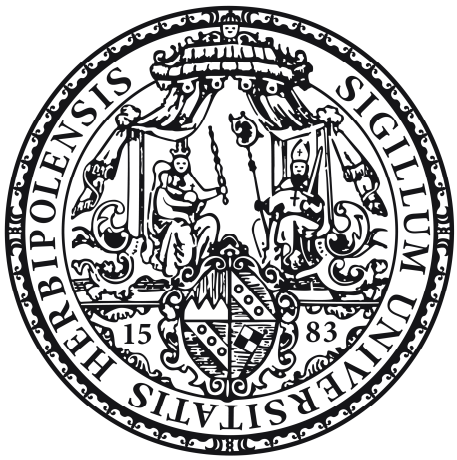
Insight into molecular mechanisms of folding and self-association of spider silk protein domains

Einblicke in molekulare Mechanismen der Faltung und Selbstassoziation von Spinnenseidenproteindomänen



Dissertation zur Erlangung des naturwissenschaftlichen Doktorgrades der Julius-Maximilians-Universität Würzburg

vorgelegt von

Julia Heiby

Geboren in Lohr am Main

Würzburg, 2019

Eingereicht am: .....

Mitglieder der Promotionskommission:

Vorsitzender: .....

Gutachter: PD Dr. Hannes Neuweiler

Gutachter: Prof. Dr. Thomas Müller

Tag des Promotionskolloquiums: .....

Doktorurkunde ausgehändigt am: .....

To my family

Never ignore coincidence.

Unless, of course, you're busy. In which case, always ignore coincidence.

— Doctor Who

Eidesstattliche Erklärungen nach §7 Abs. 2 Satz 3, 4, 5 der Promotionsordnung der Fakultät für Biologie

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, die Dissertation: „Einblicke in molekulare Mechanismen der Faltung und Selbstassoziation von Spinnenseidenproteindomänen“, eigenständig, d. h. insbesondere selbständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen, als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Weiterhin erkläre ich, dass bei allen Abbildungen und Texten bei denen die Verwertungsrechte (Copyright) nicht bei mir liegen, diese von den Rechtsinhabern eingeholt wurden und die Textstellen bzw. Abbildungen entsprechend den rechtlichen Vorgaben gekennzeichnet sind sowie bei Abbbildungen, die dem Internet entnommen wurden, der entsprechende Hypertextlink angegeben wurde.

Affidavit

I hereby declare that my thesis entitled: „Insight into molecular mechanisms of folding and self-association of spider silk protein domains” is the result of my own work. I did not receive any help or support from commercial consultants. All sources and / or materials applied are listed and specified in the thesis.

Furthermore I verify that the thesis has not been submitted as part of another examination process neither in identical nor in similar form.

Besides I declare that if I do not hold the copyright for figures and paragraphs, I obtained it from the rights holder and that paragraphs and figures have been marked according to law or for figures taken from the internet the hyperlink has been added accordingly.

Würzburg, den……

_____

Unterschrift

Summary

Spider silk is a biomaterial of extraordinary toughness paired with elasticity. The assembly of silk proteins, so-called spidroins (from "spider" and "fibroin"), generates the silk threads we typically see in our garden or the corners of our houses. Although spider webs from different species vary considerably in geometry and size, many sections of spidroin sequences are conserved. Highly conserved regions, found in all spidroins, relate to the terminal domains of the protein, i.e., the N-terminal (NTD) and C-terminal domains (CTD). Both have an essential function in the silk fibre association and polymerisation.

The NTD is a 14 kDa five-helix bundle, which self-associates via a pH-driven mechanism. This process is critical for starting the polymerisation of the fibre. However, detailed insights into how conserved this mechanism is in different species and the quantitative thermodynamic comparison between homologous NTDs was missing. For this reason, four homologous NTDs of the major ampullate gland (MaSp) from spider species *Euprosthenops australis*, *Nephila clavipes*, *Latrodectus hesperus*, and *Latrodectus geometricus* were investigated. I analysed and quantified equilibrium thermodynamics, kinetics of folding, and self-association. Methods involved dynamic light scattering (MALS), stopped-flow fluorescence and circular dichroism spectroscopy in combination with thermal and chemical denaturation experiments. The results showed conserved, cooperative two-state folding on a sub-millisecond time scale. All homologous NTDs showed a similarly fast association in the order of 10^9 $M^{-1}s^{-1}$, while the resulting equilibrium dissociation constants were in the low nanomolar range. Electrostatic forces were found to be of great importance for protein association. Monomeric protein stability increased with salt concentration while enhancing its folding speed. However, due to Debye-Hückel effects, we found intermolecular electrostatics to be shielded, which reduced the NTDs association capacity significantly at high ionic strength. Altogether, the energetics and kinetics of the NTD dimerisation was conserved for all analysed homologs.

Comparable to the NTD, the spider silks CTD is also a α-helix bundle, which covalently links two spidroins. The orientation of the domains predetermines the future fibre geometry. Here again, the detailed quantitative characterisation of the

folding and dimerisation was missing. Therefore, the CTD from the *E. australis* was analysed in-depth. The protein folded via a three-state mechanism and was placed in the family of knotted proteins.

By analysing the amino acid composition of the NTD of the MaSp1 of the *Euprosthenops australis*, we found an unusually high content of methionine residues (Met). To elucidate why this protein exhibits so many Met residues, I mutated all core Mets simultaneously to leucine (Leu). Results revealed a dramatically stabilised NTD, which now folded 50 times faster. After solving the tertiary structure of the mutant by NMR (nuclear magnetic resonance) spectroscopy, the structure of the monomeric mutant was found to be identical with the wild-type protein. However, when probing the dimerisation of the NTD, I could show that the association capacity was substantially impaired for the mutant. Our findings lead to the conclusion that Met provides the NTD with enhanced conformational dynamics and thus mobilises the protein, which results in tightly associated dimers. In additional experiments, I first re-introduced new Met residues into the Met-depleted protein at sequence positions containing native Leu. Hence, the mutated NTD protein was provided with the same number of Leu, which were previously removed by mutation. However, the protein did not regain wild-type characteristics. The functionality was not restored, but its stability was decreased as expected. To probe our hypothesis gained from the MaSp NTD, I transferred the experiment to another protein, namely the Hsp90 chaperone. Therefore, I incorporated methionine residues in the protein, which resulted in a slight improvement of its function.

Finally, trial experiments were performed aiming at the synthesis of shortened spidroin constructs containing less repetitive middle-segments than the wild-type protein. The objective was to study the findings of the terminal domains in the context of an intact spidroin. The synthesis of these engineered spidroins was challenging. Nevertheless, preliminary results encourage the assumption that the characteristics observed in the isolated domains hold true in the context of a full-length spidroin.

Zusammenfassung

Spinnenseide ist ein Biomaterial mit außergewöhnlicher Widerstandsfähigkeit welche gepaart ist mit Elastizität. Das Zusammenfügen von Seidenproteinen aus so genannten Spidroinen (ein Kunstwort aus „Spinne“ und „Fibroin“) erzeugt die Seidenfäden, die wir typischerweise in unseren Gärten oder in den Ecken unserer Häuser finden. Obwohl Spinnennetze von verschiedenen Spinnenarten in Geometrie und Größe stark variieren, sind große Teile der Spidroin-Sequenzen konserviert. Stark konservierte Bereiche, die in allen Spidroinen vorkommen, sind die endständigen Bereiche des Proteins, die N-terminale (NTD) und C-terminale Domäne (CTD) genannt werden. Beide haben wichtige Funktionen in der Assoziation der Proteine im Spinnkanal und deren Polymerisation zur Ausbildung des Seidenfadens.

Die NTD ist ein kleines 14 kDa Protein, bestehend aus einem Bündel aus fünf Helices, dessen Selbstorganisation pH-abhängig ist. Dieser Prozess leitet die Polymerisation der Faser ein. Allerdings fehlten bis heute Informationen darüber, ob dieser Mechanismus bei homologen Domänen aus verschiedenen Spinnenarten konserviert ist, da kaum quantitative biophysikalische Daten vorhanden sind. Aus diesem Grund wurden vier homologe NTDs der Spinnenarten *Euprosthenops australis*, *Nephila clavipes*, *Latrodectus hesperus* und *Latrodectus geometricus* vergleichend untersucht und deren Gleichgewichts-Thermodynamik, die Kinetik der Faltung und die Selbstassoziation quantitativ analysiert. Dazu wurden Methoden wie dynamische Mehrwinkel-Lichtstreuung (MALS), Stopped-Flow Fluoreszenzspektroskopie und Zirkulardichroismus in Kombination mit thermischen und chemischen Denaturierungsexperimenten angewandt. Die Ergebnisse lieferten die Erkenntnis einer kooperativen Zwei-Zustands-Faltung, die auf einer Zeitskala von weniger als einer Millisekunde stattfand. Alle homologen NTDs zeigten eine schnelle Assoziationsratenkonstante in der Größenordnung von 10^9 M^{-1} s^{-1}, während die Gleichgewichts-Dissoziationskonstante für alle Homologe im niedrigen nano-molaren Bereich lag. Die Proteinassoziation wurde durch elektrostatische Kräfte gesteuert, wobei hohe Salzkonzentrationen die Stabilität des monomeren Proteins und dessen Faltungsgeschwindigkeit erhöhten. Die Assoziation zweier Domänen wurde jedoch durch Abschirmung intermolekularer elektrostatischer Kräfte, dem Debye-Hückel-

Gesetz zufolge, reduziert. Die Energetik und Kinetik der NTD-Dimerisierung aller untersuchten Homologen erwies sich konserviert.

Ebenso wie die NTD, ist auch die CTD der Spinnenseide ein α-helikales Bündel, welche jedoch zwei Spidroine kovalent miteinander verbindet. Die Orientierung der verknüpften Domäne bestimmt bereits die zukünftige Faserstruktur. Ähnlich wie bei der NTD, waren Faltung und Dimerisierung der CTD bisher nicht im Detail beschrieben. Durch eine detaillierte Analyse der CTD der *E. australis* konnte gezeigt werden, dass das Protein sich in einem dreistufigen Mechanismus faltet und außerdem der Familie der geknoteten Proteine angehört.

Bei genauerer Betrachtung der Aminosäurezusammensetzung der *E. australis* NTD konnte ein ungewöhnlich hoher Anteil der Aminosäure Methionin (Met) festgestellt werden. Um diesen überraschenden Sachverhalt zu verstehen, habe ich alle im Kern liegenden Met zu Leucin (Leu) mutiert. Die Ergebnisse zeigten eine extrem stabilisierte NTD, welche sich nun 50-fach schneller faltete. Die Proteinstruktur der Mutante wurde in Lösung mittels NMR Spektroskopie ermittelt. Das Ergebnis lieferte deckungsgleiche Strukturen von Mutante und Wildtyp im monomeren Zustand. Allerdings zeigten NTD Dimerisierungs-Versuche, dass die Assoziationsfähigkeit der Mutante erheblich beeinträchtigt war. Untersuchungen der nativen Dynamik mittels NMR und Fluoreszenzkorrelationsspektroskopie zeigten, dass Met diese entscheidend verstärkt, was zu einem eng assoziierten Dimer führte. Im Versuch die Dynamik wieder künstlich herzustellen, habe ich neue Met in die Mutante eingeführt, auf Sequenzpositionen welche natürlicherweise Leu aufweisen. Somit wurde die ursprüngliche Anzahl an Met in der NTD wiederhergestellt, jedoch an anderen Positionen. Obwohl das Protein wie erwartet an Stabilität verlor, konnte dessen Funktionalität nicht wiederhergestellt werden. Um unsere Erkenntnisse auf andere Proteine zu übertragen, wurden Met Reste künstlich in ein Hsp90 Protein eingeführt. Es konnte eine leicht verbesserte Funktionalität des Proteins beobachtet werden.

Schließlich wurde versucht, die für die CTD und NTD gewonnen Erkenntnisse auf intakte, jedoch verkürzte Spidroine zu übertragen. Dazu wurden Spidroine mit weniger repetitiven Mittelsegmenten mittels rekombinanten Methoden hergestellt. Die Synthese dieser Spidroine erwies sich als Herausforderung. Allerdings zeigten die vorläufigen Ergebnisse, dass eine Verallgemeinerung der Erkenntnisse der isolierten Domänen auf das Volllängen-Spidroin möglich ist.

Table of Contents

1.Introduction

1.1. Spider silk

Spider silk is an extraordinary material. It outcompetes man-made fibres such as Kevlar or steel, being tougher, stronger, but also more flexible, considered by weight (Gosline et al. 1984). Material scientists consider it *the* fabric of the near future, as it has numerous possible applications, ranging from regenerative medicine like tissue engineering or drug delivery, to new industrial materials as nanofibres and surface coatings (Tokareva et al. 2013). Spiders can produce different types of silk, having each different properties. They are tailored to the different needs the spider encounters, including web-spinning, prey capture, shelter building and egg wrapping. The spiders form solid fibres out of proteins in solution in a fraction of seconds. Considering the different silk types, some have caught more interest than others. Especially for the toughest of them, the dragline silk, experts have tried to elucidate the detailed molecular mechanisms.

But not only spiders produce silk. Many species of arthropods, including caterpillars and mites, produce silk (Radtke 2016). When comparing the spider's silk to other silk types, particular focus lies on the silkworm *Bombyx mori*. Sanggaard and colleagues suggested 2014 that they share a common ancestor with the spiders. Interestingly, there are still significant similarities regarding the structural and functional properties of their glands and silk (Sanggaard et al. 2014). Nevertheless, spiders silk outperforms silkworm silk in toughness, extensibility and strength, which are substantially heightened for the spiders' dragline silk (table 1. 1).

Table 1. 1: Comparison of mechanical properties of different natural and man-made fibres. (Gosline et al. 1999)

Material	Strength (GPa)	Extensibility (%)	Toughness (MJ m^{-3})
Dragline silk	1.1	27	160
Bombyx mori silk	0.6	18	70
Tendon	0.15	12	7.5
Kevlar	3.6	2.7	50
Steel	1.5	0.8	6

Despite being less strong, the material of the silkworm has been used for centuries and is still prevalent today, mostly in fashion but also in clinical applications (Eisoldt et al. 2011). However, the latter recently showed problems regarding biocompatibility. Still, silkworms are easy to farm in industrial quantities, being the main reason that their silk has been used for so long. Spiders, on the other hand, are territorial, cannibalistic and cannot be farmed (Kluge et al. 2008). To harvest any significant quantity of spider silk, it takes considerable effort. During an eight-year project, textile artist Simon Peers and designer Nicholas Godley collected 1.2 million orb weavers on Madagascar together with a team of 80 people, to generate the tallest piece of clothing to date, a golden-coloured cape made with natural spider silk (Chung et al. 2012). The free-living spiders were collected every day and milked, before being released back into nature.

Such procedures cannot be made routinely as they are inefficient and not affordable. Therefore, the production of synthetical spider silk seems necessary. To date, several companies are aiming to produce and sell artificial spider silk. Most of them (for example AMSilk or Bolt Threads) desire to produce durable fabrics, which would be sustainable and environmentally friendly. In an age of global waste production, this is a huge selling point. Moreover, the creation of enhanced objects for space applications are pursued, with the need to be highly resistant and light-weighted (Tokareva et al. 2013). Companies have teamed up, like Airbus with AMSilk, aiming to build lighter aircraft, eventually leading to lower fuel consumption. To date, these projects have not made it into production despite being broadcasted extensively. It stands to reason that the silk those companies produce and use is still not sophisticated enough and is lacking the

unique qualities of natural spider silk. Curiously, most of the products are labelled as silk-like or silk-based.

Today, the most common way to produce artificial silk is by using hosts like *Escherichia coli (E. coli)*. The host gets introduced with foreign DNA encoding spider silk proteins, which the host will then produce recombinantly alongside its metabolites. With this, the main difficulty remains that artificially synthesised silk tends to aggregate easily, as natural silk proteins are big. Those proteins are called spidroins and comprise approximately 300 kDa. They are composed of a highly repetitive and long middle segment, flanked by short terminal domains (amino (N)- and carboxy (C)-terminal domain, namely NTD and CTD). The repetitive middle domain causes homologous recombination, deletions, transcription errors, accumulation in inclusion bodies and low yields during the production of recombinant spidroins in *E. coli* (Andersson et al. 2017). In comparison, there are approx. 80-100 times less repetitive segments in the middle domain of the *bombyx mori* fibroin (Mita et al. 1994). Ultimately, the mechanism of synthesis and fibre formation is yet not fully understood. There have been several studies regarding the specific chemico-physical properties of the spiders' glands, and many theories on how the spider stores and forms its silk (Vollrath and Knight 1999; Andersson et al. 2014; Kronqvist et al. 2014). However, to date, no one could reproduce artificial spidroins with the same qualities as natural spider silk.

1.2. Synthetic spider silk – state of the art

There are several approaches to the production and spinning of artificial spider silk, while man-spun fibres made from natural silk are also considered artificial. Therefore, we differentiate between regenerated and recombinantly produced fibres. Regenerated silk is reconstituted natural silk of dried protein powder, while recombinantly produced spidroins are purified proteins grown in *E. coli* and stored in aqueous solution. Regenerated silk is either wet or dry spun, while recombinant silk is always generated by wet spinning (Koeppel and Holland 2017). There are further spinning techniques like extrusion-based spinning and

electrospinning processes, which are restricted to natural silk. However, the tensile strength of, for example, electro-spun threads tend to be weak (Pérez-Rigueiro et al. 2018). Additionally, new techniques are developed to mimic the natural spinning process implying for example the pH-dependency (Andersson et al. 2017).

In short, wet-spinning needs (mostly) harsh non-solvent coagulation baths, leading to a controlled precipitation of the spidroins; dry-spinning, does not require a coagulation bath and the residual solvent clears by evaporation. The recent advance in the spinning of spider silk shows that considerable progress has been made with the recombinant wet spinning technique (Fig. 1. 1).

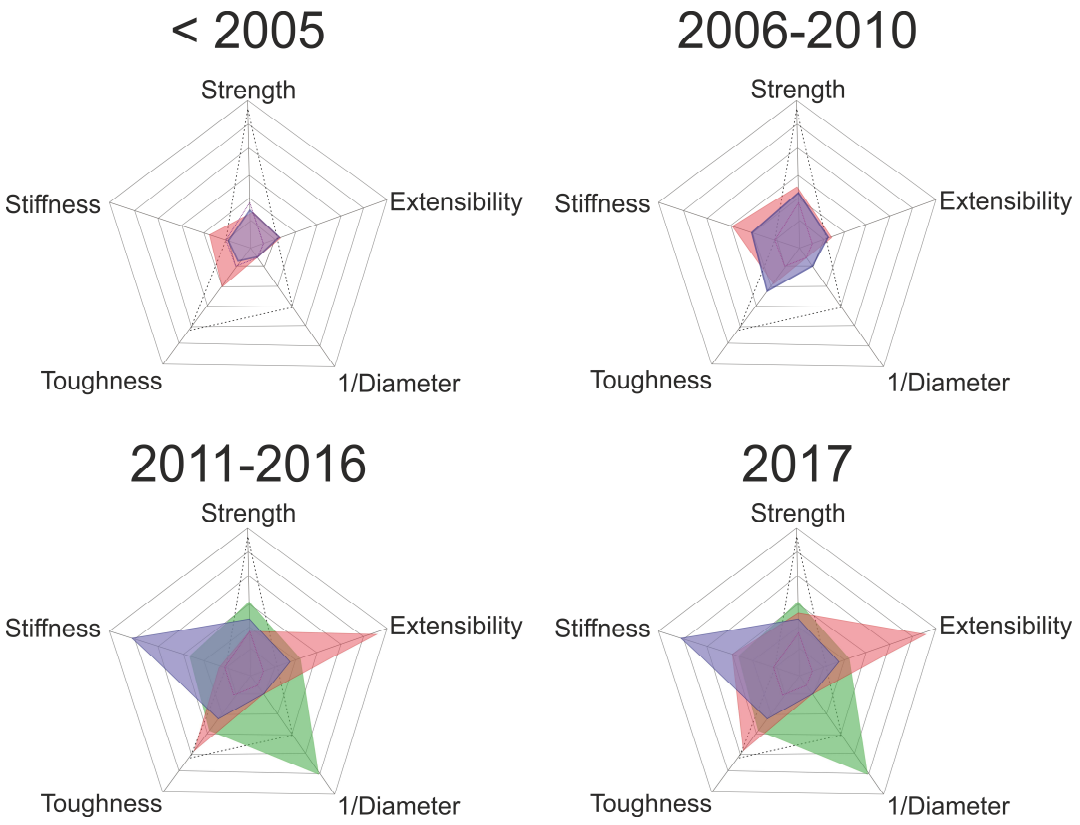


Figure 1. 1: Advance in artificial silk spinning in the last two decades, from before 2005 up to 2017, regarding mechanical properties. The different spinning techniques are coloured in blue for regenerated silk fibroin (RSF) wet spinning, in green for RSF dry spinning, in red for recombinant wet spun silk. As a comparison, the dragline silk from *N. edulis* (black dashed line) and *B. mori* (magenta dashed line) was added to the graphs. (adapted and with permission from Koeppel and Holland 2017)

The dilemma with recombinant wet spinning seems to be that the resulting threads are too thin. This result is consistent with the weaker strength those threads exhibit (Koeppel and Holland 2017). Furthermore, it could be shown that with increasing number of repetitive middle units, the strength of the resulting silk is also enhanced (Hayashi et al. 1999). The first so-called full-length spidroin (with approximately 285 kDa) was generated in 2010 by Xia et al. employing an engineered metabolic strategy, upregulating the glycyl-tRNA pool, meeting a better performance for the expression of glycine-rich proteins. The result was a stable fibre, as tough as the natural silk. The only downside to their silk was that they were missing the N- and C-terminal domains, small, but essential parts of the natural silk responsible for controlled fibre formation.

To overcome the problem of aggregation of long repetitive segments in the expressing host, Bowen and colleagues cleaved the sequence of a spidroin in half in the middle of the repetitive region (Bowen et al. 2018). They fused each half to a split intein (linker peptide) to eventually reassemble both halves after purification. The split intein covalently linked both halves of the spidroins, by a peptide bond. The lyophilised proteins were wet spun to fibres allegedly exhibiting the mechanical properties matching those of natural spider silk, but missing out, here again, the important functionally supporting roles of the naturally occurring terminal domains.

In an attempt to express spidroins recombinantly, other hosts than bacteria have been employed. It was attempted to produce silk proteins in plants, insects and even mammals. Notably, the expression of spidroins in silkworms has been a breakthrough in 2019, by combining the DNA of spider silk and silkworm, with the relatively new technique CRISPER/Cas9 (clustered regularly interspaced short palindromic repeats), a gene-editing tool to simplify mutagenesis (Zhang et al. 2019). The tremendous advantage of using this host lays in its naturally existing spinning apparatus. Additionally, silkworms can be farmed and produce large quantities of silk (Vepari and Kaplan 2007). The synthesised fibres were a mixture of spider silk and silkworm silk, resulting in spider silk-*like* fibre. These results seem promising, as the qualitative properties came close to the ones of natural spider silk.

Although it is still a considerable way to go, the production of pure synthetic spider silk seems within reach soon. This renders spider silk an equally fascinating and hot topic, where mysteries are waiting to be solved.

1.3. Dragline silk

The dragline silk is the silk type which has gathered the most attention over the past years. This silk is used by the spider to build web frames, web radii and as its lifeline (Denny 1976). It is produced by the major ampullate gland and is considered one of the toughest silk forms, in combination of being elastic. This gland produces up to 8 types of spidroins (Garb et al. 2019), all named for their generating gland and numbered from 1 to 8 (major ampullate spidroin 1, MaSp1, MaSp2, etc.) with a length between 200 and 350 kDa (Chinali et al. 2010). The main difference between MaSp1 and MaSp2, the two major components of dragline silk for orb-weavers, lays in their amino acid content of proline and the repetitive motives in the middle domain (Hu et al. 2006). The motive ratio plays a considerable role in the resulting fibre properties (Vollrath and Knight 1999; Tokareva et al. 2013). MaSp1, for example, is characterised by the motifs poly-alanine A_n, glycine-alanine GA, and GGX(X)L (L leucine), Y (tyrosine), or Q (glutamine), while MaSp2 by the motifs A_n and GPG (P proline)(Elices et al. 2009). MaSp3 does not have the typical poly-alanine motifs of MaSp1/MaSp2 and lacks the GPG motifs typical of MaSp2 (Collin et al. 2018). In contrast to MaSp1/2, the MaSp4 contains more proline and lacks alanine (reported for *Caerostris darwini*, by Garb et al. 2019). Today, new motifs are still being discovered, and further MaSp proteins are added to the list. Additionally, it seems that in analysed silk threads, more than 100 other proteins are present besides the MaSp's, among which chaperone and chaperonin proteins were found (Huang et al. 2017).

A closer look at the spidroins revealed that they are composed by a large repetitive middle region, which is flanked by highly conserved N- and C-terminal domains. The repetitive segments can encompass up to hundreds of repeats (Schacht et al. 2015; Rising and Johansson 2015). Under storage conditions in the gland, the

spidroins are mostly unstructured and undergo a conformational change to mostly β-sheets at the end of fibre formation. Both non-repetitive terminal domains comprise about 10 % of the whole sequence and play crucial roles in the association of the spidroin proteins. NTDs are composed of generally 130 amino acid residues, while the CTD is slightly smaller, with approximately 100 amino acid residues (Hedhammar et al. 2008). Spidroins form homo-dimers, linked covalently by their CTD through a disulphide bridge. The NTD, on the other hand, is monomeric under storage conditions in the spinning gland (Fig. 1. 2, 1. 3).

