Alvarado, R. J. Clark, R. A. Kyle, S. V. Rajkumar and M. A. Gertz (2010). "Serum immunoglobulin free light-chain measurement in primary amyloidosis: prognostic value and correlations with clinical features." Blood **116**(24): 5126-5129.

Kumar, S. and J. Walter (2011). Phosphorylation of amyloid beta (Aβ) peptides – A trigger for formation of toxic aggregates in Alzheimer's disease.

Kumar, V., S. Rahman, H. Choudhry, M. A. Zamzami, M. Sarwar Jamal, A. Islam, F. Ahmad and M. I. Hassan (2017). "Computing disease-linked SOD1 mutations: deciphering protein stability and patient-phenotype relations." Sci Rep **7**(1): 4678.

Kuznetsov, A. V., S. Javadov, S. Sickinger, S. Frotschnig and M. Grimm (2015). "H9c2 and HL-1 cells demonstrate distinct features of energy metabolism, mitochondrial function and sensitivity to hypoxia-reoxygenation." Biochimica et Biophysica Acta (BBA) - Molecular Cell Research **1853**(2): 276-284.

Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." Nature **227**(5259): 680-685.

LaFerla, F. M., K. N. Green and S. Oddo (2007). "Intracellular amyloid-[beta] in Alzheimer's disease." Nat Rev Neurosci **8**(7): 499-509.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson and D. G. Higgins (2007). "Clustal W and Clustal X version 2.0." Bioinformatics **23**(21): 2947-2948.

Lau, A., W. Tian, S. A. Whitman and D. D. Zhang (2013). "The Predicted Molecular Weight of Nrf2: It Is What It Is Not." Antioxidants & Redox Signaling **18**(1): 91-93.

Lavatelli, F., E. Imperlini, S. Orru, P. Rognoni, D. Sarnataro, G. Palladini, G. Malpasso, M. E. Soriano, A. Di Fonzo, V. Valentini, M. Gnecchi, S. Perlini, F. Salvatore and G. Merlini (2015). "Novel mitochondrial protein interactors of immunoglobulin light chains causing heart amyloidosis." Faseb j **29**(11): 4614-4628.

Le Bras, F., V. Molinier-Frenkel, A. Guellich, J. Dupuis, K. Belhadj, S. Guendouz, K. Ayad, M. Colombat, N. Benhaiem, C. M. Tissot, A. Hulin, A. Jaccard and T. Damy (2017). "Sequential cyclophosphamide-bortezomib-dexamethasone unmasks the harmful cardiac effect of dexamethasone in primary light-chain cardiac amyloidosis." Eur J Cancer **76**: 183-187.

Lee, C. C., A. Nayak, A. Sethuraman, G. Belfort and G. J. McRae (2007). "A three-stage kinetic model of amyloid fibrillation." Biophys J **92**(10): 3448-3458.

Lee, S. J., H. S. Lim, E. Masliah and H. J. Lee (2011). "Protein aggregate spreading in neurodegenerative diseases: problems and perspectives." Neurosci Res **70**(4): 339-348.

Lenco, J., O. Lencova-Popelova, M. Link, A. Jirkovska, V. Tambor, E. Potuckova, J. Stulik, T. Simunek and M. Sterba (2015). "Proteomic investigation of embryonic rat heart-derived H9c2 cell line sheds new light on the molecular phenotype of the popular cell model." Exp Cell Res **339**(2): 174-186.

Lesk, A. M. (2004). Introduction to protein science-architecture, funciton, and genomics., Oxford University Press.

Levinson, R. T., O. O. Olatoye, E. G. Randles, K. G. Howell, A. C. DiCostanzo and M. Ramirez-Alvarado (2013). "Role of mutations in the cellular internalization of amyloidogenic light chains into cardiomyocytes." Sci Rep **3**: 1278.

Levinthal, C. (1968). How to fold graciously. Debrunner P
Tsibris J.C.M
Münck E. . Mossbauer Spectroscopy in Biological Systems, Proceedings of a Meeting held at Allerton House, Monticello, Illinois, University of Illinois Press**:** 22.

Li, M., F. L. Simonetti, A. Goncearenco and A. R. Panchenko (2016). "MutaBind estimates and interprets the effects of sequence variants on protein-protein interactions." Nucleic Acids Res **44**(W1): W494-501.

Li, M., F. L. Simonetti, A. Goncearenco and A. R. Panchenko (2016). "MutaBind estimates and interprets the effects of sequence variants on protein–protein interactions." Nucleic Acids Research **44**(Web Server issue): W494-W501.

Li, Y., J. D. Paonessa and Y. Zhang (2012). "Mechanism of Chemical Activation of Nrf2." PLoS ONE **7**(4): e35122.

Lin, H. M., D. Seldin, A. M. Hui, D. Berg, C. N. Dietrich and E. Flood (2015). "The patient's perspective on the symptom and everyday life impact of AL amyloidosis." Amyloid **22**(4): 244-251.

Lorenzo, A. and B. A. Yankner (1996). "Amyloid fibril toxicity in Alzheimer's disease and diabetes." Ann N Y Acad Sci **777**: 89-95.

Lovell, S. C., I. W. Davis, W. B. Arendall, 3rd, P. I. de Bakker, J. M. Word, M. G. Prisant, J. S. Richardson and D. C. Richardson (2003). "Structure validation by Calpha geometry: phi,psi and Cbeta deviation." Proteins **50**(3): 437-450.

Lü, S. and J. Wang (2013). "The resistance mechanisms of proteasome inhibitor bortezomib." Biomarker Research **1**: 13-13.

Lux, K., N. Goerlitz, S. Schlemminger, L. Perabo, D. Goldnau, J. Endell, K. Leike, D. M. Kofler, S. Finke, M. Hallek and H. Büning (2005). "Green Fluorescent Protein-Tagged Adeno-Associated Virus Particles Allow the Study of Cytosolic and Nuclear Trafficking." Journal of Virology **79**(18): 11776-11787.

Mahmood, S., G. Palladini, V. Sanchorawala and A. Wechalekar (2014). "Update on treatment of light chain amyloidosis." Haematologica **99**(2): 209-221.

Maier, J. A., C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling (2015). "ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB." J Chem Theory Comput **11**(8): 3696-3713.

Malloy, M. T., D. J. McIntosh, T. S. Walters, A. Flores, J. S. Goodwin and I. J. Arinze (2013). "Trafficking of the transcription factor Nrf2 to promyelocytic leukemia-nuclear bodies: implications for degradation of NRF2 in the nucleus." J Biol Chem **288**(20): 14569-14583.

Marin-Argany, M., J. Guell-Bosch, L. M. Blancas-Mejia, S. Villegas and M. Ramirez-Alvarado (2015). "Mutations can cause light chains to be too stable or too unstable to form amyloid fibrils." Protein Sci **24**(11): 1829-1840.

Marin-Argany, M., Y. Lin, P. Misra, A. Williams, J. S. Wall, K. G. Howell, L. R. Elsbernd, M. McClure and M. Ramirez-Alvarado (2016). "Cell Damage in Light Chain Amyloidosis: FIBRIL INTERNALIZATION, TOXICITY AND CELL-MEDIATED SEEDING." J Biol Chem **291**(38): 19813-19825.

Martin, A. C., O. C. Fau, E. G. Hutchinson, S. Hutchinson Eg Fau - Jones, M. Jones S Fau - Karmirantzou, R. A. Karmirantzou M Fau - Laskowski, J. B. Laskowski Ra Fau - Mitchell, C. Mitchell Jb Fau - Taroni, J. M. Taroni C Fau - Thornton and J. M. Thornton (1998). "Protein folds and functions." Structure(0969-2126 (Print)): 875-884.

Martin, D. J. and M. Ramirez-Alvarado (2010). "Comparison of amyloid fibril formation by two closely related immunoglobulin light chain variable domains." Amyloid : the international journal of experimental and clinical investigation : the official journal of the International Society of Amyloidosis **17**(3-4): 129-136.

Martin, D. J. and M. Ramirez-Alvarado (2010). "Comparison of amyloid fibril formation by two closely related immunoglobulin light chain variable domains." Amyloid **17**(3-4): 129-136.

Martin, E. B., A. Williams, C. Wooliver, R. E. Heidel, S. Adams, J. Dunlap, M. Ramirez-Alvarado, L. M. Blancas-Mejia, R. H. Lands, S. J. Kennel and J. S. Wall (2017). "Differential recruitment efficacy of patient-derived amyloidogenic and myeloma light chain proteins by synthetic fibrils—A metric for predicting amyloid propensity." PLOS ONE **12**(3): e0174152.

McDermott, C. J. and P. J. Shaw (2008). "Diagnosis and management of motor neurone disease." BMJ : British Medical Journal **336**(7645): 658-662.

McWilliams-Koeppen, H. P., J. S. Foster, N. Hackenbrack, M. Ramirez-Alvarado, D. Donohoe, A. Williams, S. Macy, C. Wooliver, D. Wortham, J. Morrell-Falvey, C. M. Foster, S. J. Kennel and J. S. Wall (2015). "Light Chain Amyloid Fibrils Cause Metabolic Dysfunction in Human Cardiomyocytes." PLoS One **10**(9): e0137716.

Melmed, G. M. (2009). "Light chain amyloidosis: a case presentation and review." Proceedings (Baylor University. Medical Center) **22**(3): 280-283.

Meng, X., A. L. Fink and V. N. Uversky (2008). "The effect of membranes on the in vitro fibrillation of an amyloidogenic light chain variable domain SMA." Journal of molecular biology **381**(4): 989-999.

Merlini, G. (2016). "Epigallocatechingallate (EGCG) in Cardiac AL Amyloidosis (EpiCardiAL)." Retrieved 01/09/2017, 2017, from https://clinicaltrials.gov/ct2/show/NCT01511263.

Merlini, G. and V. Bellotti (2003). "Molecular mechanisms of amyloidosis." N Engl J Med **349**(6): 583-596.

Mesleh, M. F., J. M. Hunter, A. A. Shvartsburg, G. C. Schatz and M. F. Jarrold (1997). "Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential." The Journal of Physical Chemistry A **101**(5): 968-968.

Micsonai, A., F. Wien, L. Kernya, Y. H. Lee, Y. Goto, M. Refregiers and J. Kardos (2015). "Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy." Proc Natl Acad Sci U S A **112**(24): E3095-3103.

Mikhael, J. R., S. R. Schuster, V. H. Jimenez-Zepeda, N. Bello, J. Spong, C. B. Reeder, A. K. Stewart, P. L. Bergsagel and R. Fonseca (2012). "Cyclophosphamide-bortezomib-dexamethasone (CyBorD) produces rapid and complete hematologic response in patients with AL amyloidosis." Blood **119**(19): 4391-4394.

Milanesi, L., J. P. Waltho, C. A. Hunter, D. J. Shaw, G. S. Beddard, G. D. Reid, S. Dev and M. Volk (2012). "Measurement of energy landscape roughness of folded and unfolded proteins." Proceedings of the National Academy of Sciences of the United States of America **109**(48): 19563-19568.

Miller, C. R., K. H. Lee, H. A. Wichman and F. M. Ytreberg (2014). "Changing Folding and Binding Stability in a Viral Coat Protein: A Comparison between Substitutions Accessible through Mutation and Those Fixed by Natural Selection." PLOS ONE **9**(11): e112988.

Mirecka, E. A., H. Shaykhalishahi, A. Gauhar, S. Akgul, J. Lecher, D. Willbold, M. Stoldt and W. Hoyer (2014). "Sequestration of a beta-hairpin for control of alpha-synuclein aggregation." Angew Chem Int Ed Engl **53**(16): 4227-4230.

Mishra, S., J. Guan, E. Plovie, D. C. Seldin, L. H. Connors, G. Merlini, R. H. Falk, C. A. MacRae and R. Liao (2013). "Human amyloidogenic light chain proteins result in cardiac dysfunction, cell death, and early mortality in zebrafish." American Journal of Physiology - Heart and Circulatory Physiology **305**(1): H95-H103.

Moal, I. H. and J. Fernandez-Recio (2012). "SKEMPI: a Structural Kinetic and Energetic database of Mutant Protein Interactions and its use in empirical models." Bioinformatics **28**(20): 2600-2607.

Mohty, D., T. Damy, P. Cosnay, N. Echahidi, D. Casset-Senon, P. Virot and A. Jaccard (2013). "Cardiac amyloidosis: Updates in diagnosis and management." Archives of Cardiovascular Diseases **106**(10): 528-540.

Monis, G. F., C. Schultz, R. Ren, J. Eberhard, C. Costello, L. Connors, M. Skinner and V. Trinkaus-Randall (2006). "Role of Endocytic Inhibitory Drugs on Internalization of Amyloidogenic Light Chains by Cardiac Fibroblasts." The American Journal of Pathology **169**(6): 1939-1952.