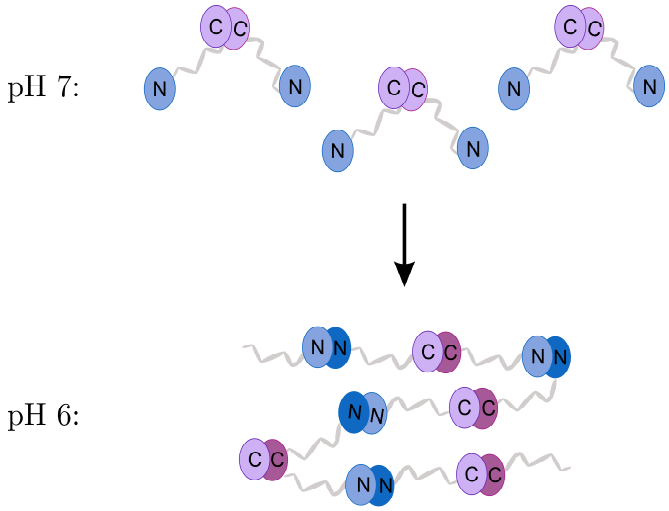


Figure 1. 2: Polymerisation scheme of the spider silk. Spidroins in their homo-dimeric form as they appear under storage conditions at pH 7, transformed under pH 6 conditions trough NTD-self-association, starting the polymerisation process of the fibre elongation in the spinning duct.

Both terminal domains are globularly folded, composed out of a 5 α-helix bundle. Their tertiary structures differ, demonstrating their different functional roles in the silk. The NTD forms an anti-parallel dimer during fibre association, whereas the CTD forms a parallel oriented dimer, each affecting the geometrical fibre formation. The NTD, therefore, plays a crucial role in the association and stabilisation of spidroins during storage (Askarieh et al. 2010). Only when undergoing the spinning duct, the NTD dimerises, starting the polymerisation reaction to form an elongated solid fibre. Changes in solution pH trigger the reaction, along with changes in salt concentration and mechanical stimuli, provided by the narrowing duct (Gaines et al. 2010; Hagn et al. 2011). The association of NTDs is an ultrafast process, including site-specific protonation

events, forming a macromolecular dipole, and conformational change (Jaudzems et al. 2012; Schwarze et al. 2013; Kronqvist et al. 2014; Otikovs et al. 2015). This combination of qualities makes the NTD a highly interesting domain. In turn, the role of the CTD is to stabilise the spidroins and keep their solubility during storage. Furthermore, CTDs pre-orient the spidroin, making them essential for the polymerisation. If not present, engineered spidroins showed less toughness in their resulting fibres (Hagn et al. 2010).

Interestingly, it is still debated today whether the highly concentrated, solute spidroins are stored in the spinning gland in either a micelle or in a liquid crystal form. The most alluring model concerning micelles, suggests that they elongate and fuse in response to increased shear forces and lowered pH (Eisoldt et al. 2011).

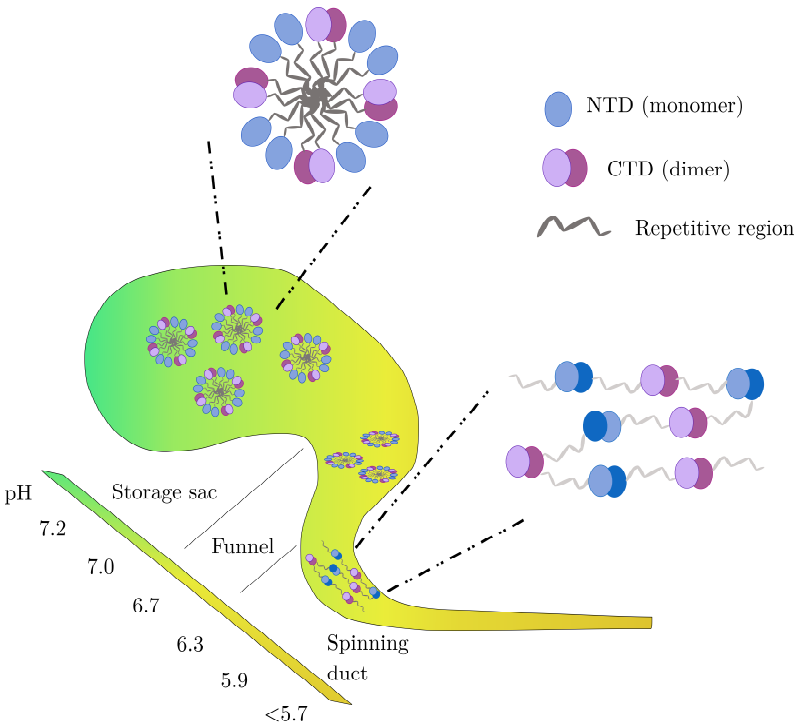


Figure 1. 3: Spidroin configuration as described for the micellar structure model. The hydrophilic N- and C-terminal domains form the outer layer of the micelles, while the hydrophobic repetitive regions are isolated in the middle, during storage conditions. While undergoing the funnel, the micelles get elongated and finally formed to a solid fibre while undergoing the spinning duct with decreasing pH value. At the left bottom of the figure, the pH value of each corresponding segment is indicated. (adapted from Kronqvist et al. 2017)

The other model suggests that the fibroins form anisotropic structures that function as a lyotropic liquid crystal (Vollrath and Knight 2001). Both models are considered to have shared parts in the spinning process (Rising and Johansson 2015). At the very end of the spinning gland, the assembled fibre is dehydrated and repetitive A_n motives align, leading to the formation of β-nanocrystals and organise to antiparallel β-sheet structures (Madurga et al. 2015). Glycine-rich segments form α-helical structures. Both are oriented in line with the thread (Vollrath and Knight 2001). Ultimately, they aggregate into a stable, insoluble spider silk thread (Koeppel and Holland 2017).

1.4. Amino acid composition of the NTD

A closer look at the composition of the N-terminal domain reveals that the most frequent amino acids are alanine, followed by serine, making up ~ 30 % of the sequence. Surprisingly, there are only a few charged side chains, although the domain is highly water-soluble. Furthermore, there is a prevalence of methionine and phenylalanine residues (Phe) in the hydrophobic core. This accumulation is remarkable since the amino acids generally have a rather rare prevalence in proteins. Methionine (Met) is integrated into a protein, by getting carried on a tRNA, which will bind during translation by a ribosome on the codon AUG from the messenger RNA (mRNA) (for more details concerning transcription and translation see section 2.1.1.). This codon functions as a start-codon for all protein synthesis in eukaryotes and with a modified Met (fMet) in prokaryotes (Kozak 1999). Methionine then gets connected to the next amino acid through a peptide bond and results in protein elongation. However, subsequently, the methionine often gets cut of the sequence or is modified, for example through acetylation of the amino group. In general, all amino acids which are coded only by one triplet codon are scarce in proteins. This is true for Met with an average prevalence of 2.5 % in proteins, but also tryptophan with 1.1 %, considered for all eukaryotes and prokaryotes (Jordan et al. 2005). All other amino acids have multiple codons, which shows that the genetic code is redundant (Sadava et al. 2019). There might be a critical evolutionary and functional reason for the relatively low prevalence of methionine residues in proteins. However, while not being produced by animals

or humans, Met is present in most living organism and is considered as an essential amino acid, which has to be assimilated with food.

To date, the active role of Met in matured proteins is still not completely understood. Up till now, the primary function was thought to be limited as antioxidant and to a regulatory function of cellular metabolism (Levine et al. 2000). To investigate and resolve the function of the methionine rich NTD is one of the principal goals of this work. A common approach to understand the role of a particular amino acid in a protein is by mutation to other amino acids. By comparing the properties of wild-type and mutant protein, one can deduce the function of the replaced amino acid. The Amino acid residues similar to methionine are leucine and isoleucine, both natural amino acids, and norleucine, an isomer of leucine but being an unnatural amino acid. Norleucine is nearly isosteric with methionine, except for the sulphur atom. The advantage of introducing norleucine is that the synthesis is straight forward, as it does not rely on extensive mutagenesis work upfront. Upon reading the mRNA, another tRNA is inserted instead of the methionine encoding triplet. Hence, protein function and structure can be studied faster, when introducing norleucine instead of methionine.

As Met is prone to oxidation, past studies have tried to optimise proteins to prevent this. Oxidised Met has been related to age-associated diseases, including Alzheimer's disease, Creutzfield-Jacob and von Willebrand disease (Valley et al. 2012). Different attempts of mutagenesis have been made to protect proteins from oxidation (for more detailed information, see section 2.1.2.). Gilles and colleges substituted methionine by norleucine back in 1988, where they transformed an adenylate kinase, which showed structural and catalytic properties similar to the wild-type protein and exhibited higher stability. In 2003, Cirino et al. replaced all 13 methionines in cytochrome P450, a fatty acid hydroxylase, by norleucine. Their results showed no increase in protein stability and a significantly reduced thermostability, but a two-fold increased peroxygenase activity (Cirino et al. 2003). With a methionine-to-norleucine substitution in a peptide, Radchenko demonstrated the convertibility of Met to norleucine, without changing its functionality. However, the protein showed a change in its interaction with

membranes, having become more rigid (Radchenko et al. 2016). Despite the advancement, the biochemical role of the Met residue remained unclear (Valley et al. 2012).

The critical difference between the proteins mentioned above and the spider silk NTD is that most of the methionine residues are present in the NTD's core, thus protected from oxidation. As norleucine does not occur naturally, the amino acid leucine becomes a suitable substitute. Nevertheless, this constitutes an imperfect change (Gellman 1991), since from a steric point of view, both amino acids are very similar, but still slightly different (Fig. 1. 4). Gellman explained that the side chain of methionine is more flexible, in comparison to leucines or isoleucines. The structural plasticity of Met would then lie in the unusual properties of the presence of the sulphur atom and the mediated torsion of the functional group. Further, he writes that due to this particular flexibility, methionine residues mould to nonpolar binding partners and that the sulphur atom is responsible for making methionine's side chain sticky, when under dispersion forces (Gellman 1991).

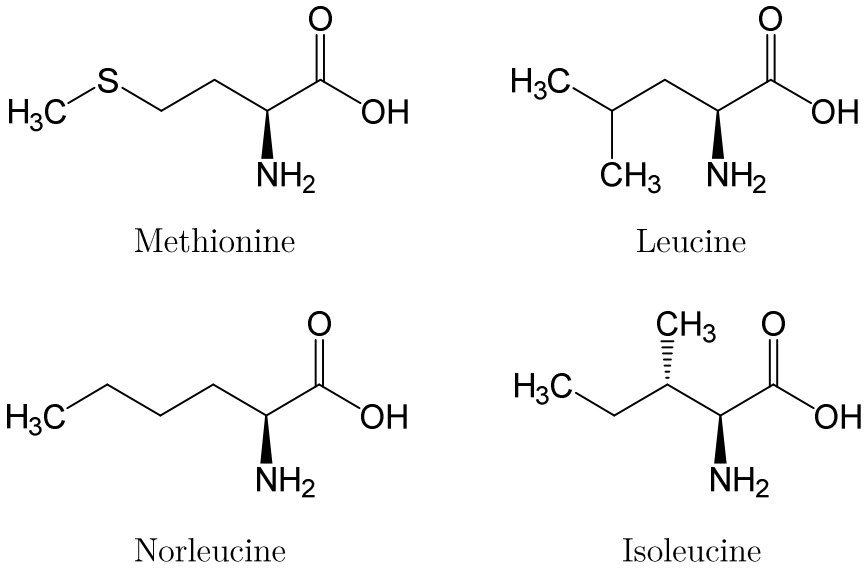


Figure 1. 4: Structural formula of the amino acids methionine, leucine, isoleucine and norleucine.

Met interacts with its surrounding amino acid residues, especially with other Met and Phe. Prediction models calculated strong Met-Met interaction, followed in strength by Met-Phe interactions. Those interactions seem to be even stronger as

the known aromatic Phe-Phe interactions (Gómez-Tamayo et al. 2016). The interaction energy between a sulphur atom and aromatic ring of Met-Phe residues is similar to a single salt bridge but over a longer distance. Interestingly, approximately one-third of all known proteins comprises an energetically stabilising Met-aromatic motif (Valley et al. 2012).

1.5. Phylogeny

Spiders have been evolving for over 400 million years. The sequence has slowly matured, leaving a highly specialised biopolymer with exceptional properties. The silks' protein sequences show strong homologies through species and glands, especially for the terminal domains, accentuating their essential functional roles in silk synthesis. Interestingly, the sequence of the CTD is slightly less conserved than the NTD, thus suggesting a more conserved function for the NTD (Garb et al. 2010).

The DNA of many spider silks have been collected for many years and can be found in genetic sequence databases such as GenBank® from the national institutes of health (NIH). Garb and colleges have united information for the NTD of spidroins across divergent spider lineages in 2010. Figure 1. 5 illustrates a systematic phylogeny of spider species, of which several terminal domains were analysed in this work.

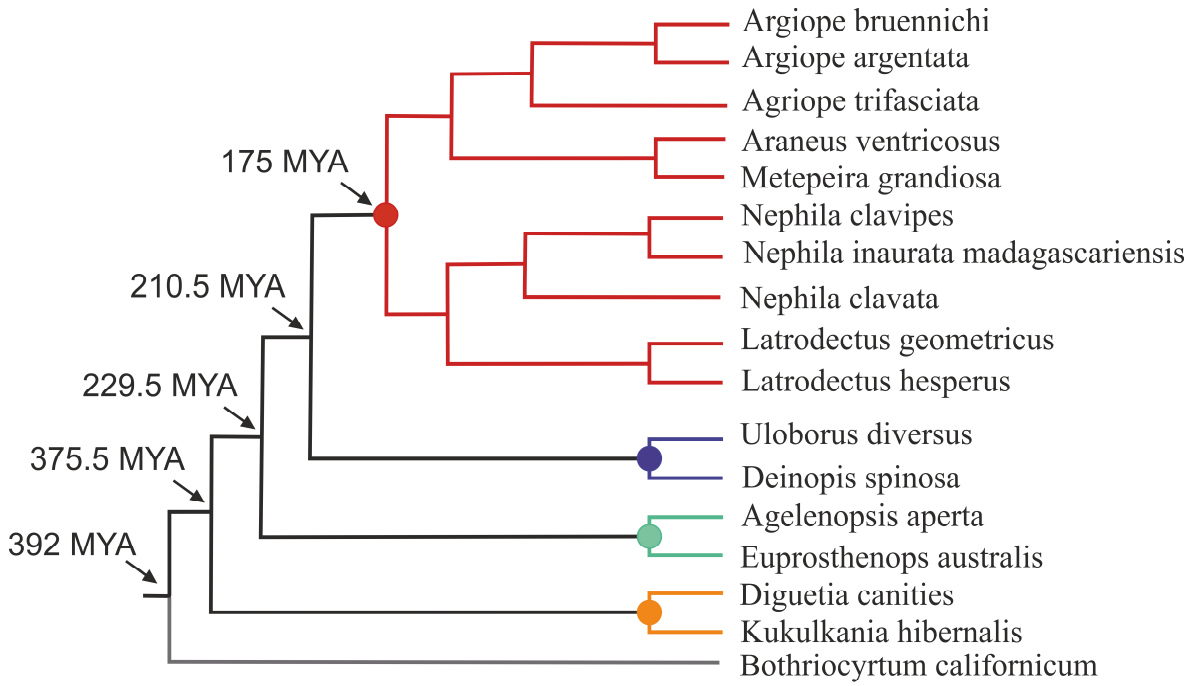


Figure 1. 5: Phylogenetic tree of spiders classified by their NTD sequences. Branch lengths are not proportional to time. Estimated divergence dates indicated with an arrow. (adapted from Garb et al. 2010)

All listed spiders are part of the arachnid class. The phylogenetic tree is categorised by the last common ancestor of the specific spider groups concerning the DNA encoding their NTDs, albeit from diver's glands.

The study showed that structural motifs associated with superior tensile strength arose earlier in history, while a motif conferring greater extensibility evolved convergently in two distantly related paralogs (Garb et al. 2010). As a result, spider's phylogenetic trees are reconstructed regularly, as the genetic data often provides more details, as physiological information can.

1.6. Scope of this work

As mentioned previously, N-terminal domains (NTDs) are highly conserved throughout spider species and glands. The domain is essential for the association of silk, by a unique mechanism, which seems analogue for all spiders.

However, one major unanswered question, which this thesis will address, is: does the conservation of the NTD reflect on its characteristics and its unusual mechanism of folding? This question is sustained by the interesting observation that even with similar sequences, the web geometries of different spider species varies considerably. In order to answer this question, I will present the quantitative comparison of four isolated homologous NTDs regarding their energetics, kinetics of folding and self-association. To put our findings for the NTD of the *E. australis* into a general context, a similar analysis of the C-terminal domains (CTD) of the same spider will be presented. Therefore, I will show how we elucidated the complex folding of the CTD and determined that the protein belongs to a class called the knotted proteins.

To correlate the results found for the NTD and CTD with a truncated form of the full-length spidroin, protein material had to be synthesised first. As large proteins tend to be "hard candidates", and especially spidroins have shown to be very tough to synthesise, the presented results are preliminary. Here, I show how spidroin material was synthesised and what additions could be made to facilitate protein expression in the future.

Following the elucidation of the isolated domain's folding mechanism, the main task of this work was to understand the role and function of the rare and particular amino acid residue methionine, in the context of spider silk. It appears that the NTD contains a high number of this special amino acid. We analysed this phenomenon by exchanging all core methionine residues with leucine through site-directed mutagenesis. I will present how we discovered that they are responsible for the dynamical range of the protein. Furthermore, I will also relate these findings to the role of the leucine residues, which have shown to be involved in the stability and rigidity of the NTD.

Based on these findings, we contemplated the potential consequences for protein engineering, in terms of creating more functional proteins or improving stability without changing the protein's structure. My results suggest that methionine residues enhance the tight binding in spider silk proteins. Moreover, we speculated if our findings could be generalised to all proteins. For this purpose, we introduced new methionine residues to a methionine-depleted NTD core and exchanged leucine with methionine residues in an Hsp90 full-length protein.

Finally, a conclusive outlook on potential follow up endeavours regarding the presented projects will be proposed. Although it is believed that one cannot correlate sequence with protein function, my preliminary results could hint towards another direction. I will give a perspective on a potentially novel classification approach, linking the phylogenetic ancestry of spider species with the share of methionine in their respective NTDs.

2.Theoretical Background

This dissertation focuses mainly on spider silk, which is crafted by complex macromolecules. In the following, I will introduce the required fundamental knowledge concerning the biology of proteins and the basis of the applied spectroscopic methods to investigate proteins properties.

2.1. Biology of proteins

2.1.1. Proteins

The term *protein* was first introduced in 1839 by Gerardus Johannes Mulder and Jacob Berzelius when they studied the constituents of eggs, fibrin and albumin. At that time, it was mistakenly thought that proteins were composed of a single type of molecule (Mulder et al. 1838; Hartley 1951). This stayed unchallenged until 1902 when Emil Fischer and Franz Hofmeister proposed that proteins were formed by amino acids, connected by peptide bonds (Fischer 1906; Hofmeister 1902).

Amino acids are organic compounds containing a hydrogen (-H), amine ($-NH_2$) and a carboxyl (-COOH) group, linked to a carbon atom. In addition, the amino acid has a characteristic side chain, defining each of the specific 21 amino acids and their properties in eukaryotes. They are the building blocks of proteins and are encoded by DNA (deoxyribonucleic acid). The DNA is composed out of four nucleotides, which consist of a phosphate group, deoxyribose sugar and one of the four bases adenine (A), cytosine (C), guanine (G) and thymine (T). The base-sequence then determines the information content of the DNA. Phosphodiester bonds link the nucleotides and have a complementary sequence forming a double strand, twisted to a helix. G-C and T-A base pairs link both strands with hydrogen bonds. DNA is considered very stable and is identical in each cell of an organism (Clark and Pazdernik 2013).

In order to synthesise proteins in a cell, first, a copy of the DNA is produced by reading and transcribing it into ribonucleic acid (RNA) with the help of an RNA polymerase. The sequence is complementary to the DNA but remains single-stranded. The resulting so-called messenger RNA (mRNA) is also composed out of 4 nucleotides, with the difference that thymine is replaced by uracil. The mRNA is then read and translated by the ribosomal machinery in the cytoplasm, thus generating proteins. For this, they follow the triplet codon sequence, where each triplet corresponds to a specific amino acid (Clark and Pazdernik 2013). The genetic code is termed canonical, as several base triplets can code for the same amino acid. In contrast to DNA, the RNA is less stable and prone to modulation but is only needed for relatively short informational transfer (Clark and Pazdernik 2013).

The transfer RNA (tRNA) is responsible for delivering the right amino acid residue. It gets coded complementary to the mRNA by the triplet codon and carries the corresponding amino acid residue. tRNAs arise from mRNA and are short nucleotide sequences. After maturation, they get covalently linked with an amino acid residue. For every protein, the initiator $tRNA_i$ is coded by the mRNA triplet AUG and is loaded by methionine (Met). A distinction is made between this initiator $tRNA_i^{met}$ and the $tRNA^{met}$ coding for methionine in the protein sequence. The ribosome does not recognise the latter as the first operating tRNA. Together with the ribosomal machinery, the mRNA gets read and associates the amino acids by peptide bonds, as an elongated polypeptide sequence arises (Clark and Pazdernik 2013).

Not all amino acids are equally frequent in proteins. Some of them are more abundant than others and show different characteristics, as amino acids can be hydrophobic or hydrophilic, charged, uncharged, polar or nonpolar, acidic or basic. These features are partially responsible for the resulting protein structure and function. Hydrophobic amino acids have carbon-rich side chains, which do not interact well with water. Hydrophilic, or polar amino acids, promote the solubility by forming hydrogen bonds with water molecules. Charged amino acids interact with oppositely charged amino acids, or with other molecules (Sadava et al. 2019; Jordan et al. 2005).

2.1.2. Mutagenesis

Mutating a protein and comparing its functionality to the wild type is an excellent tool in modern protein structure-function analysis. Thirty years ago, site-directed mutagenesis of proteins was a particularly tough challenge and a lengthy process. Today, transformed DNA, coding for mutant proteins, can be produced reasonably cheap within a few days.

At first, generating mutations was essentially non-site-specific or somewhat random, induced by the use of radiation or chemical mutagens. Only in 1978, Hutchison and colleagues used synthesised oligonucleotides as primer for the first time (Hutchison et al. 1978). In combination with DNA polymerase and the invention of the polymerase chain reaction (PCR), site-specific mutations of the DNA could be generated. For this work, Michael Smith and Kary B. Mullis were awarded the Nobel Prize in Chemistry in 1993.

Since then, several methods for the mutagenesis of DNA have been developed, among which overlap extension PCR, mega primer PCR and QuikChange™ site-directed mutagenesis are popular methods (Xia et al. 2015). For the QuikChange mutagenesis, a high-fidelity DNA polymerase is required (e.g. Q5 or Phusion DNA polymerase). This polymerase amplifies the whole plasmid, using complementary primers carrying the desired mutation. After multiple replications of the plasmid, the original DNA template is eliminated in a final step. The parental DNA is digested using DpnI enzyme, which cuts only methylated DNA (Zeng et al. 2018). For DNA amplification, the nicked DNA is transformed into competent *E. coli* cells (e.g. XL1 Blue) and cultivated. Subsequently, the DNA can be extracted with a DNA purification kit. The technique of plasmid DNA extraction by alkaline lysis was established by Birnboim and Doly in 1979 and is still widely used for the isolation of plasmid DNA from genomic DNA.

Until today, mutagenesis has had a fundamental influence on molecular biology, genetics and biochemistry but also generally in protein science. The possibility of mutating DNA, precisely and quickly, opened up a new field of mechanistic und functional protein studies (Liu and Naismith 2008). Since 2013, the development of the CRISPR/Cas9 technology has allowed genome editing *in vivo* (Zhang

2019). Based on a prokaryotic, viral defence system, the introduction of mutations into the genome of an organism can lead to the deactivation or modification of genes (Xu et al. 2015; Sadava et al. 2019).

Proteins are essential for the function and survival of every cell. If not working correctly, as they are, for example, mistranscribed or misfolded, the function of this protein gets reduced or lost. Therefore, it is essential to study proteins, their functional role, and how they are integrated into the system of an organism.

2.1.3. Protein folding

In order for a protein to be functional, it has to adopt a complex structure, starting from the linear sequence of amino acids joined by peptide bonds, called the primary structure. From this, the protein chain will form the secondary structure, classified into two motifs, α-helices and β-sheets, or remain unstructured. The α-helix is a right-handed coil, stabilised by hydrogen bonds between the amin and carboxyl groups of surrounding amino acids, while the β-sheets are formed when hydrogen bonds stabilise two or more adjacent strands of amino acids. This structural transition already begins at the end of the protein translation from the synthesising ribosome. After exiting the ribosome, chaperone molecules such as Hsp90 (heat shock protein), assist the new protein into their correct three-dimensional tertiary shape and quaternary structure. The quaternary structure is formed by two or more polypeptide chains, which organise to one functional molecule with several subunits. Most of the large proteins depend on chaperone systems, whereas small proteins can fold without them. Many proteins are globularly shaped with hydrophobic side chains sheltered in the core of the protein. The final structure influences the functionality of the protein as well as the ability of molecular interaction (Clark and Pazdernik 2013).

Most proteins fold into a single conformational state, while theoretically multiple conformations exist. In 1960, Anfinsen and colleagues formulated the hypothesis that proteins can only fold into one particular structure, defined as the state, where the global minimum of free energy is reached. Levinthal later stated, that a protein could not “find” its fold randomly, i.e., by exploring the accessible

conformational space, as this would be absurdly time-consuming (Levinthal 1969). This has become to be known as the Levinthal paradox (Zwanzig et al. 1992).

The folding of a protein can be described by a two-step process, where the protein either populates an unfolded or a folded state, which is mostly true for small proteins (Gelman and Gruebele 2014). However, for some proteins, especially bigger ones, the folding process is more complex. Here, the folding mechanism includes faster steps like the formation of the hydrophobic core and formation of secondary structures in microseconds, followed by or with simultaneous slower processes on the second timescale. Thermodynamics is the driving force of protein folding, where the free energy or Gibbs energy, consists of enthalpy and entropy. Furthermore, many proteins form folding intermediates, which restrict the conformational space at an early stage, thus controlling the entire folding process.