Morris, M. J., S. J. Craig, T. M. Sutherland, P. G. Board and M. G. Casarotto (2009). "Transport of glutathione transferase-fold structured proteins into living cells." Biochim Biophys Acta **1788**(3): 676-685.

Muchtar, E., F. K. Buadi, A. Dispenzieri and M. A. Gertz (2016). "Immunoglobulin light-chain amyloidosis: from basics to new developments in diagnosis, prognosis and therapy." Acta Haematol **135**.

Muchtar, E., F. K. Buadi, A. Dispenzieri and M. A. Gertz (2016). "Immunoglobulin Light-Chain Amyloidosis: From Basics to New Developments in Diagnosis, Prognosis and Therapy." Acta Haematologica **135**(3): 172-190.

Mukherjee, S., S. P. Pondaven, K. Hand, J. Madine and C. P. Jaroniec (2017). "Effect of amino acid mutations on the conformational dynamics of amyloidogenic immunoglobulin light-chains: A combined NMR and in silico study." Scientific Reports **7**(1): 10339.

Mukherjee, S., S. P. Pondaven, N. Hofer and C. P. Jaroniec (2009). "Backbone and side-chain (1)H, (13)C and (15)N resonance assignments of LEN, a human immunoglobulin kappaIV light-chain variable domain." Biomol NMR Assign **3**(2): 255-259.

Mulcahy, L. A., R. C. Pink and D. R. F. Carter (2014). "Routes and mechanisms of extracellular vesicle uptake." Journal of Extracellular Vesicles **3**: 10.3402/jev.v3403.24641.

Murphy, M. P. and H. LeVine, 3rd (2010). "Alzheimer's disease and the amyloid-beta peptide." J Alzheimers Dis **19**(1): 311-323.

Nelson, R., M. R. Sawaya, M. Balbirnie, A. Ø. Madsen, C. Riekel, R. Grothe and D. Eisenberg (2005). "Structure of the cross-β spine of amyloid-like fibrils." Nature **435**(7043): 773-778.

Noritake, K., T. Aki, T. Funakoshi, K. Unuma, A. Nara, C. Kato and K. Uemura (2012). "Critical roles of Rho-associated kinase in membrane blebbing and mitochondrial pathway of apoptosis caused by 1-butanol." Toxicol In Vitro **26**(6): 849-855.

Novotny, J. and E. Haber (1985). "Structural invariants of antigen binding: comparison of immunoglobulin VL-VH and VL-VL domain dimers." Proc Natl Acad Sci U S A **82**(14): 4592-4596.

Nozach, H., C. Fruchart-Gaillard, F. Fenaille, F. Beau, O. H. Ramos, B. Douzi, N. J. Saez, M. Moutiez, D. Servent, M. Gondry, R. Thai, P. Cuniasse, R. Vincentelli and V. Dive (2013). "High throughput screening identifies disulfide isomerase DsbC as a very efficient partner for recombinant expression of small disulfide-rich proteins in E. coli." Microb Cell Fact **12**: 37.

O'Connell, F. P., J. L. Pinkus and G. S. Pinkus (2004). "CD138 (syndecan-1), a plasma cell marker immunohistochemical profile in hematopoietic and nonhematopoietic neoplasms." Am J Clin Pathol **121**(2): 254-263.

O'Dell , J. R., C. E. Haire , N. Erikson , W. Drymalski , W. Palmer , P. J. Eckhoff , V. Garwood , P. Maloley , L. W. Klassen , S. Wees , H. Klein and G. F. Moore (1996). "Treatment of Rheumatoid Arthritis with Methotrexate Alone, Sulfasalazine and Hydroxychloroquine, or a Combination of All Three Medications." New England Journal of Medicine **334**(20): 1287-1291.

Ochman, H., A. S. Gerber and D. L. Hartl (1988). "Genetic applications of an inverse polymerase chain reaction." Genetics **120**(3): 621-623.

Ohage, E. C., W. Graml, M. M. Walter, S. Steinbacher and B. Steipe (1997). "Beta-turn propensities as paradigms for the analysis of structural motifs to engineer protein stability." Protein Science : A Publication of the Protein Society **6**(1): 233-241.

Olsen, K. E., K. Sletten and P. Westermark (1998). "Extended analysis of AL-amyloid protein from abdominal wall subcutaneous fat biopsy: kappa IV immunoglobulin light chain." Biochem Biophys Res Commun **245**(3): 713-716.

Onuchic, J. N., Z. Luthey-Schulten and P. G. Wolynes (1997). "Theory of protein folding: the energy landscape perspective." Annu Rev Phys Chem **48**: 545-600.

Onuchic, J. N. and P. G. Wolynes (2004). "Theory of protein folding." Current Opinion in Structural Biology **14**(1): 70-75.

Ow, S.-Y. and D. E. Dunstan (2014). "A brief overview of amyloids and Alzheimer's disease." Protein Science : A Publication of the Protein Society **23**(10): 1315-1331.

Padlan, E. A. (1994). "Anatomy of the antibody molecule." Mol Immunol **31**(3): 169-217.

Palladini, G. and G. Merlini (2016). "What is new in diagnosis and management of light chain amyloidosis?" **128**(2): 159-168.

Palladini, G., P. Milani, A. Foli, L. Obici, F. Lavatelli, M. Nuvolone, R. Caccialanza, S. Perlini and G. Merlini (2014). "Oral melphalan and dexamethasone grants extended survival with minimal toxicity in AL amyloidosis: long-term results of a risk-adapted approach." Haematologica **99**(4): 743-750.

Pandurangan, A. P., B. Ochoa-Montano, D. B. Ascher and T. L. Blundell (2017). "SDM: a server for predicting effects of mutations on protein stability." Nucleic Acids Res.

Panigrahi, P., M. Sule, A. Ghanate, S. Ramasamy and C. G. Suresh (2015). "Engineering Proteins for Thermostability with iRDP Web Server." PLOS ONE **10**(10): e0139486.

Park, S. J. and S. Y. Lee (1998). "Efficient recovery of secretory recombinant proteins from protease negative mutant Escherichia coli strains." Biotechnology Techniques **12**(11): 815-818.

Pelaez-Aguilar, A. E., L. Rivillas-Acevedo, L. French-Pacheco, G. Valdes-Garcia, R. Maya-Martinez, N. Pastor and C. Amero (2015). "Inhibition of Light Chain 6aJL2-R24G Amyloid Fiber Formation Associated with Light Chain Amyloidosis." Biochemistry **54**(32): 4978-4986.