The correct folding of proteins is of utmost importance regarding protein function. Misfolded or aggregated proteins are connected to numerous diseases. In general, the misfolding of proteins can have multitudes of causes, ranging from mutations in the DNA sequence coding for the protein of interest to external factors such as increased temperature. Ghosh and Dil found that raising the temperature of a prokaryotic cell by only 4 °C, caused 20 % of the proteins to either unfold or lose their functionality (Ghosh and Dill 2010). Additionally, if chaperon proteins were affected, they would further impact other proteins, leading to significant cell-stress, ultimately resulting in cell death. Hence, cells devote high amounts of energy in homeostasis and regulation mechanisms in order to avoid these chain reactions.

Predicting the correct fold of a protein based on amino acid sequences has been a challenge for decades. In recent years, more and more algorithms have emerged that can generate accurate protein structures based on their amino acid sequence. The underlying models include a multitude of contributing factors, for example, electrostatics, hydrogen bonding, solvation capacity and Lennard–Jones interactions (Gaines et al. 2017).

An example of a successful software is FoldIt, an online game, based on the participation of users who create new protein structures. The goal of the game is to generate a correctly folded protein with no existing homolog model (Cooper et

al. 2010). 56 of the 104 de novo user-created proteins have been successfully recombinantly synthesised in *E. coli* in 2019 and were tested for their fundamental physical properties. The creators of FoldIt draw new conclusions from these measurements, making advances in general folding understanding. They could deduce, for example, that the created proteins did not fold only by the principle of absolute low-energy design, assumed till then to be an optimisation criterion for protein design (Koepnick et al. 2019). It seems likely that software like FoldIt could be used to generate custom build proteins with desired functions in the future.

There are several other software approaches to date, predicting structures. Most of these scan their protein database for homolog matches and build homology models out of existing proteins structures. Unfortunately, this alone does not suffice to predict an unknown structure accurately. Nevertheless, those programs, like the well-known Swiss-Modelling, generally predict very well how the protein might look.

2.1.4. Protein-protein interaction

Insights in protein structures generated by methods like X-ray crystallography and structural models based on solved primary X-ray structures are essential. However, those methods have an intrinsic disadvantage, namely that they show a frozen image of the protein structure, revealing only secondary information on protein flexibility, if at all. In order to understand the biological function of those proteins and to put them into a more complex biological context, additional kinetic and dynamic data is required.

The association or dissociation of proteins, like the catalysis by enzymes or the assembly of spidroins in the spinning duct, are very fast. Methods that aim to analyse the interactions between two or more molecules, therefore, require a high temporal resolution. For this, specialised techniques like stopped-flow or nuclear magnetic resonance (NMR) spectroscopy have been developed.

Every association begins by diffusion-driven molecules, while the diffusion is generally a limiting factor. Furthermore, the proteins need particular orientation

ahead of association, which is influenced strongly by electrostatic interactions. These circumstances influence the rate constant of association strongly. The spectrum of protein-protein association rate constants ranges from 10^3 to 10^9 $M^{-1}s^{-1}$, while in comparison, the basal rate constant lays around 10^5 $M^{-1}s^{-1}$ (Schreiber et al. 2009). Rate constants faster than the basal rate constant are diffusion-limited, and rates below are slowed by conformational change. Favourable electrostatic interactions often enhance fast rate constants. The impact of the electrostatic attraction on the association can be quantified by increasing the ionic strength (*Is*) of the analysed solution. With additional salt, electrostatic charges are shielded, and the association, therefore, weakened (Alsallaq and Zhou 2008).

The dissociation rate constant, on the other hand, is only slightly affected by the ionic strength of the solution, is protein concentration-independent and generally much smaller as the association. Dissociation constants can also be measured by the stopped-flow technique, given a suited reporter system. Both rate constants of association and dissociation define the equilibrium dissociation constant K_D and therefore the association strength of the analysed molecules (see section 3.4.2., equation 3.13). The smaller the resulting K_D value, the stronger the affinity is between the two molecules. An example of a strong K_D value is the bond between an antibody and antigen, which has a K_D of only few nM.

2.1.5. Protein dynamics

As mentioned above, in order to understand the function of a protein, detailed knowledge about its intrinsic dynamics is required. For instance, the energy landscape theory states, that proteins do not necessarily adopt one, static conformation, hence a multitude of different conformational states are possible (Henzler-Wildman and Kern 2007). These conformational changes occur over diverse length and time scales. Therefore, it is ambitious to describe all atoms in a protein in detail by time and relative dimension in space. However, methods like molecular dynamics (MD) Simulations, fluorescence correlation spectroscopy (FCS) or NMR can reveal different aspects of protein structure-function

correlation. In addition, accessible local and global conformational changes also provide information on the function of the protein.

In particular, local conformational changes can be observed by PET-FCS, a technique monitoring the changes in distance between two positions of a protein over time. Here, the dynamic within a protein can be unravelled by a cleverly placed reporter system, comprising a fluorophore and a quencher (for more information see section 2.2.3). Therefore, we can analyse the flexibility of domains or secondary structures of a protein.

The characterisation of conformational states and their delimiting energy barriers requires a multi-dimensional energy landscape, to be fully described (Henzler-Wildman and Kern 2007). Reaction rates can help to determine conformational transitions and support the bigger picture. For fast reactions, methods like stopped-flow spectroscopy, temperature jump approach or flash photolysis can be applied (detailed information for stopped-flow can be found in section 3.3.5.). More about the underlying spectroscopy is introduced in the next section.

2.2. Optical spectroscopy

As most proteins are smaller than the wavelength of light, they cannot be studied easily by examining them with a microscope. Fluorescence spectroscopy has proven to be a powerful and versatile toolbox for the investigation of protein structures, function and interactions as well as for folding dynamics. For the study of the latter, particularly fast spectroscopic methods are required, as conformational dynamics take place on a timescale from seconds down to nanoseconds (Doose and Sauer 2007). There exists a wide variety of applications using fluorescently labelled proteins or molecules with intrinsic fluorescence (like tryptophan) for spectroscopic methods. Those techniques range from classic fluorescence spectroscopy to more sophisticated methods such as photoinduced electron transfer fluorescence correlation spectroscopy (PET-FCS), enabling the characterisation and analysis of molecules. The stopped-flow technique is well suited to examine fast folding kinetics in the millisecond time range (Gelman and

Gruebele 2014). For three dimensional analysis of protein structures, methods like X-ray crystallography or NMR spectroscopy are widely used.

2.2.1. Light: absorption and fluorescence

Light can be described as electromagnetic radiation, mainly defined by its wavelength λ, which is inverse proportional to its energy content. The visible, or optical range for humans spans from approximately 400 to 700 nm. Radiation with longer wavelengths has lower energy content and is called infrared light, microwaves and radio waves. Radiation with shorter wavelengths is classified as ultraviolet light (UV), X-rays and gamma rays, becoming more energetic with lower wavelengths. Commonly, the combined ultraviolet and visible spectral range is used for most spectroscopic experiments (λ= 200-700 nm).

Radiation of different wavelengths, i.e. energies, interacts differently with matter. If a molecule exhibits an electronic structure fitting a photon's energy, it can absorb the photon. Usually, a particular type of molecule will have a range of allowed photon energies, leading to an extended energy band of diverging absorption probability, the absorption spectrum (Figure 2. 1, blue line). After the molecule takes up a photon, thereby elevating an electron to an excited energy state, there are several ways the molecule can relax back to its energetic ground state. Part of the absorbed energy is internally converted to molecular vibration and rotation, i.e. heat. If the remainder of the energy is shed by the emission of a photon, the relaxation is generally called luminescence. Dependent from which exited state the photon was emitted, we distinguish between the more energetic fluorescence and the less likely phosphorescence. Because of the difference in absorbed and emitted photon energy, the emission spectrum of a fluorophore has an energy-depleted peak wavelength and is red-shifted, which is called *the stokes shift*. Together with both spectra, the spectroscopic fingerprint of a molecule is constituted, which can be used to characterise changes in the molecular micro-environment and interactions. The molecular structure of the typical fluorescent dye AttoOxa11 is depicted in Fig. 2. 1 together with its characteristic absorption (blue) and fluorescence (red) spectra.

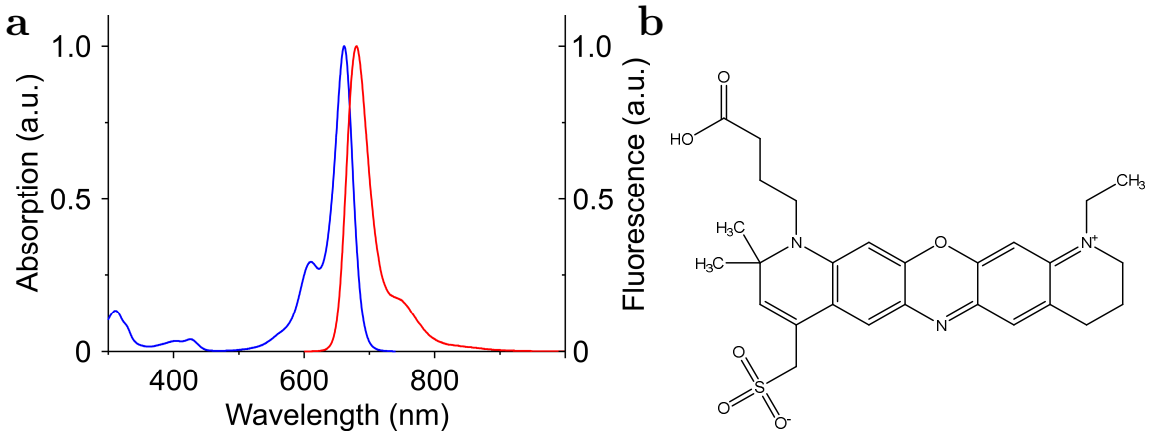


Figure 2. 1: Absorption and emission spectra of AttoOxa11. (**a**) Absorption spectrum in blue of AttoOxa11, with red-shifted emission spectrum in red. (**b**) Molecular structure of AttoOxa11 dye.

The absorption of an ensemble of molecules depends on the molar concentration c. The relation between those parameters is described by the law of Beer-Lambert:

$$OD = \log \frac{I_0}{I} = \varepsilon \cdot c \cdot d \tag{2.1}$$

where I_0 is the incident light intensity (at a particular wavelength), I is the intensity after transmission through a sample in a cuvette of a fixed length d and ε the molar extinction coefficient of the measured sample. The optical density (OD) then equals the quantity of $\log(I_0/I)$ and is a common metric for the absorption.

The aromatic amino acids phenylalanine, tyrosine and tryptophan, present in proteins, are naturally occurring fluorophores. Tryptophan, for example, has an absorption peak at a wavelength of $\lambda_{abs} = 280$ nm, which is commonly used to determine protein concentration, according to equation 2.1. In order to measure the pure emission of tryptophan, one uses the photo-selective excitation at wavelengths above 295 nm to neither excite tyrosine ($\lambda_{abs} = 275$ nm) nor phenylalanine ($\lambda_{abs} = 260$ nm) simultaneously (Eftink 1991).

Interestingly, the emission peak of tryptophan is solvatochromatic, meaning the absorption and emission spectrum are shifted with the polarity of the solvent. Therefore, the fluorescence of the amino acid residue of tryptophan is a good indicator of the local molecular environment. This inherent quality is employed

to study protein conformation. Tryptophan (Trp) is a hydrophobic amino acid, which is often found in the core of proteins. Since there are only few Trp in proteins in general (1.1 % Trp residues, Jordan et al. 2005), it is a particularly useful sensor. If the conformational state of the protein changes, the microenvironment of the individual Trp might also change. In denaturation experiments, for example, the buried tryptophan in the hydrophobic core gets solvent-exposed upon protein unfolding. The emission spectrum of tryptophan will then be red-shifted with increased solvent contact, which can be measured directly.

Proteins lacking tryptophan can be labelled with another fluorophore, functioning as a fluorescent reporter. Most often, these fluorophores are chemical compounds designed to have a bright and stable fluorescence. Most of them have several aromatic rings emitting light upon excitation (see Fig. 2. 1, with exemplary fluorophore AttoOxa11).

There are different techniques to label a protein with a fluorophore. Most of them underlie a mechanism where fluorophores are attached to proteins through a specific functional group. Thiol-, amino- or carboxyl groups are often used to create covalent bonds. The functional group thiol occurs in the amino acid cysteine, which can be used to react with maleimide modified dyes specifically. N-hydroxysuccinimide (NHS) -ester modified dyes react with amino groups of proteins, predominantly found in lysine residues. A significant disadvantage of dye labelling is the size of the fluorophores, which are often bigger than tryptophan. This can significantly alter the protein in its conformation and interaction. In turn, organic dyes are brighter than aromatic amino acids. Thus, the detection of fluorescence is comfortable and more precise. Additionally, there exists a wide range of versatile organic dyes, spanning the entire wavelength band from 300 to 800 nm.

2.2.2. Fluorescence quenching

Fluorophores can be quenched, leading to a decrease in emission intensity, detectable by different fluorescence spectroscopy techniques. This happens, when a molecule (quencher) interacts with the excited states of a fluorophore such that

alternative, non-radiative, relaxation paths are created. Here, we differentiate between the static and the collisional quenching. Quenchers that come into contact with a fluorophore by diffusion are known as collision or dynamic quenchers. In turn, certain molecules can form complexes in the ground state and are therefore not diffusion dependent. Those complexes can exhibit an energy landscape, that prevents fluorescence and are called static quenchers. In most experimental systems, both, static and dynamic quenching, coincide with different impact (Eftink 1991).

By probing fluorescence changes under different conditions, local and dynamic insight about the protein of interest can be gained. Here, quenchers can be different types of agents, like oxygen, acrylamide, iodide and other aqueous solutions. For all, steric factors have to be considered, since the distance between quencher and fluorophore is decisive for the efficiency of fluorescence quenching (Eftink 1991).

Notably, tryptophan is used both as fluorophore and quencher in this work. In case Trp was used as fluorophore, its fluorescence was altered by changes in solvent exposure or by contact to other amino acids residues, like methionine (interaction between thiol group of the methionine and the NH group of tryptophan, Pal and Chakrabarti 2001). Modulating the solvent accessibility of Trp can be achieved by thermal or chemical denaturation of the protein of interest, using the Trp fluorescence as an unfolding sensor. Methods that make use of this tryptophan characteristic include classical fluorescence spectroscopy, FCS and stopped-flow fluorescence spectroscopy (for more details see section 3.3.).

2.2.3. Photoinduced electron transfer - PET

Different types of protein dynamics occur under various premises and at different time scales. There are multiple quenching mechanisms with distinct areas of spectroscopic application, regarding temporal, spatial and spectral resolution. Of these, the photoinduced electron transfer (PET) is used in many biochemical applications, like photosynthesis of plants or photovoltaic systems (solar cells). Here, the non-radiative relaxation of the excited fluorophore is facilitated by the transfer of the excited electron, or rather its energy, to a close-by interacting

molecule. For this, the excited fluorophore F* can either be reduced by an electron-donating quencher D (reductive electron transfer) or oxidised by an electron-accepting quencher A (oxidative electron transfer) (depicted in Fig. 2. 2). Whether a molecule is an electron donor or acceptor depends on the molecules ground state redox-potential. Since the intermediate states, i.e. radical ions, of this process exhibit lower energy differentials, they can relaxate non-radiatively by charge recombination to the ground state.

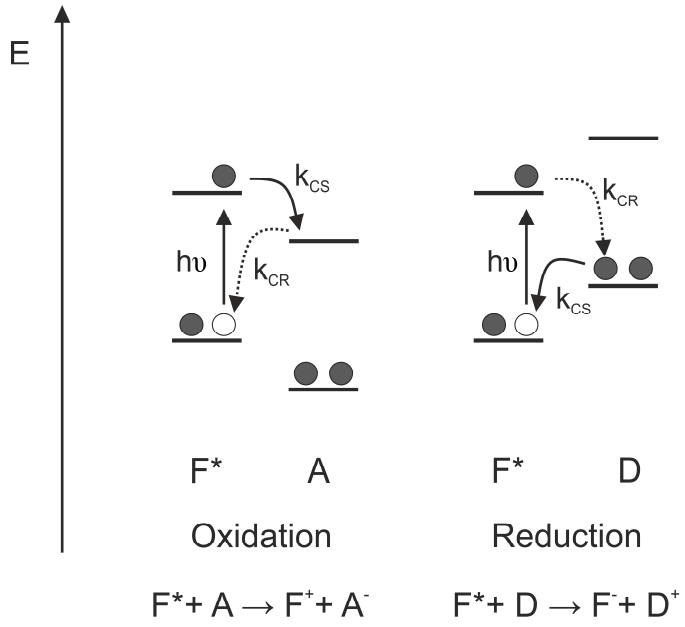


Figure 2. 2: PET reaction displayed in an energetic scheme (E indicates the relative energy). The electron transfer is shown between a fluorophore F*, exited by light (hν), and an electron acceptor A (oxidative electron transfer, left) or electron donor D (reductive electron transfer, right). Rate of charge separation k_{CS} and charge recombination k_{CR} are indicated.

Importantly, the prerequisite for an electron transfer is that both molecules, fluorophore and quencher, must be in close proximity. This is called *van der Waals* contact and is defined by an intermolecular distance of ≤ 1 nm. This distance is sufficiently near for the exchange of electrons between the electron clouds. The farther they are apart, the smaller the probability for electrons, and thus energy, to be transferred (Sauer and Neuweiler 2014; Lakowicz 2006).

The efficiency of the free energy change of charge separation ΔG^0_{CS} during the PET process, can be estimated by the Rehm-Weller equation (Rehm and Weller 1969):

$$\Delta G^0_{CS} = E_{Ox} - E_{Red} - E_{0,0} + \frac{-e^2}{\varepsilon \cdot d} \qquad (2.2)$$

Here, E_{Ox} is the oxidation potential of the donor, E_{Red} the reduction potential of the acceptor, and $E_{0,0}$ the energy difference between the ground state and excited state. The last term of the equation determines the coulombic interaction energy between two charged molecules, where e is the elementary charge, dependent on the distance d between donor and acceptor and on the solvent polarity, described by the dielectric constant ε of the surrounding.

An advantage of PET-based methods is that naturally occurring quenchers can be used. This prevents the time consuming and laborious dual labelling of the molecule of interest. Strong oxidants are, for example, the amino acid tryptophan or the nucleoside guanosine. Tryptophan is often used for PET experiments serving as electron donor interacting with fluorophores like oxazine dyes, e.g. Atto655 or AttoOxa11 (Marmé et al. 2003; Doose et al. 2009; Neuweiler et al. 2003). The PET-based quenching of an excited dye by Trp in combination with FCS can be applied to analyse dynamics of proteins (Neuweiler et al. 2010; Teufel et al. 2011; Lum et al. 2012). PET-FCS is the method of choice for the study of fast chain motions, taking place on the nano- to millisecond time scale (Sauer and Neuweiler 2014)(more about FCS in section 3.3.8.).

The main difference between ensemble measurements and single-molecule techniques like PET-FCS is the type of information extractable from the results. For bulk measurements, which include the conventional spectroscopic techniques, the result will always reflect an average of all present molecules. Hence, sub-state conformations cannot be distinguished. With the advance in fluorescence detection limits, new techniques have emerged in the past decades. Now, individual molecules can be spectroscopically probed under different conditions, with methods such as FCS and PET-FCS, single-molecule based scanning probe microscopy, single-molecule localisation microscopy or optical tweezer approaches. This opens up new possibilities for the study of mechanisms of protein

folding, conformational dynamics, kinetics of on/off-rates, association- and dissociation rate constants on the nanosecond time scale under equilibrium conditions (Kim et al. 2006; Neuweiler et al. 2005; Neuweiler et al. 2007).

3.Materials and Methods

3.1. Molecular biology

For recombinant protein synthesis of spider silk proteins, the synthetic genes of the NTD, CTD and shortened spidroins from the major ampullate gland (MaSp1 and MaSp2) were purchased from GeneArt® (Thermo Fisher Scientific, USA). Additionally, Hsp82 from *Saccharomyces cerevisiae* was chosen as a comparative protein model, described further below, and kindly provided by Andrea Schulze. In the following, this chaperone is referred to as Hsp90. For spider silk expression we selected the following genes: NTD genes from MaSp1 *Euprosthenops australis* (Ea), *Latrodectus hesperus* (Lh), *Latrodectus geometricus* (Lg) and *Nephila clavipes* (Nc); CTD gene from MaSp1 Ea. To maximise protein yield for the shortened spidroins NTD-rep4-CTD and NTD-rep17-CTD, the genes of MaSp1 from Ea and MaSp1/MaSp2 from Lh, respectively, were cloned with additional fusion proteins. We chose GroEL, Lipoyl and Hsp90-NTD as fusion proteins.

3.1.1. Preparation of chemically competent cells

Chemically competent *E. coli* cells (XL1-Blue and C41 (DE3) *E. coli*) were platted on 2x TY-agar plate (without antibiotic) and were incubated overnight at 37 °C. A bacterial pre-culture was then cultivated over-night in 100 ml 2x TY-medium without antibiotics. The 2x TY medium contained tryptone (16 g, Melford Laboratories, UK), yeast (10 g, Melford) and NaCl (5 g, Merck, DE), was solved in H_2O (1 l) and adjusted to pH 7.4 with NaOH (5 M, Merck) and was autoclaved (15 min, 121 °C). A second 2x TY-medium culture (also without antibiotics) was started with 8 ml of the over-night culture, and grown to OD_{600} 0.5 at 37 °C. The temperature was then lowered to 4 °C for 30 min and centrifuged at 4000 rpm for 10 min, 4 °C. The cells were resuspended in 100 ml precooled transformation buffer (TB). The TB was prepared with 3 g PIPES (Sigma-Aldrich, USA), 2.2 g $CaCl_2$*$2H_2O$ (Merck), 18.6 g KCl, 75 g glucose was solved in 980 ml H_2O and

adjusted to pH 6.7 with KOH (1 M, Merck). Afterwards, 10.9 g $MnCl_2*4H_2O$ (Sigma-Aldrich) was added and adjusted to 1 l, sterile filtered (single-use sterile filtration unit with 0.45 µm pore size, Acrodisc®, Pall, USA). The cells were then centrifuged again at 4 000 rpm for 10 min at 4 °C. A second resuspension step followed with 200 ml of precooled TB and incubation for 30 min on ice. The suspension was centrifuged a third time at 4 000 rpm for 10 min at 4 °C. Finally, the bacterial pellet was resuspended in 25 ml TB (4 °C) and added to 25 ml precooled 14 % DMSO (3.5 ml DMSO, 21.5 ml TB, chilled on ethanol - dry ice mixture). The suspension was aliquoted (200 µl) and quickly frozen in an ethanol-dry ice bath and stored at -80 °C (protocol adapted from Inoue et al. 1990).

3.1.2. Molecular cloning

Prior to molecular cloning, the synthetic genes were multiplied by transformation into XL1 blue *E. coli* cells, before being isolated by a plasmid isolation kit (High Pure Plasmid Isolation Kit, Roche, CH) following the manufacturers specifications. Heat-shock transformation was carried out by adding 1 µl plasmid DNA (70-180 ng/µl) to 40 µl *E. coli* cells (stored at -80 °C), incubated for 5 min on ice before 5 µl were plated onto 2x TY-agar petri dish. For TY-agar-plates, agar (15 g, Sigma-Aldrich) was added to 2x TY medium (1 l). Medium was sterilised by autoclaving (15 min, 121 °C) before being supplemented with 0.1 mg/ml ampicillin antibiotic (Melford) after the medium had cooled down (< 50 °C). With the ampicillin resistance carried on the plasmid DNA, only *E. coli* containing the desired plasmid, multiply.

The following restriction digestion was performed using BamHI and EcoRI enzymes (New England Biolabs®, USA), digesting the pRSET A vector and the vector containing the gene of interest for 3 h at 37 °C, following the manufacturer's specifications. Subsequently, a 2 % agarose gel electrophoresis was used to isolate DNA fragments (PowerEase 500, Invitrogen™, Thermo Fisher Scientific; at 120 V, 7 W, for 45 min). The gel contained 0.5x TBE-buffer (from 5x stock solution of 2.7 g Tris, Sigma-Aldrich; 1.38 g boric acid, Merck; 1 mM EDTA, Sigma-Aldrich; pH adjusted to 8) supplemented with 2 % agarose (peqGOLD universal agarose, peqlab, DE) and was stained with Safe View classic™ (G108; ABM, Applied Biological Materials®, CA). As the DNA is negatively charged, it

will migrate to the positively charged anode through the gel, distributing corresponding to their size. Afterwards, the restriction digest was added with DNA loading dye (x6, Thermo Fisher Scientific) and the gel extraction was performed according to the QIAquick kit manual (Qiagen, NL). Subsequently, the isolated plasmid was ligated to the vector, using T4 ligase (NEB®, USA) and incubated overnight at 16 °C. The ligated sample was then transformed into XL1 blue *E. coli* cells, as described above, by transforming 5 µl of ligation-sample and plating 10 µl onto 2x TY-agar plates (containing 0.1 mg/ml ampicillin). The plates were incubated overnight at 37 °C.

All genes were cloned into a modified pRSET A vector (2.9 kb, Invitrogen, Thermo Fisher Scientific) using conventional restriction digestion and ligation (vector maps in annexe C, Fig. 8. 1). The vector contained an ampicillin resistance, a T7 promotor and His6-Tag which was followed by a Thrombin recognition sequence for proteolytic cleavage of the tag.

3.1.3. Mutagenesis

Mutants of the above listed genes were generated using the QuikChange protocol from Stratagene. The underlying mechanism consists of using short amino acid sequences, complementary to the intended sequence (called primers) carrying a variation of nucleic base. When integrated successfully, another amino acid residue is integrated into the protein during the transcription of mRNA by ribosomes. The complete list of primers used in this work can be found in the supplement (annexe A). All primers were 28 - 43 base pairs long, with a 40-60 % GC content, providing higher stability upon strand dimerisation.

The QuikChange protocol allows controlled site-directed mutagenesis, implemented by polymerase chain reaction (PCR). PCR is a commonly used technique to amplify specific DNA-sequences. While the primers will hybridise to the single-stranded DNA upon denaturation, the added polymerase will subsequently read and elongate the primer to match as complementary DNA sequence. As the new DNA chain now contains the mutation, it will be copied so that with each amplification cycle more DNA carrying the mutation is generated. For detailed PCR programs see supplement annexe E, table 8. 6 and 8. 7.

The PCR contained the respective forward and reversed primer (125 ng/µl each, Sigma-Aldrich), dNTPs (desoxynucleosidtriphosphate, 25 mM, 1 µl, Agilent Technologies, USA), plasmid-DNA (1µl, 150 ng/µl), Q5 reaction buffer (NEB) and a Q5 - HF polymerase (1µl, 2U/µl, NEB).

To validate the PCR products, a 1 % agarose gel electrophoresis was performed using a 0.5x TBE buffer mentioned under molecular cloning. After successful completion of the PCR, the product was digested by restriction endonuclease DpnI (1 µl, 20 U/µl, NEB) for 2 h at 37 °C, cutting methylated DNA exclusively found in the original plasmid, not containing the mutation.

Finally, the digested PCR product was transformed into XL1 blue cells, as described above, transforming 5 µl PCR product and plating 45 µl on 2x TY-agar-plates (containing ampicillin). The grown colonies are then picked and transferred into liquid 2x TY-medium for amplification (5 ml with 0.1 mg/ml ampicillin, overnight, 37 °C). The mutated plasmid was isolated from the cells with the same plasmid isolation kit as described in chapter 3.1.2, using the instruction manual of the High Pure Plasmid Isolation kit from Roche. The results were validated by DNA sequencing (GATC, now Eurofins, DE) and analysed with BioEdit (Ibis Biosciences, CA).

3.1.4. Recombinant protein synthesis

The recombinant proteins were overexpressed in *E. coli* C41 (DE3) bacterial cells (previously heat shock transformed and plated onto agar plates, as described above) in 2x TY medium supplemented with 0.1 mg/ml ampicillin, starting in a 100 ml pre-culture overnight at 37 °C. 8 ml of the pre-culture were then transferred in a 800 ml 2x TY medium main-culture and cultivated at 37 °C, until an optical density (OD, at 600 nm) of 0.6 to 0.8 was reached.

For NMR studies, isotope labelled proteins, were cultivated and expressed in minimal media supplemented with ^{15}N ammonium chloride and ^{13}C glucose as the only nitrogen and carbon sources. The minimal medium was freshly prepared with 160 ml 5x M9 stock solution (75 g KH_2PO_4 (Merck), 169.5 g $NaHPO_4$ (Merck), 12.5 g NaCl (Merck) in 1 l H_20, autoclaved), 3.2 ml $MgSO_4$ (1 M, Merck, autoclaved), 800 µl $FeCl_3$ (10 mM, sterile filtered, 0.2 µm pore size), 800 µl

Centrum vitamin tablet (1 tablet in 20 ml H_2O, vortexed, centrifuged and sterile filtered), 8 ml trace elements ($CaCl_2$*H_2O 10 mg; $ZnSO_4$*$7H_2O$ 10 mg; $MnSO_4$*H2O 10 mg; Thiamine 250 mg; Niacin 250 mg; Biotin 5 mg solved in 50 ml H_2O and sterile filtered), the isotopes ^{13}C glucose (1.6 ml/l) and ^{15}N NH_4Cl (0.6 ml/l) and 800 µl ampicillin (0.1 mg/ml) for 800 ml of culture medium. The Centrum vitamin tablet, trace elements, $FeCl_3$ and isotopes were kindly provided by our collaboration partner Dr. Ute Hellmich (University Mainz). In case only ^{15}N labelled protein was requested, 16 ml glucose (20 % (w/w), D +, Sigma-Aldrich, autoclaved) was added instead of ^{13}C glucose. For the regular protein synthesis, the minimal medium culture was started by a pre-culture (100 ml 2x TY-medium, over-night culture supplemented with ampicillin 0.1 mg/ml) and transferred as described above in a second pre-culture (800 ml 2x TY-medium supplemented with ampicillin 0.1 mg/ml). The cells were grown to OD_{600} 0.6 and centrifuged in autoclaved containers for 10 min, 5 000 rpm at room temperature. Cells were then resuspended in minimal medium so that OD_{600} was 0.3 and grown until OD_{600} 0.6-0.8 was reached.

Translation was induced by the addition of 1 mM isopropyl-b-D-thiogalactopyranoside (IPTG, Sigma-Aldrich) and subsequently shook overnight at 20 °C. The cells were harvested by centrifugation and resuspended in pH 8.0 buffer with 50 mM phosphate and 30 mM imidazole, supplemented with DNase (50 µl/800 ml TY-medium preparation, 4 mg/ml, Applichem, DE) and lysozyme (4 mg; Sigma-Aldrich). NTDs, CTD and mini-spidroins were isolated from clarified cell lysate after sonification (19 cycles of 10 sec pulse and 30 sec pause, QSonica, CL-334, USA) by affinity chromatography using Ni-Sepharose 6 Fast-Flow resin (GE Healthcare, USA) with 300 mM imidazole (pH 8, Sigma-Aldrich/Merck) followed by proteolytic cleavage of the His6-tag using thrombin from bovine plasma (100 U, Sigma-Aldrich). For the mini-spidroin purification, protein stability was improved by adding 200 mM NaCl to both running and elution buffer of Ni-NTA. The solution pH of the buffers was adjusted through calculations of the buffer salts using the online Liverpool buffer calculator (Centre for Proteome Research, UK). All buffers used here were adjusted by this approach, if not described otherwise.

For the CTD, the dialysed thrombin-digest was applied again to Ni-Sepharose chromatography to remove the His-tagged fusion protein.

For Hsp90, the proteins were then purified by anion exchange chromatography. Therefore, the pooled fractions from Ni-NTA were diluted 1:5 with H_2O to lower the ionic strength (*Is*) of the elution buffer (300 mM imidazole) and loaded on a POROS HQ column. By setting a gradient between the running buffer (20 mM Tris, at pH 8.0 adjusted with HCl) and the elution buffer (20 mM Tris, at pH 8.0 adjusted with HCl, and 1 M NaCl), the protein would charge-dependently elute from the column at a specific salt concentration. The resulting fractions were then pooled and concentrated using a 10 kDa molecular weight cut-off (MWCO) centrifugal concentrators (Sartorius Stedim Biotech, DE).

All Proteins were finally purified to homogeneity using size-exclusion chromatography (SEC, from GE Healthcare). Details regarding column size and buffer conditions used during SEC are listed in table 3. 1.

Table 3. 1: Column size and buffer conditions used for SEC.

Protein	SEC column	Buffer conditions	Storing conditions
NTDs	Superdex 75 (HiLoad 26/60)	200 mM Ambic	Lyophilised or concentrated and mixed with 30 % Glycerin
CTD	Superdex 75 (HiLoad 26/60)	200 mM Ambic	concentrated and mixed with 30 % Glycerin
Mini-spidroins	Superdex 75, Sephacryl 400	50 mM Pi, 200 mM *Is* (KCl), pH 7.0	concentrated and mixed with 30 % Glycerin
Hsp90	Sephacryl 400	40 mM HEPES, 200 mM *Is* (KCl), pH 7.5; degassed by argon	concentrated and flash-frozen in liquid nitrogen

Cysteine mutants of NTD proteins were supplemented with 10 mM DTT 1 h prior SEC and buffers were degassed with argon gas.

Finally, protein fractions were pooled and filtered (0.2 µm; LLG®, DE) and subsequently treated protein-dependently (see table 3. 1). For protein

concentration, centrifugal concentrators with 5 to 10 kDa MWCO were used (protein size dependent). Ultimately, all proteins were stored at - 20 °C. Purity of protein samples was confirmed by SDS-PAGE (4 – 12 % from TruPAGE™ Precast Gels, Sigma-Aldrich).

3.2. Fluorescence modification of proteins

Single point cysteine mutants (NTD wild-type contains no intrinsic cysteine) were labelled with the thiol-reactive maleimide derivative of the oxazine fluorophore AttoOxa11 (MW 719 g/mol, ε_{Oxa11} 125 000 $M^{-1} \cdot cm^{-1}$, Atto-Tec, DE). Labelling was carried out using a 5-fold molar excess of fluorophore reacting for 2.5 h at 25 °C in 50 mM 3(N-morpholino)propanesulfonic acid (MOPS, Sigma-Aldrich) at pH 7.5 containing 6 M guanidinium chloride and a 10-fold molar excess of tris(2-carboxyethyl)phosphine (TCEP, Sigma-Aldrich) to prevent thiol oxidation. Labelled protein was isolated from non-reacted dye using SEC (Sephadex G-25 resin, GE Healthcare) in pH 7.0 buffer containing 10 mM phosphate. The degree of labelling (DOL) was determined according to:

$$DOL = \frac{A_{max} \cdot \varepsilon_{Protein}}{(A_{280nm} - (A_{max} \cdot CF_{280})) \cdot \varepsilon_{Dye}} \quad (3.1)$$

where A_{max} is the absorption maximum, A_{280nm} the absorption maximum at 280 nm, with $\varepsilon_{Protein}$ as the extinction coefficients of the labelled protein, and ε_{Dye} of the used dye, and CF_{280} is the dye specific correction factor at 280 nm (for AttoOxa11 CF = 0.09).

Hsp90 was labelled with SeTau647 NHS (MW 1848 g/mol, ε_{SeTau} 200 000 $M^{-1} \cdot cm^{-1}$, SETA BioMedicals, USA). N-hydroxysuccinimidyl (NHS) -esters are amine-reactive reagents, which label amino groups unspecifically. Hsp90 and mutants (+M1, +M3, +M4) were labelled with SeTau647 at equal molarity in 100 mM $NaHCO_3$ labelling buffer, for 30 min at 25 °C in H_2O. Labelled protein was isolated with the same SEC column as described above (Sephadex G-25) and eluted with a buffer containing 50 mM phosphate and 200 mM *Is* (KCl) with pH adjusted to 7. The correction factor for SeTau 647 is 0.8. All extinction coefficients $\varepsilon_{Protein}$ relevant to this work can be found in the appendix annexe A.

3.3. Spectroscopy and sample preparation

3.3.1. Steady-state fluorescence spectroscopy

Fluorescence spectroscopy has a wide range of different applications. Fluorimeters direct a beam of light of a specific wavelength (isolated and adjustable by a monochromator) to a cuvette containing the sample. Dependent on the type of instrument, the fluorimeter has a typical spectral range of 200-900 nm, can be temperature controlled and measure time-dependently. In case fluorescence is emitted by the sample, it is usually detected perpendicularly to the excitation. Excitation and emission light are, for example, gated by slit widths, allowing regulation of the signal to noise ratio.

In this work, steady-state fluorescence emission spectra were recorded with a Jasco FP-6500 spectrofluorometer. For chemical denaturation, NTD samples were analysed at 10 µM protein concentration in a 10 mm path-length fluorescence cuvette (Hellma Analytics, DE). NTD fluorescence spectra were measured at the same concentration, under monomer conditions (pH 7.0 with 50 mM phosphate and 200 mM *Is* using KCl) and under dimer conditions (pH 6.0 with 20 mM 2-(*N*-morpholino)ethanesulfonic acid (MES, Sigma-Aldrich) and 20 mM *Is* using KCl) by exiting the sample at 290 nm (for clear delimitation to tyrosine) and detecting the emission between 300 and 450 nm. For the methionine residue (Met) -depleted and the newly implemented Met NTDs, the spectra were recorded with 100 µM protein solutions. The sample temperature was controlled using a Peltier thermocouple set to 25 °C throughout all experiments.

NTD fluorescence spectra with varying pH were recorded in 50 mM MOPS buffer (pH 6.4 to 8.2) and in 50 mM MES (pH 5.2 to 7.2) with ionic strength adjusted to 20 mM (KCl).

For chemical denaturation experiments, the samples were titrated from 0 to 8 M urea in 0.25 M steps in monomer buffer (pH 7.0 with 50 mM MOPS and 200 mM *Is* using KCl). For Met-depleted samples the titration was continued to 10 M urea in the monomer buffer. Salt-dependent measurements were conducted with 2 µM protein in monomer buffer and the solution *Is* adjusted to varying values using

KCl. All chemical denaturation measurements were recorded at 330 nm emission wavelength.

3.3.2. Absorption spectroscopy

Probing the absorption of light of a sample to determine the concentration of, for example, protein or DNA, can be quantified according to the Beer-Lambert law (see theoretical background equation 2. 1). For this purpose, a NanoPhotometer® (Implen, DE), usable from ultraviolet to infrared light, was employed.

Additional to protein and DNA concentration measurements, absorption assays were performed to detect the increase or decrease of a particular substance over time. Here, the absorption was measured to quantify the ATPase activity of Hsp90 (converting ATP to ADP) using a V-650 spectrophotometer (Jasco, DE). In general, the ATPase activity indicates how functional the Hsp90 protein is (Panaretou et al., 1998)(more about ATPase activity in chapter 4. 7). The ATPase buffer consisted of 2 mM phosphoenolpyruvate (PEP, P7127, Sigma-Aldrich), 50 U/ml lactate dehydrogenase (LDH, L2625, Sigma-Aldrich), 50 U/ml pyruvate kinase (PK, P7768, Sigma-Aldrich), 0.2 mM nicotinamide adenine dinucleotide hydrate (NADH, N8129, Sigma-Aldrich), 10 mM $MgCl_2$, 5 mM DTT and 2 mM ATP (adenosinetriphosphate, A6419, Sigma-Aldrich) in 40 mM HEPES with 200 mM *Is* (KCl), pH adjusted to 7.5. NADH decrease was measured by monitoring the absorption at $\lambda = 340$ nm and relating it to the ATP hydrolysis, according to:

$$ATPase\ rate\ k = \frac{-\frac{dA_{340}}{dt}}{\varepsilon_{NADH} \cdot d \cdot c_{ATPase}} \tag{3.2}$$

where the numerator corresponds to the time-dependent change in absorption. Further, d corresponds to the optical path length inside the cuvette, the extinction coefficient $\varepsilon_{\mathrm{NADH}}$ is 6220 $\mathrm{mol^{-1} \cdot cm^{-1}}$ and c_{ATPase} the concentration of the ATPase, which equates the applied protein concentration of Hsp90.

Experiments were performed at 37 °C. To start the reaction, 5 µM Hsp90 was added to the sample. The background was determined by Hsp90 inhibition using geldanamycin (Cayman Chemical, USA), pre-incubating 5 µM Hsp90 with

150 µM geldanamycin for 30 min, followed by the addition of 2 mM ATP to start the reaction.

3.3.3. Far-UV circular dichroism spectroscopy

Circular dichroism (CD) spectroscopy is a technique widely used to characterise chiral molecules and extract their structural, folding and binding properties (Greenfield 2006). The secondary structure of proteins is susceptible to environmental changes. Variations in temperature or chemical conditions (e.g. solution pH, added denaturant) will therefore impact the folding of the protein. For CD measurements the sample has to contain optically active, i.e., chiral molecules, present in proteins as amino acids. Those will absorb circular polarised light, according to whether the molecules are left or right-handed. The detectable difference in absorption between the two is the basis for CD spectroscopy.

The resulting wavelength-dependent signal (ellipticity) is particularly characteristic for proteins being in a pure α-helical or β-sheet formation. The CD-spectrum of a purely α-helical protein will have a global minimum at 222 nm and a second local minimum at 208 nm, whereas a β-sheet formed protein has a minimum at 215 nm. Therefore, the ellipticity measured at those wavelengths indicates the content of secondary structure. A mixture of those folds lead to a superposition of the characteristic spectra (Fig. 3. 1). The detection accuracy of this method is 100 % for α-helical proteins and 75 % β-sheets (Venyaminov and Vassilenko 1994).

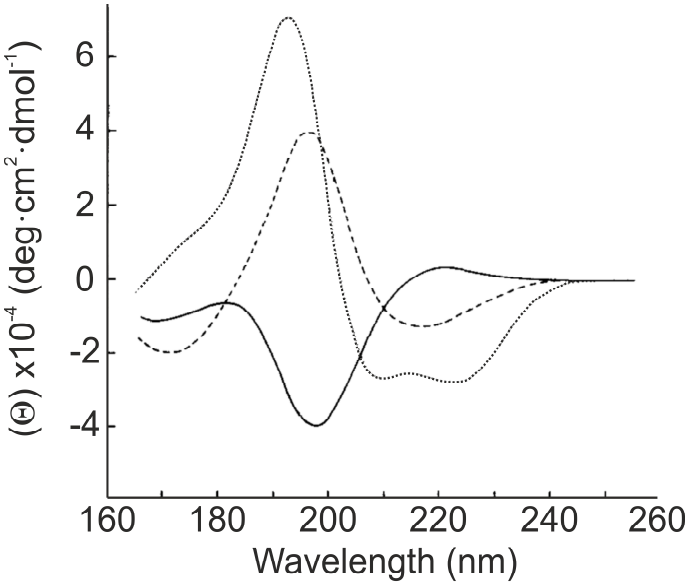


Figure 3. 1: Characteristic CD spectra of the two main types of protein secondary structure and unordered structure: α-helix, short dashed line; β-sheet, long dashed line; unordered structure, solid line. (adapted and with permission from Brahms and Brahms 1980)

Far-UV CD spectroscopy experiments were performed with a Jasco J-815 spectropolarimeter and a 1 mm path-length cuvette (Suprasil®, Hellma Analytics). For the recording of CD spectra, the samples contained 10 µM protein, except for the concentration-dependent experiments. The ellipticity was recorded from 190 to 260 nm. Chemical denaturation experiments were conducted under the same solution conditions as described above for steady-state fluorescence experiments. The signal was recorded at 222 nm, probing the α-helix secondary structure, whose share dropped gradually under denaturation conditions. Sample temperature was controlled using a Peltier thermocouple set to 25 °C (Peltier element PTC-423S/15, Jasco).

Thermal denaturation experiments were carried out in monomeric buffer (as described above) and a temperature ramp was applied with a rate of 1 °C/min from 10 °C up to 95 °C. Salt-dependent thermal denaturation data were acquired using 20 mM phosphate, pH 7.0 buffer with varying ionic strength using KCl.

3.3.4. Stopped-flow fluorescence spectroscopy

The stopped-flow (SF) method can be applied to probe how fast proteins fold and associate. Furthermore, SF can be paired with fluorescence spectroscopy, absorbance spectroscopy, circular dichroism or with synchrotron small-angle X-ray scattering (SAXS). These methods give access to a sub-second timescale, with a dead time of only 1-2 ms (Zheng et al. 2015).

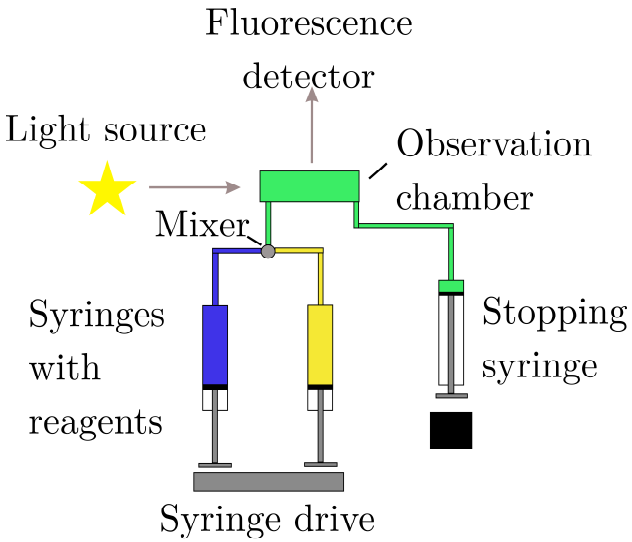


Figure 3. 2: Schematic representation of a stopped-flow setup. Two syringes are connected to a mixing zone, leading to the observation chamber, which ends with a stopping syringe. The reaction takes place in the observation chamber by fluorescence or absorbance detection. (adapted and with permission from Zheng et al. 2015)

A typical SF system is built by two syringes connected to a single detection chamber, where the content of both syringes can be rapidly mixed. The content is retained in the camber by a *stopping* syringe or valve, therefore functioning also as the eponym to this technique (Fig. 3. 2). The resulting spectroscopic signal originating from the detection chamber, e.g. fluorescence, absorption or CD-signal, can then be measured time-dependently.

Accordingly, the kinetics of NTDs were measured using an Applied Photophysics instrument (SX 18MV) equipped with a xenon-lamp as excitation source. Kinetics of association were measured by recording the native Trp fluorescence signal of the NTDs. Fluorescent control measurements were performed with N-Acetyl-L-

tryptophanamide (NATA), an analogue of tryptophan, with 1 µM in all used buffer conditions (Fig. 3. 3). Monomeric protein samples were prepared at varying concentrations in 10 mM MES at pH 7.0 with 150 mM *Is* using KCl and rapidly mixed into 20 mM MES buffer at pH 6.0, using a volumetric mixing ratio of 1:11. Each homolog was measured at three different concentrations between 0.1 µM and 1 µM.

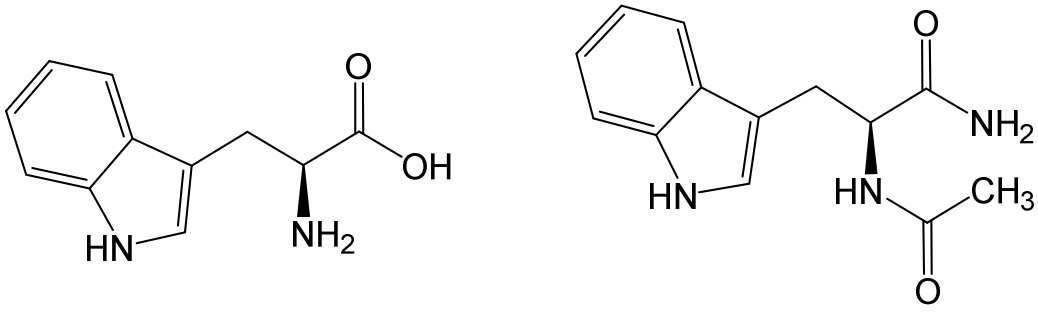


Figure 3. 3: Chemical structure of Tryptophan (left) and NATA (right).

NTD folding and unfolding kinetics were measured in chemical denaturation conditions by tracking the native Trp fluorescence signal, using an SFM-2000 BioLogic stopped-flow machine equipped with a 280 nm diode as the excitation source. The samples contained 100 µM protein and were prepared in buffered solutions with either 0 or 6 M urea in pH 7.0 buffer, with 50 mM MOPS and 200 mM *Is* using KCl. For kinetic measurements of the Met-depleted NTDs, the protein-containing solutions were mixed either in 0 or 9 M urea with the same buffer conditions as mentioned above. The samples were rapidly mixed into urea solutions of varying concentrations with a volumetric mixing ratio of 1:10. For salt-dependent measurements, the *Is* was adjusted using KCl. Samples were filtered through 0.2 mm syringe filters before measurement. All stopped-flow measurements were recorded at 25 °C, using a circulating water bath.

3.3.5. Stopped-flow far-UV circular dichroism spectroscopy

As mentioned above, SF can be combined with CD (see chapter 3.3.3). Thereby, fast kinetics of folding and unfolding of proteins can be analysed by measuring the changes in ellipticity in the far UV (222 nm). The combination of these

methods have been applied to investigate early intermediates in protein folding (Kelly and Price 2000).

Stopped-flow far-UV CD spectroscopy was carried out on an SFM-2000 stopped-flow machine (BioLogic Instruments, USA) coupled to a J-815 spectropolarimeter (Jasco) at 25 °C. Stopped-flow CD transients were recorded at 222 nm. For unfolding/refolding experiments, probing the transition of the native CTD dimer to the partially folded dimer and vice versa, 0.5 - 1.0 mM protein solution was rapidly mixed into 50 mM phosphate buffer (pH 7.0) with the ionic strength adjusted to 200 mM using KCl, containing various concentrations of urea and using a mixing ration of 1:10. Typically 10 shots were averaged to obtain transients of sufficient signal-to-noise ratio for analysis.

3.3.6. Analytical size-exclusion chromatography

Fluorescence spectroscopy and size-exclusion chromatography (SEC) can be combined in a high-performance liquid chromatography (HPLC). HPLC is primarily used in analytical chemistry. By separating molecules by size and detecting them through intrinsic tryptophan fluorescence, their association capacities can be determined.

Samples were prepared in a serial dilution style and were loaded on a Superdex-75 HR10/300 analytical gel filtration column (GE Healthcare) at 0.8 ml/min in 20 mM MES pH 6.0 with *Is* adjusted to 20 mM using KCl (monomer conditions). Sample concentration was reduced by at least tenfold for reasons of dilution on the column and in the detection flow path. Specified concentrations were measured at the highest fluorescence intensity during the experiment, corresponding to a specific elution volume. The elution of the protein was measured by UV absorbance. Measurements were carried out on a Jasco HPLC work station (LC-2000 Plus Series). For K_D estimation, 15 data points were collected, or otherwise indicated in the corresponding data table.

3.3.7. Thermophoresis

The interactions between a labelled sample and a ligand changes temperature dependently, which can be measured when a temperature gradient is applied. The fluorescence is detected over time and quantifies the binding affinity.

Samples of AttoOxa11 labelled L6-NTD-Q50C (40 nM) were analysed by mixing the sample with increasing concentrations of unlabelled L6-NTD, using a Nanotemper Monolith instrument (DE). After localisation of the standard untreated capillaries, thermophoresis was performed at 50 % heating power and 50 % illumination intensity. The data points were plotted as the ratio of fluorescence intensity before and at the end of 30 s of heating when the new stable equilibrium intensity had been achieved.

3.3.8. Fluorescence correlation spectroscopy

Fluorescence correlation spectroscopy (FCS) is a single-molecule technique, used to study conformational dynamics of biological probes at equilibrium, amongst others. As the excitation and detection volumes are nanoscopic ($0.5\text{-}1 \cdot 10^{-15}$ l, Nie and Zare 1997), only a few molecules contribute to the signal simultaneously. Therefore, the trace of each detected molecule will be thus averaged and undergoes an autocorrelation calculation. The corresponding diffusion time constant of a molecule can then be determined, characteristic for a particular size of the examined molecule. Typically, a fluorescently labelled sample is required to measure FCS. If the labelled molecule binds to another (preferably bigger) molecule, the diffusional mobility of the complex will be decreased, indicated by a shifted autocorrelation curve. By this, protein-ligand binding affinities can be analysed. Furthermore, tracking the temporal fluctuation in fluorescence can reveal information regarding the molecular dynamics of the sample, e.g. due to conformational dependent fluorescence quenching (Fig. 3. 4). Here, a labelled donor interacts with an acceptor, often engineered ingeniously in the same molecule at an appropriate distance. Time-dependent quenching and / or unquenching will then lead to on and off-states in the fluorescence trace with characteristic time constants. One important effect of quenching is based on the photoinduced electron transfer (PET, described in section 2.2.4).