Peng, Y., J. Norris, C. Schwartz and E. Alexov (2016). "Revealing the Effects of Missense Mutations Causing Snyder-Robinson Syndrome on the Stability and Dimerization of Spermine Synthase." International Journal of Molecular Sciences **17**(1): 77.

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

Perera, R. M., R. Zoncu, T. G. Johns, M. Pypaert, F.-T. Lee, I. Mellman, L. J. Old, D. K. Toomre and A. M. Scott (2007). "Internalization, Intracellular

Trafficking, and Biodistribution of Monoclonal Antibody 806: A Novel Anti-Epidermal Growth Factor Receptor Antibody." Neoplasia (New York, N.Y.) **9**(12): 1099-1110.

Peter, A. K., M. A. Bjerke and L. A. Leinwand (2016). "Biology of the cardiac myocyte in heart disease." Molecular Biology of the Cell **27**(14): 2149-2160.

Peterson, F. C., E. M. Baden, B. A. Owen, B. F. Volkman and M. Ramirez-Alvarado (2010). "A single mutation promotes amyloidogenicity through a highly promiscuous dimer interface." Structure **18**(5): 563-570.

Pinotsi, D., A. K. Buell, C. Galvagnion, C. M. Dobson, G. S. Kaminski Schierle and C. F. Kaminski (2014). "Direct observation of heterogeneous amyloid fibril growth kinetics via two-color super-resolution microscopy." Nano Lett **14**(1): 339-345.

Pires, D. E., D. B. Ascher and T. L. Blundell (2014). "mCSM: predicting the effects of mutations in proteins using graph-based signatures." Bioinformatics **30**(3): 335-342.

Pires, D. E. V., D. B. Ascher and T. L. Blundell (2014). "DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach." Nucleic Acids Research **42**(Web Server issue): W314-W319.

Pokkuluri, P. R., X. Cai, G. Johnson, F. J. Stevens and M. Schiffer (2000). "Change in dimerization mode by removal of a single unsatisfied polar residue located at the interface." Protein Sci **9**(9): 1852-1855.

Pokkuluri, P. R., D. B. Huang, R. Raffen, X. Cai, G. Johnson, P. W. Stevens, F. J. Stevens and M. Schiffer (1998). "A domain flip as a result of a single amino-acid substitution." Structure **6**(8): 1067-1073.

Pokkuluri, P. R., D. B. Huang, R. Raffen, X. Cai, G. Johnson, P. W. Stevens, F. J. Stevens and M. Schiffer (1998). "A domain flip as a result of a single amino-acid substitution." Structure **6**(8): 1067-1073.

Pokkuluri, P. R., R. Raffen, L. Dieckman, C. Boogaard, F. J. Stevens and M. Schiffer (2002). "Increasing protein stability by polar surface residues: domain-wide consequences of interactions within a loop." Biophysical Journal **82**(1 Pt 1): 391-398.

Pondaven, S. P. (2012). Conformational Flexibility and Amyloid Core Characterization of Human Immunoglobulin Light Chain Domains by Multidimensional NMR Spectroscopy. Doctor of Philosophy The Ohio State University.

Poshusta, T. L., N. Katoh, M. A. Gertz, A. Dispenzieri and M. Ramirez-Alvarado (2013). "Thermal stability threshold for amyloid formation in light chain amyloidosis." Int J Mol Sci **14**(11): 22604-22617.

Poshusta, T. L., L. A. Sikkink, N. Leung, R. J. Clark, A. Dispenzieri and M. Ramirez-Alvarado (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**(4): e5169.

Pras, M., M. Schubert, D. Zucker-Franklin, A. Rimon and E. C. Franklin (1968). "The characterization of soluble amyloid prepared in water." J Clin Invest **47**(4): 924-933.

Prokaeva, T., B. Spencer, M. Kaut, A. Ozonoff, G. Doros, L. H. Connors, M. Skinner and D. C. Seldin (2007). "Soft tissue, joint, and bone manifestations of AL amyloidosis: clinical presentation, molecular features, and survival." Arthritis Rheum **56**(11): 3858-3868.

Puckett, C. A. and J. K. Barton (2009). "Fluorescein Redirects a Ruthenium-Octaarginine Conjugate to the Nucleus." Journal of the American Chemical Society **131**(25): 8738-8739.

Qin, Z., D. Hu, M. Zhu and A. L. Fink (2007). "Structural characterization of the partially folded intermediates of an immunoglobulin light chain leading to amyloid fibrillation and amorphous aggregation." Biochemistry **46**(11): 3521-3531.

Qin, Z., Y. Sun, B. Jia, D. Wang, Y. Ma and G. Ma (2017). "Kinetic Mechanism of Thioflavin T Binding onto the Amyloid Fibril of Hen Egg White Lysozyme." Langmuir **33**(22): 5398-5405.

Raffen, R., L. J. Dieckman, M. Szpunar, C. Wunschl, P. R. Pokkuluri, P. Dave, P. Wilkins Stevens, X. Cai, M. Schiffer and F. J. Stevens (1999). "Physicochemical consequences of amino acid variations that contribute to fibril formation by immunoglobulin light chains." Protein Sci **8**(3): 509-517.

Raffen, R., P. W. Stevens, C. Boogaard, M. Schiffer and F. J. Stevens (1998). "Reengineering immunoglobulin domain interactions by introduction of charged residues." Protein Engineering, Design and Selection **11**(4): 303-309.

Rajkumar, S. V., A. Dispenzieri and R. A. Kyle (2006). "Monoclonal gammopathy of undetermined significance, Waldenstrom macroglobulinemia, AL amyloidosis, and related plasma cell disorders: diagnosis and treatment." Mayo Clin Proc **81**(5): 693-703.

Rambaran, R. N. and L. C. Serpell (2008). "Amyloid fibrils: Abnormal protein assembly." Prion **2**(3): 112-117.

Ramirez-Alvarado, M. (2007). Protein Misfolding, Aggregation and Conformational Diseases. A. L. F. Vladimir N. Uversky, Springer**:** 188-190.

Ramirez-Alvarado, M. (2012). "Amyloid formation in light chain amyloidosis." Curr Top Med Chem **12**(22): 2523-2533.

Ramirez-Alvarado, M. (2012). "Amyloid Formation in Light Chain Amyloidosis()." Current topics in medicinal chemistry **12**(22): 2523-2533.