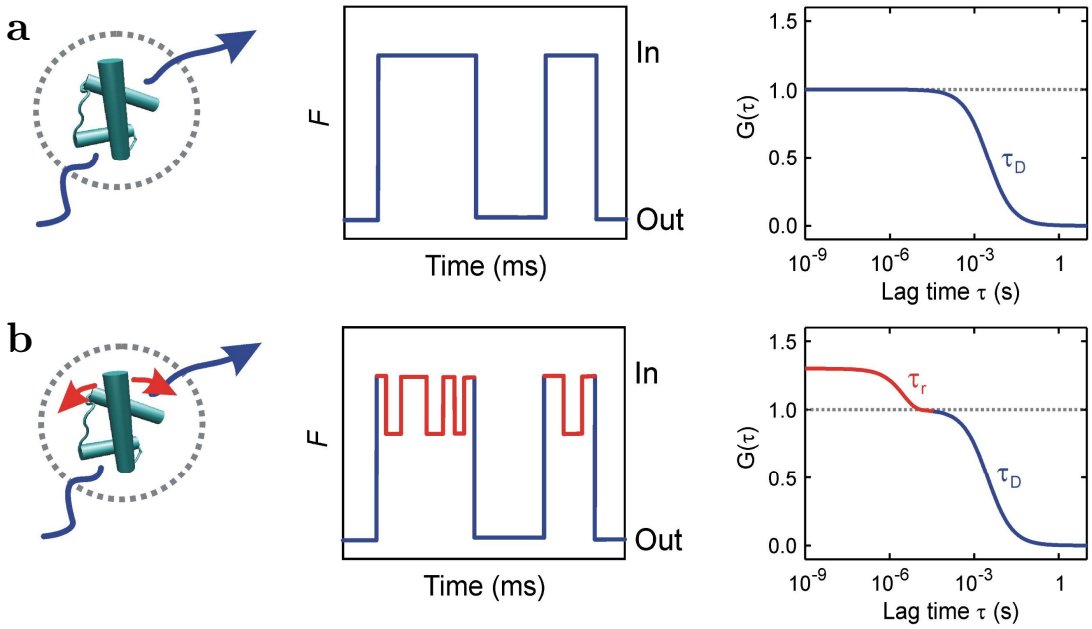


Figure 3. 4: Protein dynamics measured with PET-FCS using fluorescently labelled proteins. (**a**) The protein diffuses freely through the detection volume (left, dotted circle), causing on and off states of the fluorescent signal for the detector (F, middle), here labelled in and out, respectively. The autocorrelation curve (right) exhibiting only one decay. (**b**) Protein diffusion with simultaneous intrinsic dynamics, shown in red, creating an additional decay of the autocorrelation curve. (with permission from Jensen et al. 2011)

By setting a logarithmic scale for the lag time **τ**, the correlation curve has a sigmoidal shape. This leads to a straightforward display of the characteristic diffusion constant, being located at the mid-point of the normalised correlation curve. The relative amplitude of the signal is inversely proportional to the number of the observed molecules. An additional time constant and amplitude may be obtained from a PET effect.

Here, PET-FCS was performed on a home-built confocal fluorescence microscope setup that is based on an inverse microscope (Zeiss Axiovert 100 TV, DE) and was described in detail elsewhere (Ries et al. 2014). In short, the diode laser beam (637 nm, Coherent Cube, USA) was coupled to a dichroic beam splitter (645DLRP, Omega Optics, USA) before leading into an oil-immersion objective lens (Zeiss Plan Apochromat, 63x, NA 1.4). The fluorescence signal was collected by the same objective, filtered by a band-pass filter (675RDF50, Omega Optics), and imaged onto two avalanche photodiode detectors (fibre-coupled APDs;

SPCM-AQRH-15-FC; Perkin Elmer, USA). The beam was split using cubic non-polarizing (Linos, DE) and multi-mode optical fibres (100 µm diameter). The APDs signals were recorded in the cross-correlation mode, employing a digital hard-ware correlator device (ALV 5000/60X0, multiple tau digital real correlator) bypassing the detector dead-time and after-pulsing effects. Average laser power was adjusted using an optical density filter to 400 µW before entering the back aperture of the microscope.

Fluorescently labelled NTDs were diluted to 1 nM protein concentration in buffered solutions containing 0.3 mg/ml BSA (Sigma-Aldrich) and 0.05 % Tween-20 (Sigma-Aldrich) as additives to suppress glass surface interactions. The buffer also contained 50 mM phosphate and 200 mM *Is* (with KCl) with pH adjusted to 7.0. For SeTau647 labelled Hsp90 samples, 0.1 nM protein concentration was used, with the same buffer conditions as for the NTD measurements. Samples were filtered before measurements (0.2 µm syringe filter) and transferred onto a microscope slide (with mould, 76x26 mm, Ø15 mm; Thermo Fisher Scientific) with a coverslip (high-precision, 24x60 mm, 170±5 µm; Marienfeld, DE). Sample temperature was set to 25 °C by a custom-built objective heater. For each sample, three individual measurements were performed varying from 3 to 10 min each, equal for each measurement set.

3.4. Data analysis

For all data analyses, the software OriginPro was used (OriginPro 2016G).

3.4.1. Equilibrium denaturation

For analysis of the equilibrium denaturation data, a two-state thermodynamic model was used (Santoro and Bolen 1988):

$$S(P) = \frac{\alpha_N + \beta_N \cdot P + (\alpha_D + \beta_D \cdot P) \cdot \exp(-\Delta G_{D-N}(P)/RT)}{1 + \exp(-\Delta G_{D-N}(P)/RT)} \quad (3.3)$$

where S is the spectroscopic signal as a function of P, either being the denaturant concentration (in M) or temperature (in K); α_N, β_N, α_D, and β_D are the linear baseline slopes of native and denatured states, R is the gas constant

($1.987 \cdot 10^{-3}$ kcal $\cdot$ K^{-1} $\cdot$ mol^{-1}), T the temperature of measurement (25 °C), and ΔG_{D-N} the difference in free energy between native and denatured state.

For probing the folding equilibrium, urea was used as denaturant (P = [urea]) in chemical denaturation experiments. ΔG_{D-N} as a function of chemical denaturant is described by the following linear-free energy equation (Tanford 1968):

$$\Delta G_{D-N}([urea]) = \Delta G_{D-N} - m_{D-N}[urea] \quad (3.4)$$

where ΔG_{D-N} (kcal $\cdot$ mol^{-1}) is the change in free energy dependent by urea concentration (M) and m_{D-N} (kcal $\cdot$ M^{-1} $\cdot$ mol^{-1}) is the equilibrium m-value describing the surface accessibility of the folded protein to the denaturant. Experimental errors of ΔG were determined from propagated errors of fitted values of mid-point concentrations of urea ([urea]$_{50\%}$; in M) and m_{D-N} values.

Under thermal denaturation condition, the folding equilibrium is perturbed by heat ($P = T$). The free energy of unfolding ΔG_{D-N} (kcal $\cdot$ mol^{-1}) is described by rearrangement of the Gibbs Helmholtz relationship with equation 3.3 (Jackson and Fersht 1991; Nicholson and Scholtz 1996) as:

$$\Delta G_{D-N}(T) = \Delta H_m \cdot \left(1 - \frac{T}{T_m}\right) - \Delta c_p \left(T_m - T + T \ln\left(\frac{T}{T_m}\right)\right) \quad (3.5)$$

where ΔH_m (kcal $\cdot$ mol^{-1}) is the enthalpy of unfolding temperature dependently, T_m (K) the mid-point temperature and Δc_p is the difference in heat capacity between native and denatured state (14 cal K^{-1} mol^{-1} per amino acid, Myers et al. 1995).

Changes in the free energy of unfolding $\Delta\Delta G$ (kcal $\cdot$ mol^{-1}), which is caused by the change of solution ionic strength Is (M), was estimated with the Schellman formalism (Becktel and Schellman 1987):

$$\Delta\Delta G = \frac{\Delta H_m}{T_m} \cdot \Delta T_m \quad (3.6)$$

where ΔH_m (kcal $\cdot$ mol^{-1}) is the enthalpy of unfolding and ΔT_m (K) is the difference in the mid-point temperatures of the denaturation curves.

3.4.2. Kinetics of folding and association

Stopped flow transients of folding and unfolding experiments were fitted to a single exponential function containing a linear baseline drift:

$$S(t) = Aexp(-k_{obs}t) + bt + c \tag{3.7}$$

with S as fluorescence signal as a function of time, A is the amplitude and k_{obs} the observed rate constant of the transient, which is the sum of the microscopic rate constants for folding and unfolding (k_f and k_u, in s^{-1}). The parameters b and c describe the linear drift and offset of the baseline.

The development of k_{obs} as a function of urea concentration was calculated by fitting the data to the chevron model for a barrier-limited two-state transition following a linear-free-energy relationship (Jackson and Fersht 1991):

$$log\ k_{obs}([urea]) = \log[k_f \exp\left(-\frac{m_f[urea]}{RT}\right) + k_u \exp\left(\frac{m_u[urea]}{RT}\right)] \tag{3.8}$$

where m_f and m_u are the kinetic m-values of folding and unfolding (in $kcal \cdot mol^{-1} \cdot M^{-1}$), k_f and k_u are the microscopic rate constants of folding and unfolding under standard solvent conditions in the absence of denaturant, respectively. The resulting equilibrium free energy ΔG is then:

$$\Delta G = -\mathrm{RT} \ln\left(\frac{k_u}{k_f}\right) \tag{3.9}$$

The combination of kinetic and thermodynamic data indicates the similarity between the transition state and the native state (Tanford 1970). For this, the Tanford β-value was calculated as:

$$\beta_T = \frac{m_f}{m_f + m_u} \tag{3.10}$$

NTDs self-association rate constants were obtained from fitting kinetic transients to a reaction model of protein dimerisation (Mallam and Jackson 2007):

$$2A \overset{k_{ass}}{\longleftrightarrow} A_2 \tag{3.11}$$

$$\frac{d[A_2]}{dt} = k_{ass}[A]^2 \tag{3.12}$$

where A describes a folded, monomeric NTD, A_2 is the dimeric NTD and k_{ass} is the bimolecular rate constant of association ($M^{-1} \cdot s^{-1}$). The differential equation can be solved to (Mallam and Jackson 2007):

$$S(t) = S_{t_0} + S\frac{(k_{obs}t)}{(1+ k_{obs}t)} \quad (3.13)$$

where $S(t)$ is the time-dependent signal, S is the change in signal amplitude, and S_{t0} is the signal at $t = 0$. The observed rate constant is proportional to the association rate constant k_{ass}, like (Mallam and Jackson 2007):

$$k_{obs} = c_A \cdot k_{ass} \quad (3.14)$$

where c_A is the monomer protein concentration.

The equilibrium dissociation rate constants of the NTD were obtained by fitting kinetic transients of chasing experiments to a mono-exponential rise function. The equilibrium dissociation constant K_D (in M) was calculated by:

$$K_D = \frac{k_{diss}}{k_{ass}} \quad (3.15)$$

where k_{diss} is the dissociation constant (s^{-1}). The kinetic transients of the folding and unfolding transition of the CTD (N_2 native dimeric folded state to I_2 dimeric intermediate unfolded state) were measured by stopped-flow far-UV CD spectroscopy. The transients were fitted to a mono-exponential decay function as in equation 3.7 (without baseline drift) and analysed by a barrier-limited two-state transition containing the same linear-free energy relationship as in equation 3.8. Errors of values derived from kinetic experiments are standard errors from regression analysis and propagated standard errors.

3.4.3. NTD dimerisation assays

In thermophoresis experiments, fluorescently modified NTDs associated with non-modified NTD present at excess concentration. Equilibrium dimerisation was therefore described by the model for a conventional bi-molecular binding isotherm ($N + N^* = NN^*$). The complex is composed of fluorescently modified NTD and non-modified NTD. The concentration of the complex is described as (Hulme and Trevethick 2010):

$$[NN^*] = \frac{[N^*]_t[N]}{[N]+K_D} \quad (3.16)$$

where NN^* is the concentration of the NTD dimer, N^*_t is the total concentration of fluorescently modified NTD and N is the non-modified NTD concentration.

The K_D corresponds to the equilibrium dissociation constant. The thermophoresis ratio of fluorescence signal was modelled as:

$$\frac{S-S_u}{S_b-S_u} = \frac{[N]}{[N]+K_D} \quad (3.17)$$

where S is the observed signal, S_b is the signal of the bound state and S_u the signal of the unbound state.

Analytical SEC experiments generated binding isotherms. Therefore, the elution volume V_E (ml) was plotted against the NTD concentration (M). The binding isotherms were fitted using a model based on a thermodynamic equilibrium of dimerisation ($2N = N_2$), where N and N_2 denote for NTD monomer and dimer, respectively. The equilibrium dissociation constant K_D corresponds to:

$$K_D = \frac{[N_2]}{[N]^2} \quad (3.18)$$

V_E values are composed of:

$$V_E = V_E^N + F_{N2}(V_E^{N2} - V_E^N) \quad (3.19)$$

where V_E^N and V_E^{N2} are the elution volumes of monomer and dimer, respectively. F_{N2} is the fraction of dimer, which is described as:

$$F_{N2} = \frac{[N_2]}{c_t-[N_2]} \quad (3.20)$$

where c_t is the total concentration of monomeric protein. N_2 is described as:

$$[N_2] = \frac{K_D}{8}\left(1 - \left(\left(1 + \frac{4c_t}{K_D}\right)^2 - \frac{16c_t^2}{K_D^2}\right)^{0.5}\right) + \frac{c_t}{2} \quad (3.21)$$

3.4.4. PET-FCS

Autocorrelation functions (ACFs) were fitted based on an analytical model for translational diffusion of a molecule that manifests chemical relaxations (Krichevsky and Bonnet 2002). ACFs were fitted to a function for diffusion of a globular molecule with no (eq. 3. 22), two (eq. 3. 23) or three (eq. 3. 24) independent, single-exponential relaxations, described as:

$$G(\tau) = \frac{1}{N\left(1+\frac{\tau}{\tau_D}\right)} \quad (3.\,22)$$

$$G(\tau) = \frac{1}{N\left(1+\frac{\tau}{\tau_D}\right)}\left(1 + a_1\exp\left(-\frac{\tau}{\tau_1}\right) + a_2\exp\left(-\frac{\tau}{\tau_2}\right)\right) \quad (3.\,23)$$

$$G(\tau) = \frac{1}{N\left(1+\frac{\tau}{\tau_D}\right)}\left(1 + a_1\exp\left(-\frac{\tau}{\tau_1}\right) + a_2\exp\left(-\frac{\tau}{\tau_2}\right) + a_3\exp\left(-\frac{\tau}{\tau_3}\right)\right) \quad (3.\,24)$$

where τ is the lag time (s), τ_D is the experimental diffusion time constant (s), N is the average number of molecules in the detection volume, a_1 and a_2 are the observed amplitudes of the first and the second relaxation, and τ_1 and τ_2 are the according time constants. The equation 3.22 was used for molecules showing only simple diffusion with no additional relaxation. For all NTD measurements equation 3.23 was used, while for all K_D titration experiments of Hsp90 three exponential relations were fitted to the ACFs with equation 3.24. The two-dimensional model is valid for three-dimensional diffusion since the lateral components of the volume are often very small in relation to the axial component.

The microscopic time constants of the NTD Trp (τ_{out} and τ_{in}, in s) were calculated from the observed amplitudes a_1 and time constants τ_1 from the ACF with (Schulze et al. 2016):

$$\tau_1 = \tau_{out} + \tau_{in} \quad (3.\,25)$$

$$a_1 = \frac{\tau_{out}}{\tau_{in}} \quad (3.\,26)$$

where τ_{out} describes the time constants of the Trp moving out of the NTD core and τ_{in} the time constant into the NTD core. Errors are standard errors from regression analysis and propagated standard errors.

4.Results and Discussion

The following chapter is based on three published manuscripts and additional results, unpublished by the time of completion of this thesis. In case specific results, required for the integrity of the presented argumentation, originated from experiments, not performed by myself, the person responsible is clearly acknowledged. The respective person is named by their initials. If no such specific information is provided, the respective experiments, analyses, interpretation and illustrations were conducted by myself.

4.1. Folding and association of homologous spidroin NTDs

The mechanism of silk synthesis is thought to be conserved for all spiders. However, webs from different species show substantial differences in size, geometry and overall mechanical properties. Those differences could be caused by sequence variations in the repetitive middle segments of spidroins, chemical and / or physical conditions during fibre formation (Vollrath 2000; Vollrath and Knight 2001; Heim et al. 2009), but also by the non-repetitive terminal domains. The N-terminal domain (NTD) sequence is highly conserved in various spider species. Therefore, the question arose if the sequence conservation is also reflected in the silks mechanism of fibre formation. What is the contribution of the terminal domains to the spider silk, apart from its function as a starting point of fibre polymerisation? In the following chapter, I will present the results of the comparison of the qualitative characteristics of four NTD homologs, namely the *Euprosthenops australis* (Ea), *Latrodectus hesperus* (Lh), *Latrodectus geometricus* (Lg) and *Nephila clavipes* (Nc).

4.1.1. Association of spidroin NTDs

All NTDs share a particular structural feature: a single tryptophan (Trp) wedged in the centre of the protein core. This position seems to be conserved for all analysed NTDs under monomeric conditions (see also chapter 4. 4). Upon dimerisation, the tryptophan gets released from the core into a more solvent-exposed position, which has been shown for Ea NTD by crystallography and NMR spectroscopy (PDB-code for monomer 2lpj by Jaudzems et al. 2012, and dimer 3lr2 by Askarieh et al. 2010). The fluorescence signal of the tryptophan can be used as an indicator of the NTD association and therefore, dimerisation. In the monomeric NTD form, the Trp shows a strong fluorescent signal, compared to the dimeric form. A change of solution pH triggers the dimerisation of the NTD. In the spinning gland, the pH drops gradually from 7 to 5 (Andersson et al. 2016). We analysed this phenomenon for four homologues NTDs. Therefore, we probed each homolog in different buffers with pH ranging from 5.2 to 8.2 to determine the exact pH conditions of dimerisation (Fig. 4. 1. 1 a).

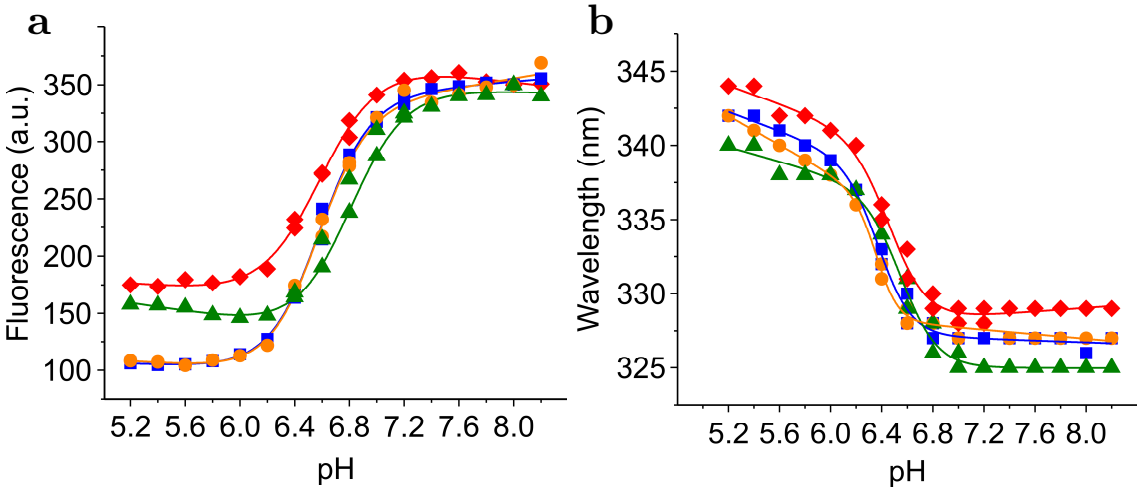


Figure 4. 1. 1: Dimerisation of NTD homologs. (**a**) Tryptophan fluorescence emission intensities measured in different pH solutions at 330 nm. The fluorescence intensities are normalised to the maximum fluorescence of Ea at pH 8, for reasons of clarity. (**b**) Wavelengths of the measured maximal fluorescence intensities in different pH solutions. Colour code: Lh - blue squares; Lg - orange spheres; Nc - green triangles; Ea - red diamonds. Solid lines are guide to the eye. (Heiby et al. 2017)

The Trp fluorescence emission showed that all NTD homologs were dimeric under pH 6.0 conditions and mostly monomeric under pH 7.0 conditions. In order to achieve comparability between the buffers, only small salt concentrations were used (ionic strength adjusted to 20 mM with KCl), leading to an incomplete monomeric state at high pH. Yet, in the spinning gland, a higher salt concentration aids the NTDs to stay monomeric by shielding electrostatic forces (Askarieh et al. 2010). Therefore, in subsequent experiments, monomeric buffers have an ionic strength (*Is*) of 200 mM, if not stated otherwise.

Based on the above-described results, the Trp fluorescence signal was used to characterise the association mechanism of the NTDs. By measuring the Trp emission at different pH values, we could track its bathochromic shift in more detail (Fig. 4. 1. 1 b). They all showed maximal fluorescence at pH 5.2 between 340 and 345 nm, while dropping with increased pH value from 6.2 to 6.6, and settling at constant emission with wavelengths between 325 to 330 nm at pH 7.0. For more detailed visualisation, fluorescence spectra of the homologues are displayed at pH 7 and pH 6 in Fig. 4. 1. 2 a. For better comparability, the results were normalised to the maximum fluorescence intensity at pH 7. Lg and Lh are equivalent in their Trp fluorescence signature, with a relative quantum yield of ~30 % at pH 6. Nc and Ea behaved similarly but showed less Trp quenching at pH 6, with a quantum yield of 40 - 60 %, respectively, indicating that the Trp gets less solvent-exposed. To ensure that monomers and dimers are purely present under each pH condition and no sub-populations are measured, we determined the molecular mass, together with our collaborator Dr. Chris Johnson (Fig. 4. 1. 2 b).

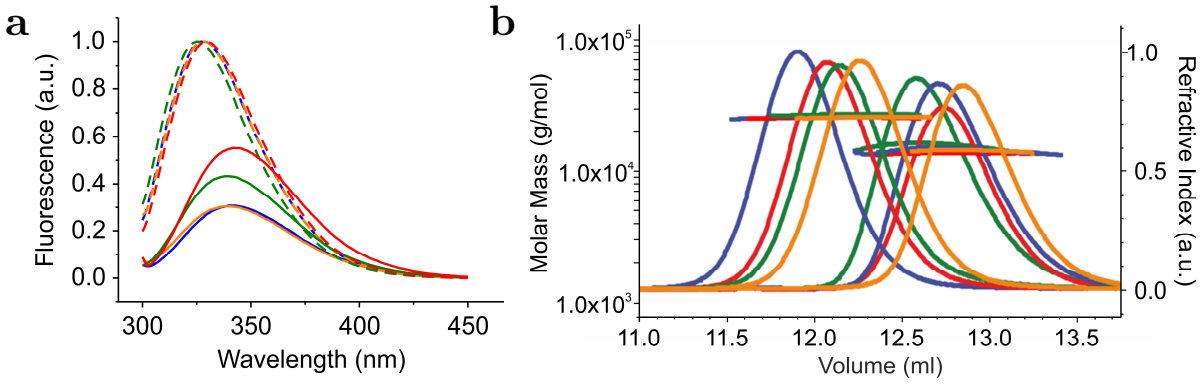


Figure 4. 1. 2: Monomeric and dimeric characterisation of NTDs. (**a**) Tryptophan fluorescence spectra at pH 7 (dashed lines) to pH 6 (solid lines) for all homologues. (**b**) SEC-MALS elution bands at pH 7 and pH 6 solution conditions and measured molecular mass of the homologues as monomer and as a dimer, pH dependently, shown by the horizontal lines. Colour code: Lh – blue; Lg – orange; Nc – green; Ea – red. ((a) from Heiby 2015; Figure (b) measured and created by CJ; Heiby et al. 2017)

All homologs eluted from analytical size-exclusion chromatography in combination with MALS spectroscopy (SEC-MALS) with approximately 14 kDa under monomeric conditions (pH 7 with 50 mM phosphate and 200 mM *Is* adjusted using KCl), and as a dimer with approximately 26 kDa under dimeric conditions (pH 6 with 20 mM MES and 60 mM *Is*), as expected (Fig. 4. 1. 2 b).

To explore the kinetics of NTD association and dissociation, we used stopped-flow spectrometry, recording the native Trp fluorescence signal. In order to investigate the rate constant of association k_{ass}, the protein was rapidly mixed from a pH 7 solution into pH 6 buffer (Fig. 4. 1. 3 a). The single traces were then fitted using a kinetic model of dimerisation (table 4. 1. 1). On the other hand, fluorescently modified protein (G3C-NTD labelled with AttoOxa11) contained in a pH 6.0 buffer solution, was rapidly mixed in the same buffer condition containing a tenfold of unlabelled protein, to measure dissociation rate constants k_{diss} (Fig. 4. 1. 3 b). The chasing experiment relied on the dissociation and re-association of the dimer with competitive monomers carrying a label, reforming under equilibrium conditions.

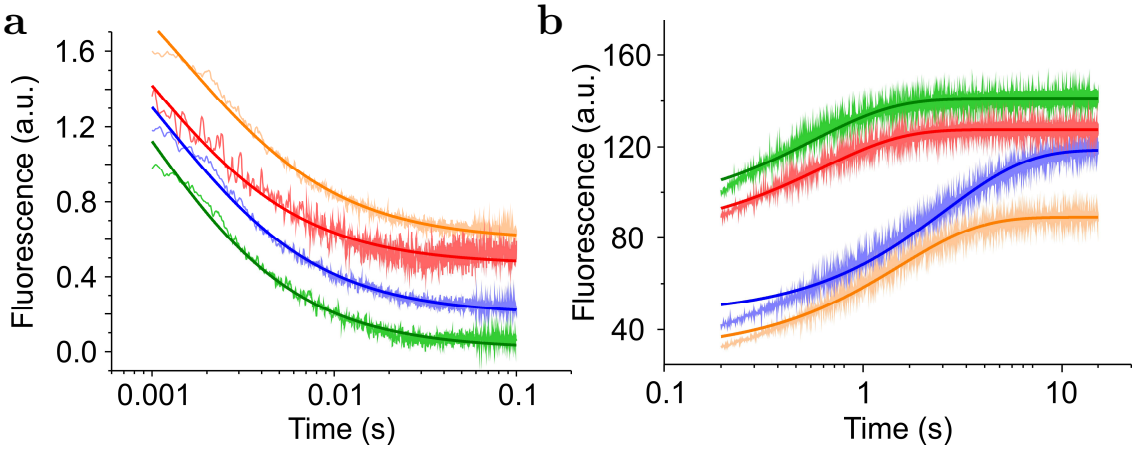


Figure 4. 1. 3: Kinetics of association and dissociation of NTD homologs. (**a**) Stopped flow measured kinetics of association as representative averaged transients, by tracked tryptophan fluorescence. Dimerisation measured by mixing the NTD protein from pH 7 to pH 6. The traces are shifted by 0.2 each au for clarity. Data were fitted with a kinetic model of dimerisation, represented by dark-coloured lines. (**b**) Stopped flow measured kinetics of dissociation as representative averaged transients, by tracking fluorescently modified NTD proteins. The labelled NTDs were rapidly mixed with a tenfold excess of unlabelled NTD, in pH 6 solutions. The traces are shifted arbitrarily along the y-axis for clarity. The data were fitted with a mono-exponential fit, represented by dark-coloured lines. Colour code: Lh – blue; Lg – orange; Nc – green; Ea – red. (Heiby et al. 2017)

The kinetic transients were analysed by a model of dimerisation using exponential functions (table 4. 1. 1). Slight deflections of the fits can be seen on the fast time scale of the dissociation curves, which can be explained by the presence of minor subpopulations of oligomeric species formed under these solution conditions (Heiby et al. 2017), but did not influence the measured rate constants significantly. From the rate constants k_{ass} and k_{diss} we then calculated the equilibrium dissociation constants K_D.

Table 4. 1. 1: Rate constants of association k_{ass}, dissociation k_{diss} and equilibrium dissociation constant K_D of the homologous NTDs. (Heiby et al. 2017)

NTD	k_{ass} (10^9 $M^{-1} \cdot s^{-1}$)	k_{diss} (s^{-1})	K_D (10^{-9} M)
Lh	2.19 ± 0.30	0.31 ± 0.02	0.14 ± 0.01
Lg	1.39 ± 0.12	0.64 ± 0.02	0.46 ± 0.03
Nc	2.58 ± 0.20	1.88 ± 0.05	0.73 ± 0.04
Ea	1.58 ± 0.24	1.68 ± 0.43	1.07 ± 0.11

As listed in table 4. 1. 1, the microscopic association rate constants k_{ass} of all homologs were similarly fast in the range of 1.4 to 2.6 10^9 M^{-1} s^{-1}, constituting a very fast pH-dependent dimerisation, while the dissociation rate constant k_{diss} ranged between 0.3 and 1.9 s^{-1}. The resulting K_D values ranged from 0.14 nM for the Lh NTD up to 1.07 nM for the Ea NTD. This reveals that the NTDs are strong dimers, in the same categories as high-affinity antibodies, binding in the low nanomolar range (Pan et al. 2016).

4.1.2. Equilibrium denaturation of NTDs

To probe the folding stability of the spidroin NTDs, we performed thermal and chemical denaturation experiments in combination with circular dichroism (CD) and fluorescence spectroscopy. During denaturation, the protein loses its secondary structure which can be monitored by the ellipticity at 222 nm by CD spectroscopy and the gain of Trp fluorescence by fluorescence spectroscopy, while the Trp gets gradually solvent-exposed upon unfolding.

Thermal denaturation experiments were conducted under monomeric conditions, at pH 7.0 (50 mM phosphate, 200 mM *Is* adjusted with KCl) (Fig. 4. 1. 4). Folding data could be described by a barrier-limited two-state transition model, from which we extracted mid-point melting temperatures T_m and enthalpy of unfolding ΔH (table 4. 1. 2). The thermal denaturation of Lg shown in Fig. 4. 1. 4 was measured by Suhaila Rajab within the scope of her bachelor thesis, co-supervised by myself.

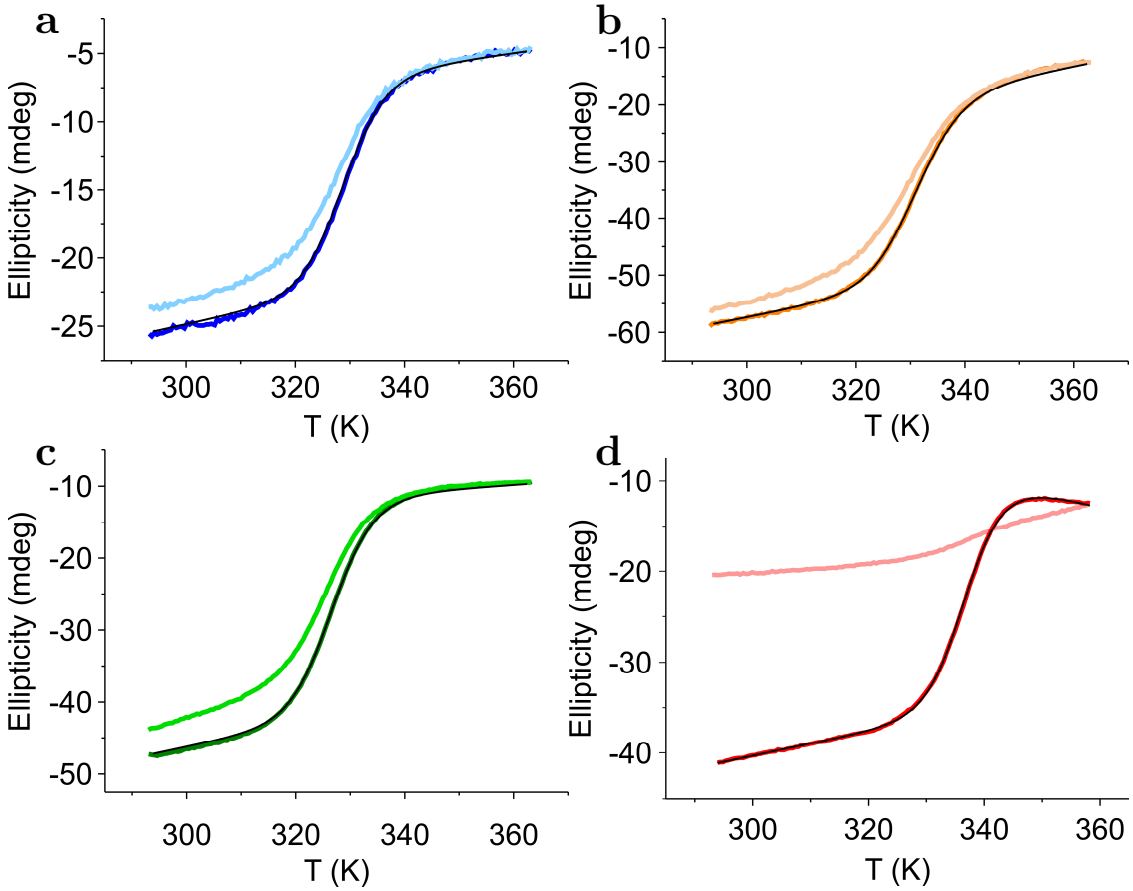


Figure 4. 1. 4: Thermal denaturation of homologous NTDs under equilibrium conditions measured by CD. (**a**) Lh – blue; (**b**) Lg – orange; (**c**) Nc – green; (**d**) Ea – red. Denaturation (bold coloured lines) and renaturation (lighter coloured lines) measured at pH 7 (200 mM *Is*) at 222 nm. Temperature (T) indicated in K. Melting curves are fitted using a model for a two-state equilibrium (black lines).

The denatured proteins did not fully regain the original ellipticity upon renaturation. Lg, Lh and Nc refolded nearly completely, while Ea denatured irreversibly. Thermal denaturation showed Lh and Lg have similar melting temperatures with approximately 330 K. Nc was slightly more unstable with 327 K, and Ea was most stable with 337 K. The enthalpies of unfolding (ΔH) were similar for Lh, Lg and Nc. A strong, non-covalent intramolecular interaction stabilised their fold. The fold of Ea was slightly more stable than the others, which could be explained by the irreversibility of thermal unfolding (Heiby et al. 2017).

Table 4. 1. 2: Equilibrium thermal denaturation data. Mid-point melting temperatures T_m and enthalpies of unfolding ΔH_m at the transition midpoint of the homologous NTDs. (Heiby et al. 2017)

NTD	*Lh*	*Lg*	*Nc*	*Ea*
T_m (K)	329	331	327	337
ΔH_m (kcal/mol)	54.8 ± 0.7	50.1 ± 0.4	54.3 ± 0.2	67.5 ± 0.4

Chemical denaturation was performed using urea as a denaturant. The denaturation was probed by CD and fluorescence spectroscopy and results were combined in Fig. 4. 1. 5.

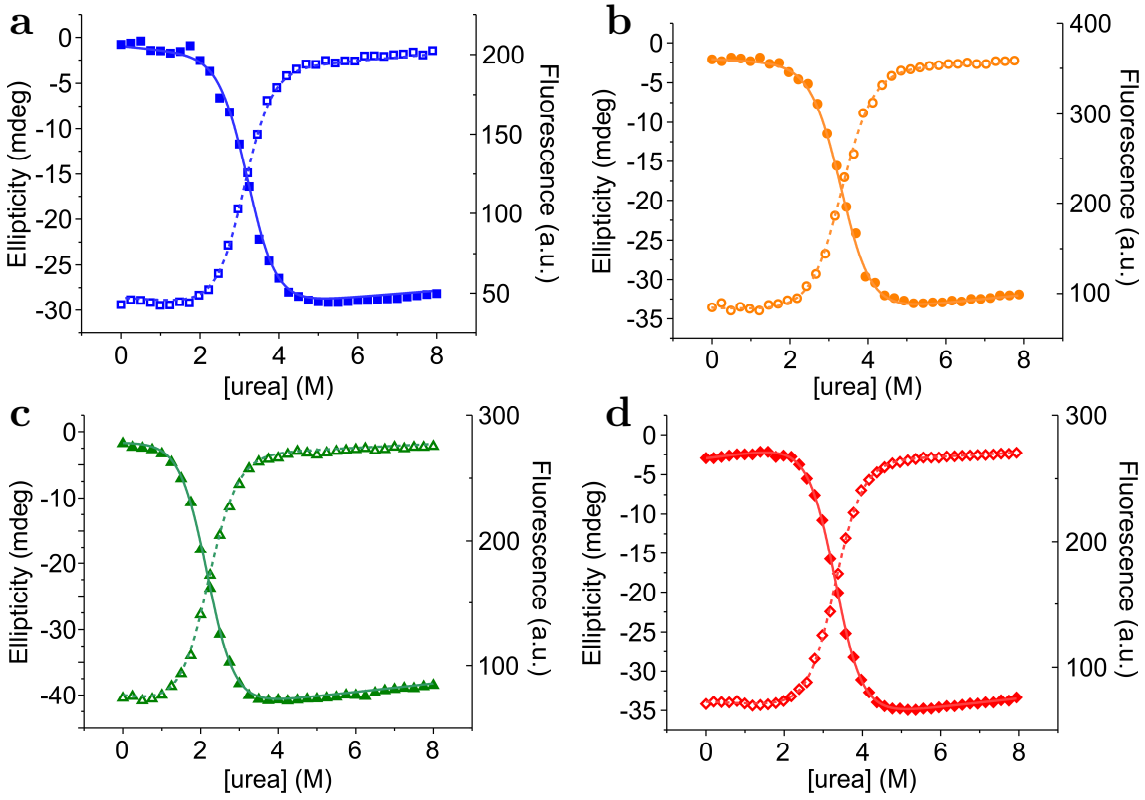


Figure 4. 1. 5: Chemical denaturation of homologous NTDs under equilibrium conditions measured by CD and fluorescence spectroscopy. (**a**) Lh – blue; (**b**) Lg – orange; (**c**) Nc – green; (**d**) Ea – red. Titration of NTD in urea in pH 7 buffered solution (200 mM *Is*) and measured at 222 nm with CD (closed symbols), and tryptophan fluorescence at 330 nm (open symbols). Data are fitted with a model for a two-state equilibrium (solid lines, CD; dashed lines, Trp fluorescence).

All NTD homologs could be described by a two-state model. Therefore all titration curves were fitted using a thermodynamic model for a two-state equilibrium. The chemical denaturation experiments were performed by Suhaila Rajab (Lg, Rajab 2015), Charlotte Rat (Nc, Rat 2015), Ries et al. 2014 (Ea) and myself (Lh, Heiby 2015). The stability of the NTDs appeared to be similar, with an exception for the Nc, which showed a reduced stability, consistent with the thermal denaturation experiments. The resulting transition mid-points and m-values were in good agreement between the CD and fluorescence spectroscopy data (table 4. 1. 3). Furthermore, the calculated free energies of unfolding were consistent with the previously discussed results for the thermal denaturation. Here again, judging by ΔG, Ea was the most stable NTD homolog, followed by Lh and Lg which were very similar to each other and the Nc being less stable in terms of transition mid-point and therefore the resulting free energy.

Table 4. 1. 3: Equilibrium chemical denaturation data from CD and fluorescence spectroscopy (F). Equilibrium $m_{D\text{-}N}$-values, mid-point concentration of urea $[urea]_{50\%}$ and the difference in free energy between native and denatured state $\Delta G_{D\text{-}N}$ of the homologous NTDs. (Lh from Heiby 2015, Lg from Rajab 2015, Nc from Rat 2015 and Ea from Ries et al. 2014)

NTD	Signal	$m_{D\text{-}N}$ ($kcal \cdot M^{-1} \cdot mol^{-1}$)	$[urea]_{50\%}$ (M)	$\Delta G_{D\text{-}N}$ ($kcal \cdot mol^{-1}$)
Lh	CD	1.64 ± 0.04	3.09 ± 0.01	5.07 ± 0.14
	F	1.58 ± 0.08	3.20 ± 0.02	5.06 ± 0.29
Lg	CD	1.46 ± 0.05	3.39 ± 0.02	4.95 ± 0.20
	F	1.53 ± 0.06	3.34 ± 0.02	5.10 ± 0.23
Nc	CD	1.59 ± 0.04	2.19 ± 0.02	3.48 ± 0.12
	F	1.67 ± 0.04	2.18 ± 0.01	3.64 ± 0.10
Ea	CD	1.69 ± 0.03	3.30 ± 0.01	5.60 ± 0.12
	F	1.69 ± 0.02	3.31 ± 0.01	5.60 ± 0.08

As demonstrated by two independent spectroscopic methods, the cooperativity of the folding seems to be conserved for the four NTD homologs. Particular small proteins generally tend to fold in a two-state manner (Jackson 1998).

4.1.3. Folding kinetics of NTDs

After resolving how well the protein folds, we analysed how fast this process is. Here, the folding kinetics should be in accordance with the equilibrium folding data in order to confirm the above-described results of cooperativity and the two-state model. Therefore, we performed another set of stopped-flow experiments to probe the folding and unfolding kinetics of the NTD homologs.

Samples were prepared in pH 7 buffer containing either zero or six molar urea, which was rapidly mixed to the higher or lower denaturant concentration, respectively. The intrinsic Trp fluorescence of the NTD was then tracked over time. Similar to the chemical denaturation experiments, the Trp fluorescence increased or decreased depending on the molecular surrounding, getting quenched or fluorescent upon folding or unfolding, respectively (Fig. 4. 1. 6 a). By fitting the transients with a mono-exponential function, the observed rate constants k_{obs} of all applied denaturant concentration were extracted.

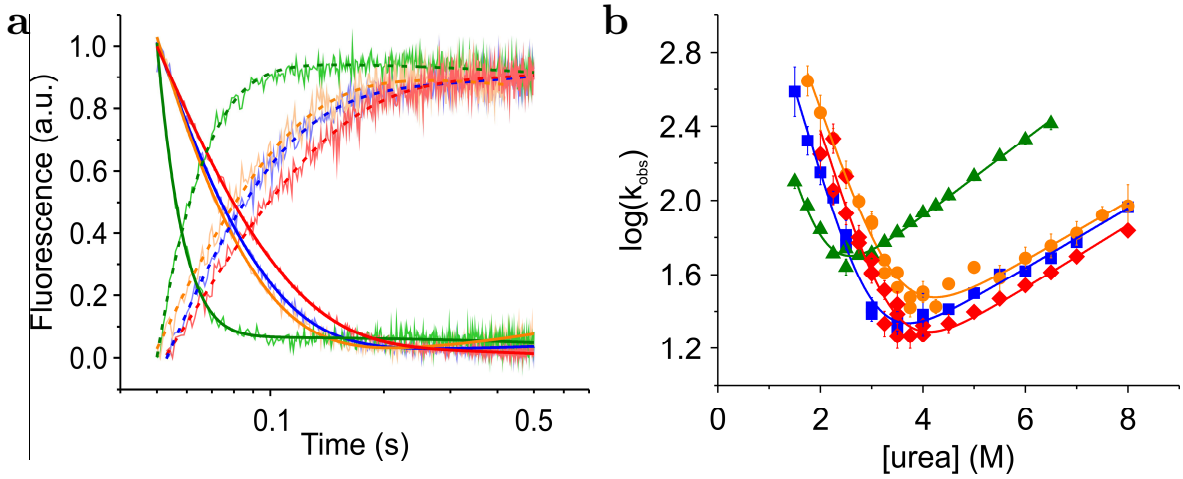


Figure 4. 1. 6: Kinetics of folding and unfolding measured by stopped-flow tryptophan fluorescence spectroscopy. (**a**) Representative single shot transients of all homologs showing folding kinetics (decay curves), which are measured with 3 M urea and unfolding kinetics (rise curves) with 5 M urea. Folding and unfolding data are fitted with mono-exponential functions, in solid and dashed lines, respectively. Data are normalised for clarity. (**b**) Chevron analysis of observed rate constants measured at different urea concentrations. The data is fitted to a model for a barrier-limited two-state transition. Colour code: Lh – blue; Lg – orange; Nc – green; Ea – red. (Heiby et al. 2017)

The fit of the transients contained a linear baseline drift, factoring photobleaching or sample diffusion in the mixing zone of the stopped-flow machine. The logarithmic k_{obs} were plotted against the applied urea concentration resulting in a chevron plot (Fig. 4. 1. 6 b). By applying a barrier-limited two-state transition fit to the chevron plot, the microscopic rate constants of folding k_f and unfolding k_u could be extrapolated (table 4. 1. 4). Folding and unfolding rate constants ranged between k_f = 10 000-13 000 s^{-1} and k_u = 3-5 s^{-1}, for Lh, Lg and Ea. Solely the Nc had a lower folding rate constant with 3 600 s^{-1} and a higher unfolding rate constant with 12 s^{-1} resulting in a lower (equilibrium) free energy ΔG value, compared to the other homologs. Those findings are in good agreement with the results from the equilibrium experiments (table 4. 1. 3 and 4. 1. 4).

Table 4. 1. 4: Folding kinetics of homologous NTDs. Data extracted from chevron plot with rate constant of folding k_f and unfolding k_u (in s^{-1}) extrapolated to zero denaturant, kinetic m-values of folding m_f and unfolding m_u (in kcal · mol^{-1} · M^{-1}), the calculated difference in free energy $\Delta G_{D\text{-}N}$ (kcal · mol^{-1}), and β Tanford value β_T. (Heiby et al. 2017)

NTD	k_f	m_f	k_u	m_u	$\Delta G_{D\text{-}N}$	β_T
Lh	10 123±2 278	1.29±0.06	4.67±0.70	0.22±0.01	4.55±0.16	0.85±0.01
Lg	11 899±3 035	1.09±0.06	5.34±1.16	0.22±0.02	4.56±0.20	0.83±0.01
Nc	3 652±1 328	1.39±0.13	12.16±1.02	0.28±0.01	3.38±0.22	0.83±0.01
Ea	13 257±6 348	1.20±0.11	3.28±1.14	0.23±0.03	4.92±0.35	0.84±0.02

Summing the m-values of the resulting folding and unfolding kinetics revealed that all NTD m-values were in good agreement with the m-values established from the equilibrium data and were further within error of previously published temperature-jump experiments for Ea NTD (m_{eq}^{Ea} = 1.69±0.03 kcal mol^{-1} M^{-1}; m_{kin}^{Ea} = 1.7±0.2 kcal mol^{-1} M^{-1}; Ries et al. 2014). The free energies $\Delta G_{D\text{-}N}$ values were also in good agreement with equilibrium data (see table 4. 1. 3). Therefore, the data support the theory of two-state transition of folding (Jackson and Fersht 1991). For most two-state folders, the β_T value lays between 0.6 and 0.9 (Tanford 1970). According to the calculated β-Tanford values (β_T), the transitional state during the folding process must be structurally close to the native state, as the

values are all high and similar for all NTDs. Additionally, there was no indication for folding intermediates.

4.1.4. Influence of solution ionic strength on folding

In order to analyse the influence of ionic strength on the fold of the NTD, we performed denaturation experiments with CD and fluorescence spectroscopy comparing the stability of the NTDs with increasing salt concentration. For thermal denaturation experiments, salt concentrations ranging from 41 mM to 1 M *Is* adjusted with potassium chloride, in a pH 7 buffer were applied. Salt concentrations below 41 mM could not be prepared due to the inherent ionic strength of the buffer. For fluorescence spectroscopy measurements, the ionic strength was adjusted to 100, 200, 350 and 500 mM (with KCl). The results showed higher stability with increasing ionic strength for all homologs under thermal and chemical denaturation (Fig. 4. 1. 7 a, Fig. 4. 1. 8 a). The thermal denaturation experiments under varying *Is* solution condition were performed by Suhaila Rajab (Lg, Rajab 2015), Charlotte Rat (Nc, Rat 2015) and myself (Ea, Lh from Heiby 2015).

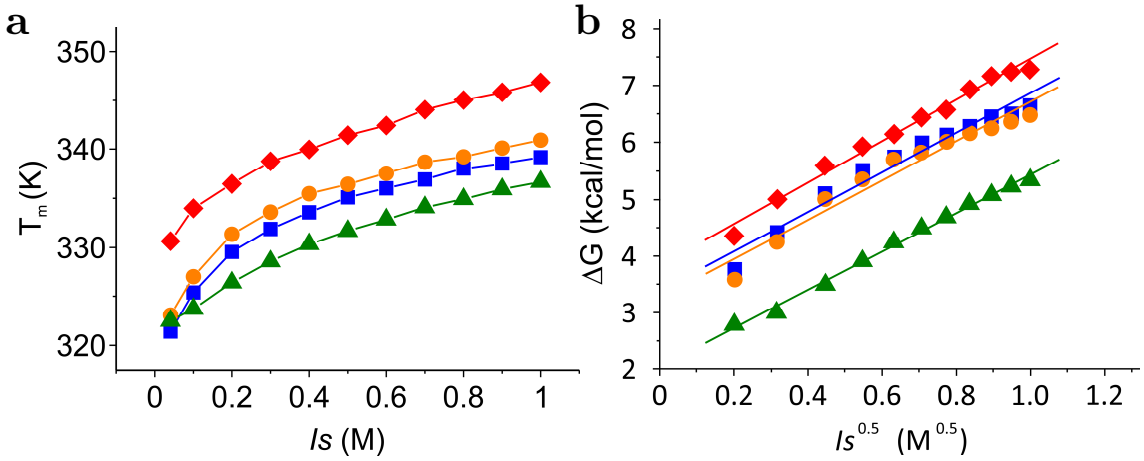


Figure 4. 1. 7: Thermal denaturation of homologous NTDs under the influence of salt. (**a**) Relation between mid-point denaturation temperature T_m and salt concentration of the buffer (*Is*), measured in pH 7 condition with increasing amount of salt by CD spectroscopy at 222 nm. Solid lines connect the data points to guide the eye. (**b**) Difference in free energy ΔG is plotted against the square root of the applied ionic strength. Data points were linearly fitted. Colour code: Lh – blue; Lg – orange; Nc – green; Ea – red. (Heiby et al. 2017)

The observed effect is the result of Debye-Hückel screening of point charges (Debye and Hückel 1923). This assumption can be made due to the linearity of the calculated ΔG plotted against the square root of the ionic strength (Fig. 4. 1. 7 b). The resulting slope of the plot from ΔG against $Is^{0.5}$ corresponds to the equilibrium m-value m_{eq} (de los Rios and Plaxco 2005). The m_{eq} values of the homologs were constant with 3.6±0.2 kcal mol^{-1} $M^{-0.5}$ for Ea, Nc with 3.4±0.1 kcal mol^{-1} $M^{-0.5}$, Lh with 3.5±0.2 kcal mol^{-1} $M^{-0.5}$, and Lg with 3.4±0.3 kcal mol^{-1} $M^{-0.5}$. These findings were validated by chemical denaturation experiments of Ea NTD, measured with different solution *Is*. By plotting the calculated ΔG against $Is^{0.5}$, the resulting m_{eq} was found as 4.0±0.8 kcal mol^{-1} $M^{-0.5}$, being in reasonable agreement with the values obtained from thermal denaturation experiments (Fig. 4. 1. 8 b).