Ramirez-Alvarado, M., J. K. De Stigter, E. M. Baden, L. A. Sikkink, R. W. McLaughlin and A. L. Taboas (2007). Immunoglobulin Light Chain and Systemic Light-Chain Amyloidosis. Protein Misfolding, Aggregation, and Conformational Diseases: Part B: Molecular Mechanisms of Conformational Diseases. V. N. Uversky and A. L. Fink. Boston, MA, Springer US**:** 183-197.

Randles, E. G., J. R. Thompson, D. J. Martin and M. Ramirez-Alvarado (2009). "Structural alterations within native amyloidogenic immunoglobulin light chains." Journal of molecular biology **389**(1): 199-210.

Randles, E. G., J. R. Thompson, D. J. Martin and M. Ramirez-Alvarado (2009). "Structural alterations within native amyloidogenic immunoglobulin light chains." J Mol Biol **389**(1): 199-210.

Rapp, C. S. and R. M. Pollack (2005). "Crystal packing effects on protein loops." Proteins **60**(1): 103-109.

Ravn, P., C. Madhurantakam, S. Kunze, E. Matthews, C. Priest, S. O'Brien, A. Collinson, M. Papworth, M. Fritsch-Fredin, L. Jermutus, L. Benthem, M. Gruetter and R. H. Jackson (2013). "Structural and pharmacological characterization of novel potent and selective monoclonal antibody antagonists of glucose-dependent insulinotropic polypeptide receptor." J Biol Chem **288**(27): 19760-19772.

Redler, R. L., J. Das, J. R. Diaz and N. V. Dokholyan (2016). "Protein Destabilization as a Common Factor in Diverse Inherited Disorders." Journal of molecular evolution **82**(1): 11-16.

Reece, D. E., V. Sanchorawala, U. Hegenbart, G. Merlini, G. Palladini, J.-P. Fermand, R. A. Vescio, X. Liu, Y. A. Elsayed, A. Cakana and R. L. Comenzo (2009). "Weekly and twice-weekly bortezomib in patients with systemic AL amyloidosis: results of a phase 1 dose-escalation study." Blood **114**(8): 1489-1497.

Repnik, U., M. H. Česen and B. Turk (2013). "The Endolysosomal System in Cell Death and Survival." Cold Spring Harbor Perspectives in Biology **5**(1): a008755.

Richards, D. B., L. M. Cookson, A. C. Berges, S. V. Barton, T. Lane, J. M. Ritter, M. Fontana, J. C. Moon, M. Pinzani, J. D. Gillmore, P. N. Hawkins and M. B. Pepys (2015). "Therapeutic Clearance of Amyloid by Antibodies to Serum Amyloid P Component." New England Journal of Medicine **373**(12): 1106-1114.

Riek, R. and D. S. Eisenberg (2016). "The activities of amyloids from a structural perspective." Nature **539**(7628): 227-235.

Robinson, G. W. and C. H. Cho (1999). "Role of hydration water in protein unfolding." Biophysical Journal **77**(6): 3311-3318.

Rocco Moretti, L. N. (2012). "Documentation for relax pdb with all-atom constraints / prepare pdb for rosetta with relax." Retrieved 16/08, 2017, from https://www.rosettacommons.org/manuals/archive/rosetta3.4_user_guide/d9/d2e/prepare_pdb_for_rosetta_with_relax.html.

Rochet, J.-C. and P. T. Lansbury (2000). "Amyloid fibrillogenesis: themes and variations." Current Opinion in Structural Biology **10**(1): 60-68.

Rognoni, P., F. Lavatelli, S. Casarini, G. Palladini, L. Verga, P. Pedrazzoli, G. Valentini, G. Merlini and V. Perfetti (2013). "A strategy for synthesis of pathogenic human immunoglobulin free light chains in E. coli." PLoS One **8**(9): e76022.

Rose, G. D., P. J. Fleming, J. R. Banavar and A. Maritan (2006). "A backbone-based theory of protein folding." Proceedings of the National Academy of Sciences **103**(45): 16623-16633.

Rose, G. D., P. J. Fleming, J. R. Banavar and A. Maritan (2006). "A backbone-based theory of protein folding." Proc Natl Acad Sci U S A **103**(45): 16623-16633.

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

Ross, C. A. and M. A. Poirier (2005). "What is the role of protein aggregation in neurodegeneration?" Nat Rev Mol Cell Biol **6**(11): 891-898.

Roy, V. (2012). "Autologous Stem Cell Transplant for AL Amyloidosis." Bone Marrow Research **2012**: 5.

Saccon, R. A., R. K. A. Bunton-Stasyshyn, E. M. C. Fisher and P. Fratta (2013). "Is SOD1 loss of function involved in amyotrophic lateral sclerosis?" Brain **136**(8): 2342-2358.

Saelices, L., L. M. Johnson, W. Y. Liang, M. R. Sawaya, D. Cascio, P. Ruchala, J. Whitelegge, L. Jiang, R. Riek and D. S. Eisenberg (2015). "Uncovering the Mechanism of Aggregation of Human Transthyretin." The Journal of Biological Chemistry **290**(48): 28932-28943.

Sanchorawala, V. (2006). "Light-Chain (AL) Amyloidosis: Diagnosis and Treatment." Clinical Journal of the American Society of Nephrology **1**(6): 1331-1341.

Sanchorawala, V., E. Blanchard, D. C. Seldin, C. O'Hara, M. Skinner and D. G. Wright (2006). "AL amyloidosis associated with B-cell lymphoproliferative disorders: frequency and treatment outcomes." Am J Hematol **81**(9): 692-695.

Sarroukh, R., E. Goormaghtigh, J. M. Ruysschaert and V. Raussens (2013). "ATR-FTIR: a "rejuvenated" tool to investigate amyloid proteins." Biochim Biophys Acta **1828**(10): 2328-2338.

Sawaya, M. R., S. Sambashivan, R. Nelson, M. I. Ivanova, S. A. Sievers, M. I. Apostol, M. J. Thompson, M. Balbirnie, J. J. Wiltzius, H. T. McFarlane, A. O. Madsen, C. Riekel and D. Eisenberg (2007). "Atomic structures of amyloid cross-beta spines reveal varied steric zippers." Nature **447**(7143): 453-457.

Scheinost, J. C., G. E. Boldt and P. Wentworth (2012). Protein Misfolding and Disease. Chemical Biology, John Wiley & Sons, Inc.**:** 379-400.

Schroder, R. and R. P. Linke (1999). "Cerebrovascular involvement in systemic AA and AL amyloidosis: a clear haematogenic pattern." Virchows Arch **434**(6): 551-560.

Schrodinger, LLC (2015). The PyMOL Molecular Graphics System, Version 1.8.