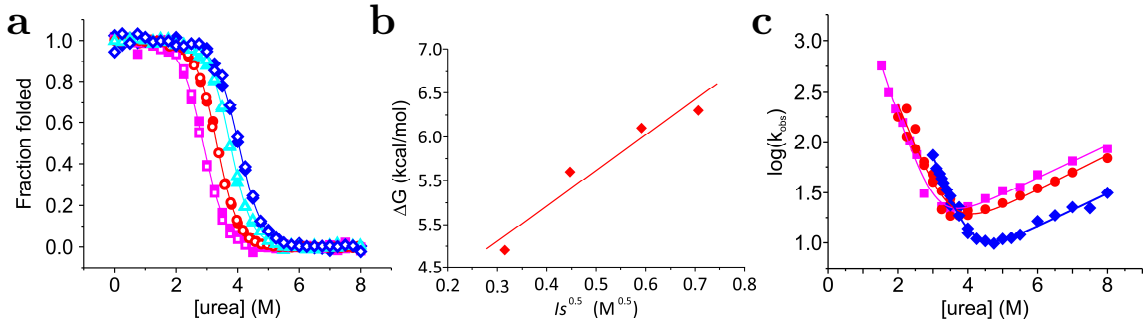


Figure 4. 1. 8: Salt dependency on the stability and fold of the Ea NTD. (**a**) Chemical denaturation measured by CD (closed symbols) and Trp fluorescent spectroscopy (open symbols) of Ea under 100 mM *Is* (magenta), 200 mM *Is* (red), 350 mM *Is* (cyan), and 500 mM *Is* (blue) in a pH 7 buffered solution. Data is normalised to the folded fraction and fitted by a thermodynamic model for a two-state equilibrium. (**b**) The calculated change in free energy ΔG from the chemical denaturation measurements in (a) plotted against the square root of the ionic strength *Is*. The data is extrapolated by a linear fit. (**c**) Chevron analysis of Ea NTD with 100 mM *Is* (magenta), 200 mM *Is* (red), and 500 mM *Is* (blue) under pH 7 buffered solution conditions (350 mM *Is* measurements not included). The observed rate constants are plotted against the applied urea concentration. The data are fitted by a model for a barrier-limited two-state transition. ((a), (b) Heiby et al. 2017)

The m_{eq} values are relatively high when compared to other proteins of the same size. Fyn-SH3, for example, a kinase participating in the lymphocyte-specific signalling pathway, has an observed m_{eq} value of 2.90±0.07 kcal mol^{-1} $M^{-0.5}$ (de los

Rios and Plaxco 2005). The clustered point charges of the NTDs, which are also responsible for the anti-parallel dimerisation, could be the reason for higher m_{eq} values of NTDs since the overall stability is reduced by them. Furthermore, salt can reduce the intermolecular electrostatic strain either by direct binding or through Debye-Hückel screening, shielding coulombic forces, which results in a stabilised fold.

For reasons of completeness, the results from the chemical and thermal denaturations were also compared with the m-values determined by stopped-flow measurements (Fig. 4. 1. 8 c). All results were in good agreement within error margins. The m-values, mid-point transitions $[urea]_{50\%}$ and calculated ΔG values from Ea NTD are listed in table 4. 1. 5.

Table 4. 1. 5: Salt dependent chemical denaturation and chevron analyses data from CD, Trp fluorescence (F) and stopped-flow (SF) spectroscopy of Ea NTD. Equilibrium $m_{D\text{-}N}$-values, mid-point concentration of urea $[urea]_{50\%}$ and the difference in free energy between native and denatured state $\Delta G_{D\text{-}N}$ of Ea NTD.

Ionic strength	Signal	$m_{D\text{-}N}$ ($kcal \cdot M^{-1} \cdot mol^{-1}$)	$[urea]_{50\%}$ (M)	$\Delta G_{D\text{-}N}$ ($kcal \cdot mol^{-1}$)
100mM	CD	1.59 ± 0.11	2.88 ± 0.03	4.54 ± 0.36
	F	1.72 ± 0.06	2.85 ± 0.02	4.91 ± 0.17
	SF	1.64 ± 0.11	3.04 ± 0.07	4.98 ± 0.22
200mM	CD *	1.69 ± 0.03	3.30 ± 0.01	5.58 ± 0.12
	F *	1.69 ± 0.02	3.31 ± 0.01	5.59 ± 0.08
	SF	1.43 ± 0.15	3.44 ± 0.11	4.92 ± 0.35
350mM	CD	1.61 ± 0.06	3.73 ± 0.02	6.01 ± 0.23
	F	1.65 ± 0.05	3.76 ± 0.01	6.21 ± 0.20
500mM	CD	1.53 ± 0.06	4.03 ± 0.02	6.17 ± 0.29
	F	1.55 ± 0.08	4.06 ± 0.03	6.28 ± 0.37
	SF	1.56 ± 0.09	4.04 ± 0.06	6.31 ± 0.27

* 200 mM *Is* CD and Trp fluorescence data from Ries et al. 2014.

The extracted stability mid-point values and free energy of the stopped-flow experiments were in good agreement with the equilibrium measurements, showing

increased stability with augmented salt concentration and having a higher free energy linked to salt.

Interestingly, the folding and unfolding rate constants of Ea NTD changed under varying ionic strength conditions. Due to extended extrapolation of the folding and unfolding arms of the chevron plot, the fit inherits a considerable error. For this reason, the rate constants of folding and unfolding estimated in buffers containing a concentration of 100 and 200 mM ionic strength, were in a similar range. With 500 mM *Is*, the folding rate k_f was enhanced, while the unfolding rate k_u was decreased significantly, showing a strong influence of salt on the folding kinetics (table 4. 1. 6).

Table 4. 1. 6: Salt dependent folding kinetics of Ea NTD. Data measured from chevron plot with rate constant of folding k_f and unfolding k_u (in s^{-1}) extrapolated to zero denaturant and kinetic m-values of folding m_f and unfolding m_u (in $kcal \cdot mol^{-1} \cdot M^{-1}$) at varying ionic strengths.

Ionic strength	k_f	m_f	k_u	m_u
100 mM	21 942 ± 6 555	1.42 ± 0.08	4.85 ± 1.07	0.22 ± 0.02
200 mM *	13 257 ± 6 348	1.20 ± 0.11	3.28 ± 1.14	0.23 ± 0.03
500 mM	54 746 ± 21 879	1.32 ± 0.07	1.29 ± 0.23	0.24 ± 0.02

* 200 mM *Is* data from Heiby et al. 2017.

The kinetics of folding were on the fast sub-millisecond time scale for all homologs, with 200 mM *Is* (table 4. 1. 4 and 4. 1. 6). With the increasing amount of salt concentration the folding kinetics got faster (table 4. 1. 6).

Higher salt concentrations seem to stabilise the monomeric folded protein. However, the dimeric protein is weakened with additional salt due to the intermolecular electrostatics being shielded by Debye-Hückel effects. This phenomenon is also reflected in the spinning duct, where the salt is resorbed from the spinning solution as the dope approaches the end of the duct. Thus, the proteins in monomeric form are stabilised with high salt concentrations. The dimerisation at the end of the spinning duct, where the fibres are formed, is then stabilised by a reduced salt concentration (Askarieh et al. 2010; Hagn et al. 2011).

4.2. Folding kinetics of a spidroin CTD

The C-terminal domain (CTD) is a five-helix bundle, which is covalently linked as a parallel homo-dimer. This configuration is of the utmost importance, as it pre-defines the conformation of the spidroins in the spinning gland. The CTD is responsible for maintaining the solubility of the spidroins under storage conditions (Eisoldt et al. 2011). In the spinning duct, the spidroins are assembled to form a solid fibre, as a result of mechanically- and chemically-induced forces. Like the N-terminal domain, the CTD is highly conserved across spider species, which is likely reflected in a functional role for the spidroins.

Interestingly, the CTD seems to fold along a very particular pathway and belongs to the family of knotted proteins. This is a class of proteins where, until now, little is known about the functional advantages of the knot and how they fold (Lim and Jackson 2015). Together with Charlotte Rat and Hannes Neuweiler, we elucidated the underlying folding mechanism of the CTD in 2018. Most notably, we found that the protein unfolded in multiple steps under denaturing conditions (Rat 2018). For detailed analysis of the folding mechanism of the CTD of spidroin MaSp1 of the nursery web spider *Euprosthenops australis*, far UV CD spectroscopy was combined with the stopped-flow technique to measure its kinetics. The results showed a fast primary folding, paired with a simultaneous association of the monomeric domains. The subsequent complete folding to the native-like dimer was much slower (Rat et al. 2018). Refolding experiments of denatured and reduced CTD material were performed and analysed by C. Rat. Her results showed a bi-exponential time constant, where the first, very fast, phase is most likely concentration-dependently related to the formation of an intermediate I_2 state from the denatured 2 D form (time constants could not be acquired due to detection limitation). The second, slower, exponential decay showed a concentration-independent mono-molecular folding event, reporting the formation of N_2 (native state) from I_2.

To elucidate in more detail the slow folding and unfolding events of the CTD, I performed chevron analysis of chemical denaturation experiments, using far-UV-

CD stopped-flow spectroscopy (Fig. 4. 2. 1). The measurements were conducted under non-reducing conditions, retaining the natural disulphide-bridge of the CTD. The observed rate constants were plotted against the applied urea concentration and analysed with a chevron plot (Fig. 4. 2. 1 b).

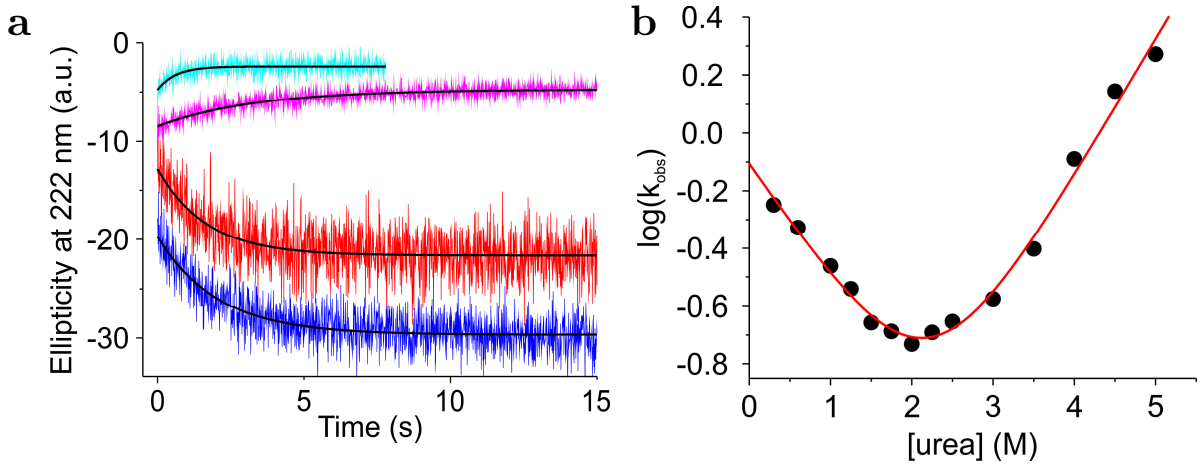


Figure 4. 2. 1: Kinetics of folding and unfolding of N-terminal helices of Ea CTD measured by far-UV-CD stopped-flow spectroscopy. (**a**) Representative transients of single shots, showing folding kinetics (decay curves) which are measured with 0.3 M (red) and 1.0 M (blue) urea, and unfolding kinetics (rise curves) with 2.5 M (magenta) and 4.5 M (cyan) urea. Folding and unfolding data are fitted with mono-exponential functions (black lines). The transients are redistributed along the y-axis for clarity. (**b**) Chevron analysis of observed rate constants measured at different urea concentrations. The data were fitted by a model for a barrier-limited two-state transition (red line). (Rat et al. 2018)

Microscopic rate constants of folding and unfolding were calculated by extrapolation to zero denaturant of the chevron plot, using a kinetic model for a barrier-limited two-state transition and the results are listed in table 4. 2. 1.

Table 4. 2. 1: Folding kinetics of Ea CTD. Kinetic data measured from chevron plot with rate constant of folding k_f and unfolding k_u extrapolated to zero denaturant, kinetic m-values of folding m_f and unfolding m_u, and the calculated difference in free energy $\Delta G_{D\text{-}N}$. (Rat et al. 2018)

Chevron plot analysis	
k_f (s^{-1})	0.78 ± 0.07
m_f ($kcal \cdot mol^{-1} \cdot M^{-1}$)	0.56 ± 0.06
k_u (s^{-1})	9 ± 0.002
m_u ($kcal \cdot mol^{-1} \cdot M^{-1}$)	0.65 ± 0.04
ΔG ($kcal \cdot mol^{-1}$)	2.70 ± 0.20

The sum of kinetic m-values (m_f and m_u) was in good agreement with the m-value obtained from equilibrium denaturation experiments ($m_{N2\text{-}I2}^{eq}$ = 1.2±0.2 kcal mol^{-1} M^{-1} measured by CR; Rat et al. 2018). The free energy calculated from kinetics was also in very good agreement with the determined ΔG from equilibrium chemical denaturation experiments ($\Delta G_{N2\text{-}I2}^{eq}$ = 2.6±0.5 kcal mol^{-1}; Rat et al. 2018). These results confirmed that the slow kinetic phase could be assigned to a barrier-limited two-state folding/unfolding transition between I_2 and N_2.

Figure 4. 2. 2 depicts an overview of the transitions between the folded and the associated phase. The transition between the fully denatured state of monomers (2 D) to the intermediate state (I2) is a fast association paired with a transitional folding of the dimer interface. The final transformation from I2 to the native state (N2) is, in contrast to the first step, surprisingly slow, namely on the time scale of seconds.

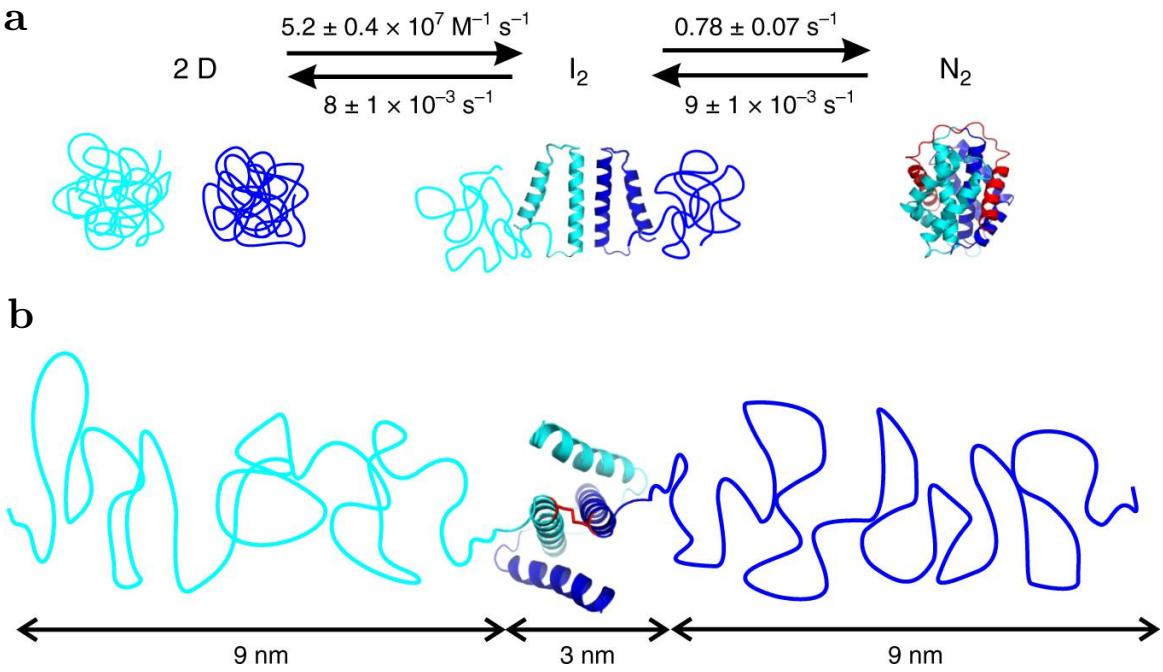


Figure 4. 2. 2: Reaction model depicting time constants of CTD assembly and cartoon diagrams of the different states. (**a**) Detailed kinetics of the CTD by a two-step mechanism of folding and assembly, where the denatured state of the monomers D folds to an intermediate state I_2 and transitions to a native state of the dimer N2. (**b**) Cartoon overview of the I_2 state with indicated chain dimensions. The helices 1-3 remain unfolded while helices 4-5 are already folded and linked by a disulphide bridge (red). (Figure created by HN; Rat et al. 2018)

The fast dimerisation process can be explained by orientational constraint of a pre-folded dimer interface or by long-range electrostatics, which would accelerate protein association (Lim and Jackson 2015). By contrast, the last folding step is slow, probably because the helices clamp and thread themselves into the final CTD form. Because of this molecular clamp, we classified the CTD in the category of knotted proteins (Rat et al. 2018).

In comparison to the CTD, the spidroin NTD associates two orders of magnitude faster ($1.4\text{-}2.6 \cdot 10^9$ M^{-1} s^{-1}; Heiby et al. 2017)(see chapter 4.1.). This fact also clearly reflects the different roles of the domains in the spidroins. The CTD is covalently linked and ready for spidroin polymerisation, while the NTD has to associate promptly for this process. Furthermore, the folding of the NTD is very fast as well, with on average four orders of magnitude greater time constants than for the CTD (0.78 s^{-1} for CTD, 13 000 s^{-1} for NTD). For quality silk, correct folding

of the molecular clamp is necessary. In the ampulla of the spider gland, the dope stays long enough that there is sufficient time for correct folding. However, due to its slower kinetics, the CTD is prone to be unstable, which could also be an advantage in the reformation of the silk at the end of the spinning duct, where the repetitive middle domain rearranges to β-sheet structures (Madurga et al. 2015). Also, spider silk is very elastic when considering its toughness. This elasticity has also been linked to the CTD, as synthesised CTD depleted spidroins often lack extensibility (Gnesa et al. 2012). Therefore, the mechanism demonstrated above could be contributing to the elasticity of the spider silk (Rat et al. 2018).

4.3. Synthesis of miniaturised spidroins

To validate the finding found for the isolated NTDs and CTD in a broader context, it is necessary to transfer the experiments to the full-length spidroins. However, the synthesis of spidroins in heterologous expression systems is still a major challenge. Thus a trade-off solution is needed. The first promising results to synthesise spidroin material implied to reduce the number of repetitive segments to only four and link them to C- and N-terminal domains (Fig. 4. 3. 1). In comparison, the full-length spidroin has up to hundreds of repeats (Rising and Johansson 2015).

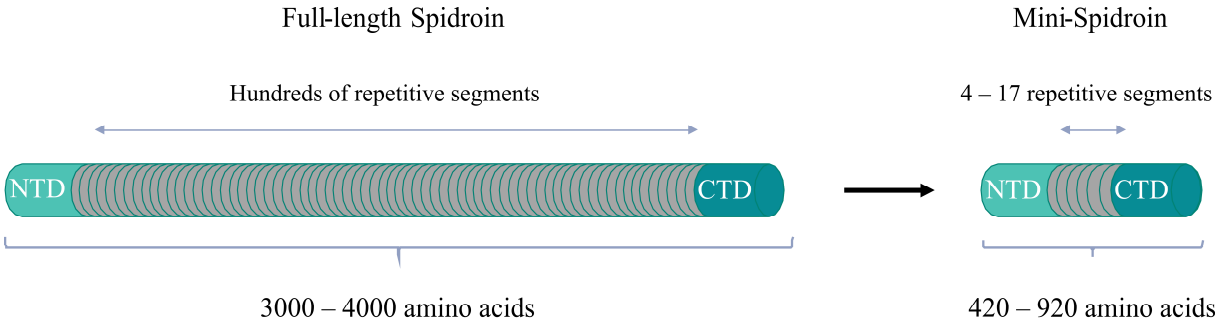


Figure 4. 3. 1: Cartoon overviews of full-length spidroin (left) and shortened mini-spidroin (right), synthesised to overcome expression difficulties. The spidroins contain N- and C-terminal domains (cyan and petrol, respectively), as well as multiple repetitive segments (grey). The length of the full-length and mini-spidroin is given in amino acid number at the bottom as well as in numbers of repetitive segments on top of the cartoon.

Here, I shortened the originally highly repetitive MaSp DNA sequences, while the terminal domains were left unchanged. Therefore, the coding sequences were synthesised by GeneArt. The DNA sequence was then introduced into an expression optimised vector through molecular cloning procedures. The mini-spidroins containing four repetitive units expressed at low yield, but basic functionality tests revealed that the dimerisation of the NTD was still functional, consistent to isolated conditions (Heiby 2015). The tryptophan (Trp) got quenched in a pH 6 solution condition, where Trp is more solvent-exposed as

dimer NTD (29 % quenching for Ea mini-spidroin; 44 % for isolated Ea NTD). Scattering of the signal at 300 nm indicated association (or aggregation) of the proteins. Therefore, the results gained for isolated terminal domains might be assigned to the full-length spidroin as well. Askarieh and colleagues could also show in 2010 that their results on isolated domains correlated with shortened variants of spidroins (Askarieh et al. 2010). Because the experimental data predicted lower strength for the heavily truncated versions of spidroins (Andersson et al. 2017), we then inserted up to 17 repetitive segments. MaSp1 and 2 of the *Latrodectus hesperus* (74.8 and 80.3 kDa respectively) were combined with different fusion-proteins to tune up the expression levels, such as the N-terminal domain of the chaperone protein Hsp90 (28.1 kDa). The latter showed promising preliminary results, but not enough protein material could be purified to do further experiments.

Other fusion-proteins were added to the vector of interest, for example, the Lipoyl and GroEL protein. The spidroin expression was also tested without such fusion proteins, only supplemented by a His-Tag, but then mainly inclusion bodies were harvested after sonification and the subsequent centrifugation (Fig. 4. 3. 2).

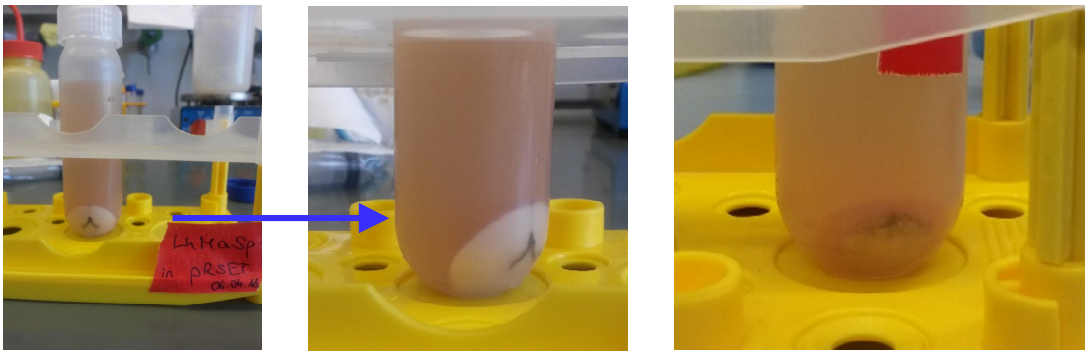


Figure 4. 3. 2: Inclusion bodies after ultra-centrifugation step, showing a white pellet (left and centre image). The right image demonstrates how a regular cell scrap pellet looks like (coloured light to dark brown).

The fusion protein should assist the miniaturised spidroin to stay in solution (Stark et al. 2007) and could later be cleaved of by thrombin recognition. After a potentially successful purification, the proteins could then be fluorescently modified and visualised with optical microscopy techniques, such as laser scanning microscope (LSM) or direct stochastical optical resolution microscopy

(*d*STORM), to elucidate the process of dimerisation. Therefore, two different fluorophores would be attached to the C- and N-terminal domain. By changing the solution pH from 7 to 6 (and below), the direct visualisation of not only the conformation in storage conditions of the spidroins would be possible, but also the dynamic process of polymerisation. Unfortunately, not enough protein material was expressed in order to conduct these experiments. Most of the spidroin dope stayed in inclusion bodies, which after purification could not be tracked on the SDS-PAGE any more. To tune up the mini-spidroin production even more, the co-expression of whole chaperones, like Hsp90, would be interesting. It is known that smaller proteins such as the isolated spider silk NTD or CTD do not rely on chaperone systems to be folded correctly. However, large proteins like the spidroins, and even mini-spidroins with multiple repetitive units, probably need chaperones to achieve a correct fold. The best chaperone system would be the spiders' own chaperones occurring in the spinning gland and duct. A multitude of proteins and chaperones were found to coexist with the spidroins in the spinning duct, believed to be involved in spidroin synthesis and storage, including Hsp70, Hsp60, Hsp20.8 A (Huang et al. 2017). The chosen chaperone should have a different size than the expressed spidroin, to separate them during size-exclusion chromatography.

Furthermore, the future optimisation of the protein expression protocol is essential. There are several published expression protocols reporting the synthesis of miniaturised spidroin-material. In those, one critical point being the down-regulation of their culturing temperature to 303 K instead of the standard 310 K. After the purification step by Ni-NTA, the protein material could be dialysed against 20 mM Tris at pH 8, at 277 K overnight, using a Spectra/Por dialysis membrane with a 6–8 kDa molecular-weight cut-off. Size-exclusion would then be performed with a running buffer containing 20 mM Tris, 150 mM NaCl and 1 mM EDTA at pH 8.0 (for NTD-2Rep-CTD; Andersson et al. 2017). Other protocols suggest to add + 0.4 % v/v glycerol, 17 mM KH_2PO_4, and 72 mM K_2HPO_4 to the cultivation medium (for full-length spidroin split in two with intein linker, without terminal domains; Bowen et al. 2018). In addition to protocol and sequence optimisation, cellular glycyl tRNA levels could be upregulated to meet their higher demand necessary for spidroin overexpression.

With this work as a starting point for future endeavours and existing expression protocols from other labs, the synthesis of miniaturised spidroins seems to be possible. For a systematic spectroscopic study of spidroins, a collaboration with a group already being able to synthesis protein material could also be a promising option. In order to advance towards the full picture of the actual molecular fibre polymerisation, we could directly visualise the fibre formation with the above described spectroscopic tools. This could ultimately lead to a major leap towards the (large-scale) production of synthetic spider silk in the future.