Schroeder, H. W. and L. Cavacini (2010). "Structure and Function of Immunoglobulins." The Journal of allergy and clinical immunology **125**(2 0 2): S41-S52.

Schymkowitz, J., J. Borg, F. Stricher, R. Nys, F. Rousseau and L. Serrano (2005). "The FoldX web server: an online force field." Nucleic Acids Res **33**(Web Server issue): W382-388.

Selkoe, D. J. (2003). "Folding proteins in fatal ways." Nature **426**(6968): 900-904.

Selkoe, D. J. (2013). "Folding proteins in fatal ways." Nature **426**(1476-4687 (Electronic)): 900-904.

Serpell, L. C. (2000). "Alzheimer's amyloid fibrils: structure and assembly." Biochim Biophys Acta **1502**(1): 16-30.

Serpell, L. C. (2000). "Alzheimer’s amyloid fibrils: structure and assembly." Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease **1502**(1): 16-30.

Shaykhalishahi, H. (2015). Selection and Characterization of Engineered Binding Proteins to a-Synuclein, Universität Düsseldorf.

Shaykhalishahi, H., E. A. Mirecka, A. Gauhar, C. S. Gruning, D. Willbold, T. Hard, M. Stoldt and W. Hoyer (2015). "A beta-hairpin-binding protein for three different disease-related amyloidogenic proteins." Chembiochem **16**(3): 411-414.

Shi, J., J. Guan, B. Jiang, D. A. Brenner, F. del Monte, J. E. Ward, L. H. Connors, D. B. Sawyer, M. J. Semigran, T. E. Macgillivray, D. C. Seldin, R. Falk and R. Liao (2010). "Amyloidogenic light chains induce cardiomyocyte contractile dysfunction and apoptosis via a non-canonical p38α MAPK pathway." Proceedings of the National Academy of Sciences of the United States of America **107**(9): 4188-4193.

Shin, J. T., J. E. Ward, P. A. Collins, M. Dai, H. L. Semigran, M. J. Semigran and D. C. Seldin (2012). "Overexpression of human amyloidogenic light chains causes heart failure in embryonic zebrafish: a preliminary report." Amyloid **19**(4): 191-196.

Shvartsburg, A. A. and M. F. Jarrold (1996). "An exact hard-spheres scattering model for the mobilities of polyatomic ions." Chemical Physics Letters **261**(1): 86-91.

Sikkink, L. A. and M. Ramirez-Alvarado (2010). "Cytotoxicity of amyloidogenic immunoglobulin light chains in cell culture." Cell Death and Dis **1**: e98.

Simpson, E. R., E. M. Herold and J. Buchner (2009). "The Folding Pathway of the Antibody VL Domain." Journal of Molecular Biology **392**(5): 1326-1338.

Singh, A., V. Upadhyay, A. K. Upadhyay, S. M. Singh and A. K. Panda (2015). "Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process." Microbial Cell Factories **14**: 41.

Singh, G. (2017). "Serum Free Light Chain Assay and κ/λ Ratio: Performance in Patients With Monoclonal Gammopathy-High False Negative Rate for κ/λ Ratio." Journal of Clinical Medicine Research **9**(1): 46-57.

Sipe, J. D., M. D. Benson, J. N. Buxbaum, S. I. Ikeda, G. Merlini, M. J. Saraiva and P. Westermark (2016). "Amyloid fibril proteins and amyloidosis: chemical identification and clinical classification International Society of Amyloidosis 2016 Nomenclature Guidelines." Amyloid **23**(4): 209-213.

Sipe, J. D. and A. S. Cohen (2000). "Review: history of the amyloid fibril." J Struct Biol **130**(2-3): 88-98.

Sivakumar, S. and S. Niranjali Devaraj (2014). "Tertiary structure prediction and identification of druggable pocket in the cancer biomarker – Osteopontin-c." Journal of Diabetes and Metabolic Disorders **13**: 13-13.

Smith, D. P., S. E. Radford and A. E. Ashcroft (2010). "Elongated oligomers in β2-microglobulin amyloid assembly revealed by ion mobility spectrometry-mass spectrometry." Proceedings of the National Academy of Sciences **107**(15): 6794-6798.

Smith, David P., Lucy A. Woods, Sheena E. Radford and Alison E. Ashcroft (2011). "Structure and Dynamics of Oligomeric Intermediates in β(2)-Microglobulin Self-Assembly." Biophysical Journal **101**(5): 1238-1247.

Smith, J. F., T. P. Knowles, C. M. Dobson, C. E. Macphee and M. E. Welland (2006). "Characterization of the nanoscale properties of individual amyloid fibrils." Proc Natl Acad Sci U S A **103**(43): 15806-15811.

Solomon, A. (1985). "Light chains of human immunoglobulins." Methods Enzymol **116**: 101-121.

Soskine, M. and D. S. Tawfik (2010). "Mutational effects and the evolution of new protein functions." Nat Rev Genet **11**(8): 572-582.

Soto, C. (2001). "Protein misfolding and disease; protein refolding and therapy." FEBS Lett **498**(2-3): 204-207.

Souillac, P. O., V. N. Uversky and A. L. Fink (2003). "Structural transformations of oligomeric intermediates in the fibrillation of the immunoglobulin light chain LEN." Biochemistry **42**(26): 8094-8104.

Souillac, P. O., V. N. Uversky, I. S. Millett, R. Khurana, S. Doniach and A. L. Fink (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**(15): 12657-12665.

Sowmya, G., E. J. Breen and S. Ranganathan (2015). "Linking structural features of protein complexes and biological function." Protein Science : A Publication of the Protein Society **24**(9): 1486-1494.

Sowmya, G. and S. Ranganathan (2015). "Discrete structural features among interface residue-level classes." BMC Bioinformatics **16**(Suppl 18): S8-S8.

Spencer, N. Y. and J. F. Engelhardt (2014). "The basic biology of redoxosomes in cytokine-mediated signal transduction and implications for disease-specific therapies." Biochemistry **53**(10): 1551-1564.

Srinivasan, R. and G. D. Rose (1999). "A physical basis for protein secondary structure." Proc Natl Acad Sci U S A **96**(25): 14258-14263.

Stefani, M. (2004). "Protein misfolding and aggregation: new examples in medicine and biology of the dark side of the protein world." Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease **1739**(1): 5-25.

Stefani, M. and S. Rigacci (2013). "Protein Folding and Aggregation into Amyloid: The Interference by Natural Phenolic Compounds." International Journal of Molecular Sciences **14**(6): 12411-12457.

Stefani, M. and S. Rigacci (2013). "Protein folding and aggregation into amyloid: the interference by natural phenolic compounds." Int J Mol Sci **14**(6): 12411-12457.

Stevens, F. J. (2000). "Four structural risk factors identify most fibril-forming kappa light chains." Amyloid **7**(3): 200-211.

Stevens, F. J. and M. Schiffer (1981). "Computer simulation of protein self-association during small-zone gel filtration. Estimation of equilibrium constants." Biochem J **195**(1): 213-219.

Stevens, F. J. and M. Schiffer (1995). Structure and Properties of Human Immunoglobulin Light-Chain Dimers. Antibody Engineering Protocols. S. Paul. Totowa, NJ, Humana Press**:** 51-81.

Stevens, P. W., R. Raffen, D. K. Hanson, Y. L. Deng, M. Berrios-Hammond, F. A. Westholm, C. Murphy, M. Eulitz, R. Wetzel, A. Solomon 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**(3): 421-432.

Studer, R. A. (2015). "Estimating the stability effect of a mutation with FoldX." Retrieved 15/08/2017, 2017, from https://evosite3d.blogspot.co.uk/2015/03/tutorial-estimating-stability-effect-of.html.

Studer, R. A., P. A. Christin, M. A. Williams and C. A. Orengo (2014). "Stability-activity tradeoffs constrain the adaptive evolution of RubisCO." Proc Natl Acad Sci U S A **111**(6): 2223-2228.

Su, Y. and P. T. Chang (2001). "Acidic pH promotes the formation of toxic fibrils from beta-amyloid peptide." Brain Res **893**(1-2): 287-291.

Taniguchi, A., Y. Shimizu, K. Oisaki, Y. Sohma and M. Kanai (2016). "Switchable photooxygenation catalysts that sense higher-order amyloid structures." Nat Chem **8**(10): 974-982.

Teng, J., W. J. Russell, X. Gu, J. Cardelli, M. L. Jones and G. A. Herrera (2004). "Different types of glomerulopathic light chains interact with mesangial cells using a common receptor but exhibit different intracellular trafficking patterns." Lab Invest **84**(4): 440-451.

The, T. H. and T. E. W. Feltkamp (1970). "Conjugation of fluorescein isothiocyanate to antibodies: I. Experiments on the conditions of conjugation." Immunology **18**(6): 865-873.

Theobald, D. L. and D. S. Wuttke (2006). "THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures." Bioinformatics **22**(17): 2171-2172.

Thiltgen, G. and R. A. Goldstein (2012). "Assessing Predictors of Changes in Protein Stability upon Mutation Using Self-Consistency." PLoS ONE **7**(10): e46084.

Tokuriki, N., F. Stricher, J. Schymkowitz, L. Serrano and D. S. Tawfik (2007). "The stability effects of protein mutations appear to be universally distributed." J Mol Biol **369**(5): 1318-1332.

Tokuriki, N., F. Stricher, L. Serrano and D. S. Tawfik (2008). "How Protein Stability and New Functions Trade Off." PLoS Computational Biology **4**(2): e1000002.

Tomovic, A. and E. J. Oakeley (2008). "Computational Structural Analysis: Multiple Proteins Bound to DNA." PLoS ONE **3**(9): e3243.

Tosi, P. (2013). "Diagnosis and treatment of bone disease in multiple myeloma: spotlight on spinal involvement." Scientifica (Cairo) **2013**: 104546.

Toyama, B. H. and J. S. Weissman (2011). "Amyloid Structure: Conformational Diversity and Consequences." Annual review of biochemistry **80**: 10.1146/annurev-biochem-090908-120656.

Trinkaus-Randall, V., M. T. Walsh, S. Steeves, G. Monis, L. H. Connors and M. Skinner (2005). "Cellular Response of Cardiac Fibroblasts to Amyloidogenic Light Chains." The American Journal of Pathology **166**(1): 197-208.

Tsolis, A. C., N. C. Papandreou, V. A. Iconomidou and S. J. Hamodrakas (2013). "A Consensus Method for the Prediction of 'Aggregation-Prone' Peptides in Globular Proteins." PLOS ONE **8**(1): e54175.

Turoverov, K. K., I. M. Kuznetsova and V. N. Uversky (2010). "The Protein Kingdom Extended: Ordered and Intrinsically Disordered Proteins, Their Folding, Supramolecular Complex Formation, and Aggregation." Progress in biophysics and molecular biology **102**(2-3): 73-84.

Uversky, V. N., J. Li and A. L. Fink (2001). "Metal-triggered structural transformations, aggregation, and fibrillation of human alpha-synuclein. A possible molecular NK between Parkinson's disease and heavy metal exposure." J Biol Chem **276**(47): 44284-44296.

Vallejo, L. F. and U. Rinas (2004). "Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins." Microbial Cell Factories **3**: 11-11.

Van Durme, J., J. Delgado, F. Stricher, L. Serrano, J. Schymkowitz and F. Rousseau (2011). "A graphical interface for the FoldX forcefield." Bioinformatics **27**(12): 1711-1712.

Verma, M., A. Vats and V. Taneja (2015). "Toxic species in amyloid disorders: Oligomers or mature fibrils." Annals of Indian Academy of Neurology **18**(2): 138-145.

Vincent, M. and S. Schnell (2016). "A collection of intrinsic disorder characterizations from eukaryotic proteomes." Scientific Data **3**: 160045.

Voisine, C., J. S. Pedersen and R. I. Morimoto (2010). "Chaperone networks: tipping the balance in protein folding diseases." Neurobiol Dis **40**(1): 12-20.

Volkamer, A., D. Kuhn, F. Rippmann and M. Rarey (2012). "DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment." Bioinformatics **28**(15): 2074-2075.

Vrana, J. A., J. D. Gamez, B. J. Madden, J. D. Theis, H. R. Bergen and A. Dogan (2009). "Classification of amyloidosis by laser microdissection and mass spectrometry–based proteomic analysis in clinical biopsy specimens." Blood **114**(24): 4957-4959.

Wall, J. S., S. J. Kennel, A. Williams, T. Richey, A. Stuckey, Y. Huang, S. Macy, R. Donnell, R. Barbour, P. Seubert and D. Schenk (2012). "AL Amyloid Imaging and Therapy with a Monoclonal Antibody to a Cryptic Epitope on Amyloid Fibrils." PLOS ONE **7**(12): e52686.

Walsh, D. M., B. P. Tseng, R. E. Rydel, M. B. Podlisny and D. J. Selkoe (2000). "The oligomerization of amyloid beta-protein begins intracellularly in cells derived from human brain." Biochemistry **39**(35): 10831-10839.

Walsh, I., F. Seno, S. C. E. Tosatto and A. Trovato (2014). "PASTA 2.0: an improved server for protein aggregation prediction." Nucleic Acids Research **42**(W1): W301-W307.

Wang, Y., D. Li, H. Fan, L. Tian, Y. Zhong, Y. Zhang, L. Yuan, C. Jin, C. Yin and D. Ma (2006). "Cellular uptake of exogenous human PDCD5 protein." J Biol Chem **281**(34): 24803-24817.

Watkins, S. J., G. M. Borthwick and H. M. Arthur (2011). "The H9C2 cell line and primary neonatal cardiomyocyte cells show similar hypertrophic responses in vitro." In Vitro Cell Dev Biol Anim **47**(2): 125-131.

Wechalekar, A. D., H. J. Lachmann, M. Offer, P. N. Hawkins and J. D. Gillmore (2008). "Efficacy of bortezomib in systemic AL amyloidosis with relapsed/refractory clonal disease." Haematologica **93**(2): 295-298.

Wei, A. and S. Juneja (2003). "Bone marrow immunohistology of plasma cell neoplasms." Journal of Clinical Pathology **56**(6): 406-411.

Westermark GT, J. K., Westermark P. (1999). "Staining methods for identification of amyloid in tissue." methods in enzymology(0076-6879 (Print)).

Westermark, P. (2005). "Aspects on human amyloid forms and their fibril polypeptides." Febs j **272**(23): 5942-5949.

Westermark, P. (2012). "Localized AL amyloidosis: A suicidal neoplasm?" Upsala Journal of Medical Sciences **117**(2): 244-250.

Wilkins, M. R., E. Gasteiger, A. Bairoch, J. C. Sanchez, K. L. Williams, R. D. Appel and D. F. Hochstrasser (1999). "Protein identification and analysis tools in the ExPASy server." Methods Mol Biol **112**: 531-552.

Wilkins, M. R., I. Lindskog, E. Gasteiger, A. Bairoch, J. C. Sanchez, D. F. Hochstrasser and R. D. Appel (1997). "Detailed peptide characterization using PEPTIDEMASS--a World-Wide-Web-accessible tool." Electrophoresis **18**(3-4): 403-408.

Wilson, M. R., J. J. Yerbury and S. Poon (2008). "Potential roles of abundant extracellular chaperones in the control of amyloid formation and toxicity." Mol Biosyst **4**(1): 42-52.

Winn, M. D., C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson (2011). "Overview of the CCP4 suite and current developments." Acta Crystallogr D Biol Crystallogr **67**(Pt 4): 235-242.

Wirths, O. and T. A. Bayer (2003). "Alpha-synuclein, Abeta and Alzheimer's disease." Prog Neuropsychopharmacol Biol Psychiatry **27**(1): 103-108.

Witek, P., A. Korga, F. Burdan, M. Ostrowska, B. Nosowska, M. Iwan and J. Dudka (2016). "The effect of a number of H9C2 rat cardiomyocytes passage on repeatability of cytotoxicity study results." Cytotechnology **68**(6): 2407-2415.

Witvliet, D. K., A. Strokach, A. F. Giraldo-Forero, J. Teyra, R. Colak and P. M. Kim (2016). "ELASPIC web-server: proteome-wide structure-based prediction of mutation effects on protein stability and binding affinity." Bioinformatics **32**(10): 1589-1591.

Woods, L. A., S. E. Radford and A. E. Ashcroft (2013). "Advances in ion mobility spectrometry-mass spectrometry reveal key insights into amyloid assembly." Biochim Biophys Acta **1834**(6): 1257-1268.

Wright, P. E. and H. J. Dyson (2015). "Intrinsically disordered proteins in cellular signalling and regulation." Nat Rev Mol Cell Biol **16**(1): 18-29.

Xiang, Z. (2006). "Advances in Homology Protein Structure Modeling." Current protein & peptide science **7**(3): 217-227.

Xu, Z., R. Paparcone and M. J. Buehler (2010). "Alzheimer's Aβ(1-40) Amyloid Fibrils Feature Size-Dependent Mechanical Properties." Biophysical Journal **98**(10): 2053-2062.

Yasuda, S., T. Yoshidome, H. Oshima, R. Kodama, Y. Harano and M. Kinoshita (2010). "Effects of side-chain packing on the formation of secondary structures in protein folding." J Chem Phys **132**(6): 065105.

Yin, S., F. Ding and N. V. Dokholyan (2007). "Modeling backbone flexibility improves protein stability estimation." Structure **15**(12): 1567-1576.

Yoshimura, Y., Y. Lin, H. Yagi, Y. H. Lee, H. Kitayama, K. Sakurai, M. So, H. Ogi, H. Naiki and Y. Goto (2012). "Distinguishing crystal-like amyloid fibrils and glass-like amorphous aggregates from their kinetics of formation." Proc Natl Acad Sci U S A **109**(36): 14446-14451.

Yu, H., M.-J. Wang, N.-X. Xuan, Z.-C. Shang and J. Wu (2015). "Molecular dynamics simulation of the interactions between EHD1 EH domain and multiple peptides." Journal of Zhejiang University. Science. B **16**(10): 883-896.

Zambrano, R., M. Jamroz, A. Szczasiuk, J. Pujols, S. Kmiecik and S. Ventura (2015). "AGGRESCAN3D (A3D): server for prediction of aggregation properties of protein structures." Nucleic Acids Research **43**(W1): W306-W313.

Zhang, N. and D. Li (2017). "Druggability analysis of membrane proteins by DoGSiteScorer." PeerJ Preprints **5**: e2868v2861.

Zhang, Z., V. Martiny, D. Lagorce, Y. Ikeguchi, E. Alexov and M. A. Miteva (2014). "Rational Design of Small-Molecule Stabilizers of Spermine Synthase Dimer by Virtual Screening and Free Energy-Based Approach." PLOS ONE **9**(10): e110884.

Zhou, S., W. Sun, Z. Zhang and Y. Zheng (2014). "The Role of Nrf2-Mediated Pathway in Cardiac Remodeling and Heart Failure." Oxidative Medicine and Cellular Longevity **2014**: 260429.

Zhu, M., P. O. Souillac, C. Ionescu-Zanetti, S. A. Carter and A. L. Fink (2002). "Surface-catalyzed amyloid fibril formation." J Biol Chem **277**(52): 50914-50922.