4.4. The role of methionine in the *E. australis* NTD

Spiders have been evolving for 400 million years (Elices et al. 2009), entrusting a highly optimised spider silk fibre. As described in chapter 4.1. and 4.2. of this work, especially the terminal domains are highly conserved between species. Even more intriguing is the unusual amino acid composition of the N-terminal domain compared to most other proteins. A thorough examination of the aligned sequences of different NTDs from various glands and species (Fig. 4. 4. 1), revealed that the domains contain a high amount of methionine residues (Met) and that some of them seem to have conserved sequence positions. Furthermore, also leucine (Leu) and isoleucine (Ile) residues show a similar pattern. But most importantly, it seemed that the sequence positions where Met was not *highly* conserved, we predominantly found leucines and isoleucines. This suggested that in conserved sequence positions, there it is either Met, Leu or Ile.

Methionine residues are known to be rare in proteins, being mostly dispersed through sequences. The total share of Met in proteins lies typically around 2.5 % (Jordan et al. 2005), and even lower in the spidroins of dragline silk with approximately 0.2 - 0.4 %. By contrast, the amount of Met is significantly higher in the NTD of said spidroins with 5 - 7 %. Visualising the sequences of diverse NTD spidroins and their structure with PyMOL (The PyMOL Molecular Graphics System), we deduced that most Mets were located in the core of the protein or were involved in tertiary interactions (e.g. Fig. 4. 4. 3). To our knowledge, this particular arrangement was noticed for the first time for the N-terminal domain of spidroins (Heiby et al. 2019).

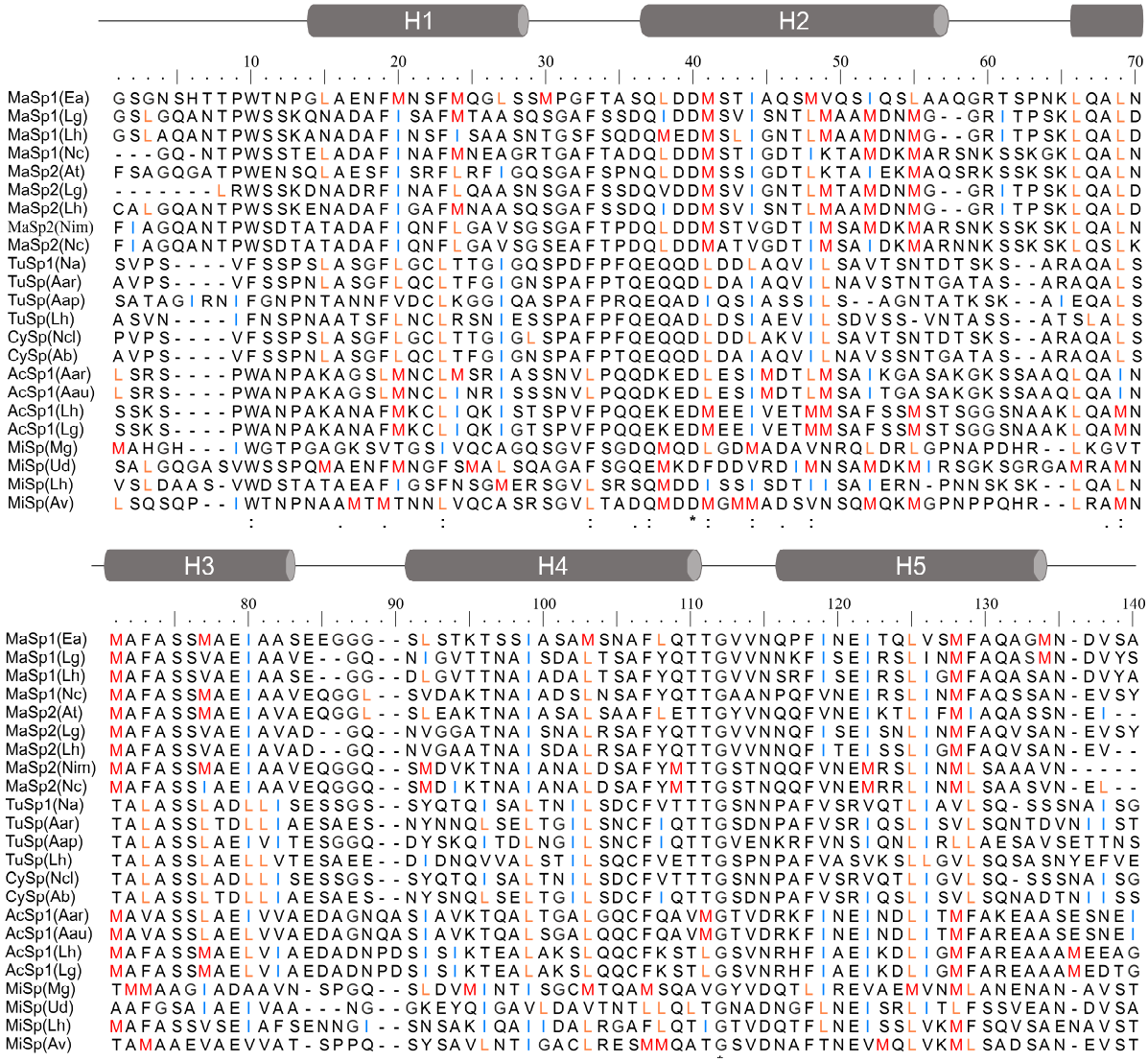


Figure 4. 4. 1: Sequence alignment of N-terminal domains originating from different spiders and glands. Position conservation of the amino acid residues methionine (M), leucine (L) and isoleucine (I) are highlighted in red, orange and blue, respectively. α-helices are indicated by grey bars on top. Amino acid conservation is denoted identical (*), very similar (:) and similar (.) at the bottom, calculated with Clustal Omega. Spider species and gland abbreviations: Ea *Euprosthenops australis*, Lg *Latrodectus geometricus*, Lh *Latrodectus hesperus*, Nc *Nephila clavipes*, At *Argiope trifasciata* and Nim *Nephila inaurata madagascariensis*, Na *Nephila antipodiana*, Aar *Argiope argentata*, Aap *Agelenopsis aperta*, Ncl *Nephila clavata*, Ab *Argiope bruennichi*, Aau *Argiope aurantia*, Mg *Metepeira grandiose*, Ud *Uloborus diversus*, Av *Araneus ventricosus* of the MaSp major ampullate spidroin, CySp and TuSp cylindrical silk/ tubuliform spidroin, AcSp aciniform spidroin, MiSp minor ampullate spidroin. (Heiby et al. 2019)

Dependent on their originating gland, we calculated the exact content of methionine, leucine and isoleucine residues in the NTDs. We could see, that MaSp (major ampullate spidroin), MiSp (minor ampullate spidroin) and AcSp (aciniform spidroin) NTDs have a high content of methionine residues, whereas TuSp (tubuliform spidroin) has none. Furthermore, the fractions of Met, Ile and Leu seem to be conserved independently of the originating gland, with approximately 18 % overall amino acid share. Independent of the Met content, the total content of Met, Leu and Ile residues seems compensated by Leu and Ile residues, as seen for TuSp and CySp spidroin NTDs (Fig. 4. 4. 2).

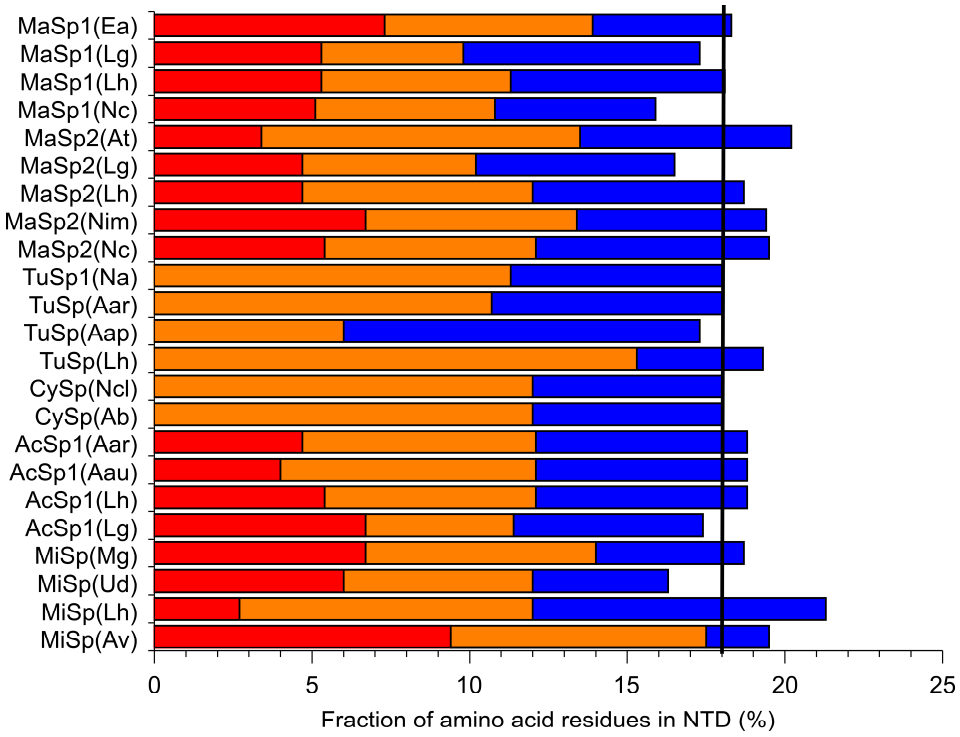


Figure 4. 4. 2: Met, Leu and Ile content in spidroin NTDs. Spider species and glands are denoted consistent with Fig. 4. 4. 1. Met, Leu and Ile frequency in homologues NTDs is shown as a fraction of the total amino acid content. The averaged content of Met, Leu and Ile of 18 % is indicated by a black vertical bar. Colour code: Met – red; Leu – orange; Ile – blue. (Heiby et al. 2019)

Trying to correlate the number of specific amino acid residues with its impact on the physical properties of spider silk seems daring. But strikingly, the only gland shown in Fig. 4. 4. 2 without Met in its NTD sequence, originates from the tubuliform gland and is known to be the least strongest spidroin form (Blackledge and Hayashi 2006).

Although early attempts to understand the role of Met and Leu have been made, mainly in order to regulate oxidation of Met, a decisive breakthrough was missing, until now (Gilles et al. 1988; Cirino et al. 2003). While some reports may have had a similar angle as our study, most of the experimental approaches have been kept simple. Since then, methionine residues have not regained significant interest, as there are not many proteins with as many Mets as the NTD of the MaSp (Fig. 4. 4. 3).

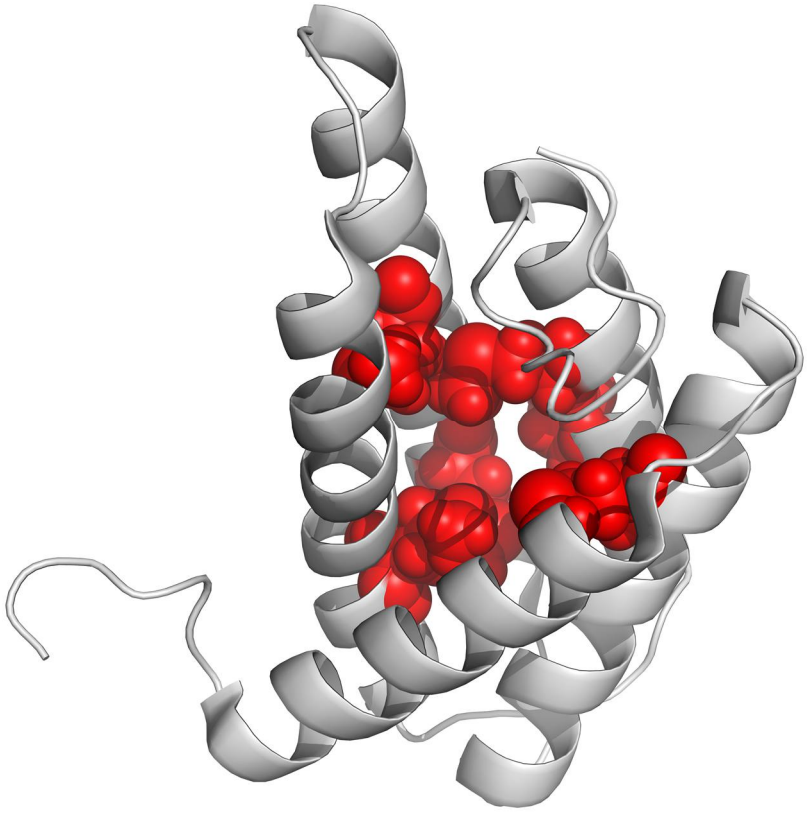


Figure 4. 4. 3: PyMOL cartoon diagram of *Euprosthenops australis* NTD (grey) with highlighted methionine core (red). PDB-code: 2lpj.

Methionine, leucine and isoleucine residues are very similar in structure, essentially diverging only by a sulphur atom. To investigate why this particular set of amino acid residues appear predominantly in the core of the NTD, we exchanged all core Met by Leu by site-directed mutagenesis. Specifically, we generated one to six tuple mutants in accumulating fashion, subsequently named L1, L2, L3, L4, L5 and L6-NTD.

4.4.1. Core methionine – folding at equilibrium state

All generated mutants expressed surprisingly well and seemed to be structurally intact. We examined the secondary structure of the wild type and its mutants by circular dichroism (CD) spectroscopy (Fig. 4. 4. 4). Our results showed a nearly perfect match of all CD spectra, confirming that the structures of all mutants were retained (L6-NTD and WT shown exemplarily in Fig. 4. 4. 4).

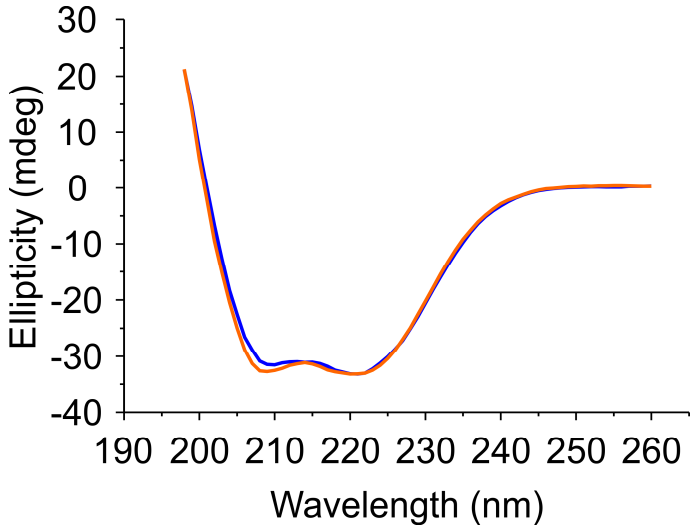


Figure 4. 4. 4: Far-UV CD spectra of Ea WT (blue) and L6-NTD (orange). (Heiby et al. 2019)

With the mutant proteins seemingly intact, we analysed their folding stability and kinetics under equilibrium conditions. First, we probed stability under chemical and thermal denaturation conditions with CD and fluorescence spectroscopy (Fig. 4. 4. 5). To our surprise, the L6-NTD gained dramatically in stability.

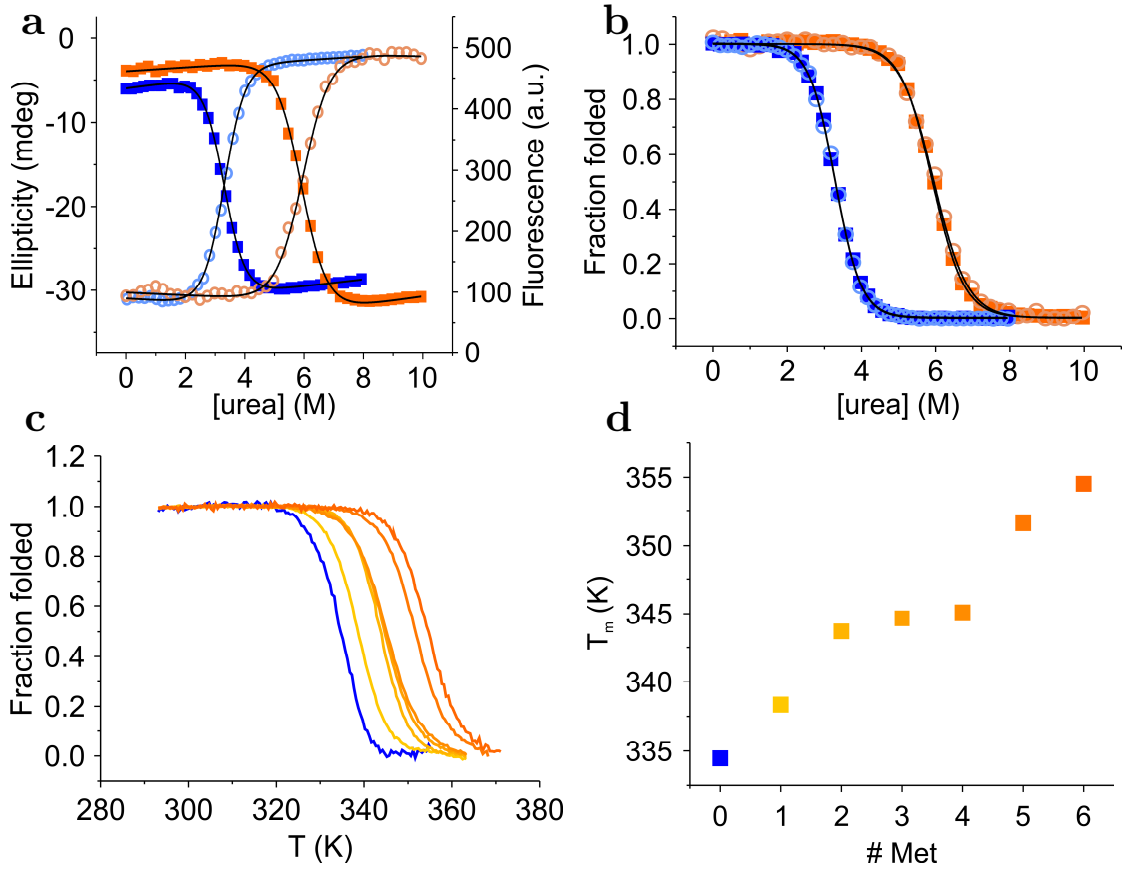


Figure 4. 4. 5: Chemical and thermal denaturation of Ea WT and L6-NTD under equilibrium conditions measured by far-UV CD and fluorescence spectroscopy. (**a**) Titration of NTD in urea in pH 7 buffered solution (200 mM *Is*), measured at 222 nm with CD (light open circles), and tryptophan fluorescence at 325 nM (closed squares). Data points were fitted with a model for a two-state equilibrium (solid lines). (**b**) Normalised data from (a) displayed as a fraction of folded protein. (**c**) Thermal denaturation of Ea WT and cumulative Met-to-Leu mutants, measured with CD under pH 7 (200 mM *Is*) condition at 222 nm. Data were normalised to the fraction of folded protein. (**d**) Transition mid-point melting temperatures T_m dependent on the number of exchanged Met. Colour code: Ea WT NTD – blue; L1 to L6 NTD – shades of orange. (Heiby et al. 2019)

The results showed a gain of 2.2 kcal/mol for chemical and 20 K for thermal folding stability of L6-NTD, compared to the wild type. In contrast to wild type, all Met depleted mutants were denatured reversibly at pH 7 conditions but did not fully regain the initial fold. The more Met were mutated to Leu, the more complete the protein folded back (progressive series). The L6-NTD almost reached the original CD signal upon renaturation. The accumulating series of mutants L1-L6 NTD showed an additive stabilising effect, with every further introduced

leucine (Fig. 4. 4. 5 c, table 4. 4. 1). Thus, not a single mutation was responsible for the strong stabilising effect for L6-NTD, but rather the cumulative mutations.

Table 4. 4. 1: Mid-point transition temperature T_m of thermal denaturation and relative stability gain compared to wild type, ordered by increasing number of Met to Leu mutations. Sequence positions of mutations indicated in brackets, relative to PDB-code sequence 2lpj.

#Met/Leu Exchange	Ea-NTD (mutated seq. position)	T_m (K)	Rel. stability gain, against WT ref. (K)
0	WT	334.5±0.1	--
1	L1 (20)	340.0±0.1	+ 5.5
2	L2 (20,24)	343.7±0.1	+ 9.3
3	L3 (20,24,41)	344.7±0.1	+10.2
4	L4 (20,24,41,48)	345.1±0.1	+10.6
5	L5 (20,24,41,48,77)	351.7±0.1	+17.2
6	L6 (20,24,41,48,77,101)	354.5±0.2	+20.1

Even though every newly introduced leucine had a stabilising effect, L4 seemed to have the smallest impact. The reason could be, that the mutated methionine on position M48L had direct contact to the single tryptophan (Trp) of the NTD. When Met and Trp are in the right distance and orientation, they are stabilised between S-NH groups of the residues (Pal and Chakrabarti 2001). By exchanging M48 with Leu, the Leu caused a stability gain only to be compensated by the loss of the Met-Trp induced stability. The result is only a slight increase in the observed T_m value.

From the denaturation data (table 4. 4. 2; Fig. 4. 4. 5 a - b) we could derive the m-values, indicating the accessible surface area of a protein upon folding (Myers et al. 1995). Our data suggested a compaction of the L6-NTD since the m-value for the mutant was slightly reduced compared to the WT. These findings were underlined by stopped-flow (SF) measurements (Fig. 4. 4. 6) and NMR spectroscopy, which are described below. Additionally, we also observed the clear shift in stability and a distinct change of folding and unfolding rates by analysing the folding kinetics of the L6-NTD mutant, measured by SF spectroscopy (Fig. 4. 4. 6 b). Previous measurements of the Ea wild type showed that the NTD folds

rapidly on the 100 µs time scale through a barrier-limited two-state transition (Heiby et al. 2017).

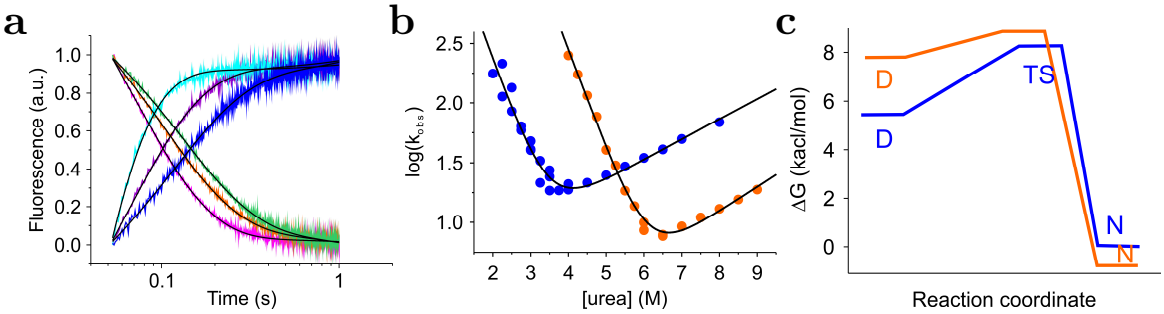


Figure 4. 4. 6: Kinetics of folding and unfolding of Ea WT and L6-NTD measured by stopped-flow tryptophan fluorescence spectroscopy. (**a**) Representative transients of L6-NTD showing folding kinetics (decay curves) measured with 5.0 M (cyan), 5.5 M (violet), 6.0 M (blue) urea and unfolding kinetics (rise curves) with 6.5 M (green), 7.5 M (orange), 8.5 M (magenta) urea. Data points were normalised for clarity and fitted with mono-exponential function (black lines). (**b**) Chevron analysis of observed rate constants measured at different urea concentrations (Ea WT NTD – blue; L6 NTD –orange). Data points were fitted to a model for a barrier-limited two-state transition (black lines). (**c**) Reaction scheme of the free-energy profile during folding. The different folding states of denatured state (D), transition state (TS) and native state (N) are indicated. ΔG of the native state of WT-NTD was set to zero as a reference. Colour code from (b) applies. ((c) created by HN; Heiby et al. 2019)

In the case of the L6-NTD, the extrapolated rate constants k_f and k_u indicated 40 times faster folding and 15 times slower unfolding compared to the wild type (table 4. 4. 2). The calculated free energy of L6-NTD was consistent with the data observed for equilibrium denaturation measurements with CD and fluorescence spectroscopy (table 4. 4. 2). The m-value estimated by stopped-flow spectroscopy was slightly lower than the values measured with other spectroscopic methods, but in reasonable agreement, considering error margins. These relatively large errors are a consequence of the long extrapolation in the chevron plot.

The landscape of the free energy profile of L6-NTD has changed significantly, as depicted in Fig. 4. 4. 6 c. The L6-NTD's energy barrier of unfolding to the transition state is significantly lowered, and the energy barrier between the transition state and native state is higher, compared to the wild type. Hence, the native state is stabilised.

Table 4. 4. 2: Equilibrium chemical denaturation and kinetic data from far-UV CD, fluorescence (F) and stopped-flow (SF) spectroscopy. Equilibrium $m_{D\text{-}N}$ values, mid-point concentration of urea $[urea]_{50\%}$ and the difference in free energy between native and denatured state $\Delta G_{D\text{-}N}$ of Ea WT and L6-NTD. Folding kinetics extracted from chevron plot with rate constants of folding k_f and unfolding k_u, extrapolated to zero denaturant, kinetic m-values of folding m_f and unfolding m_u, and β Tanford value β_T. Errors originate from regression analysis. (Heiby 2019)

	WT-NTD*	L6-NTD
$m_{D\text{-}N}^{CD}$ (kcal mol^{-1} M^{-1})	1.69±0.03	1.30±0.05
$m_{D\text{-}N}^{F}$ (kcal mol^{-1} M^{-1})	1.69±0.02	1.37±0.03
$m_{D\text{-}N}^{SF}$ (kcal mol^{-1} M^{-1})	1.43±0.15	1.43±0.09
$[urea]_{50\%}^{CD}$ (M)	3.30±0.01	5.99±0.02
$[urea]_{50\%}^{F}$ (M)	3.31±0.01	5.95±0.01
$[urea]_{50\%}^{SF}$ (M)	3.44±0.11	6.1±0.1
$\Delta G_{D\text{-}N}^{CD}$ (kcal mol^{-1})	5.6±0.1	7.8±0.3
$\Delta G_{D\text{-}N}^{F}$ (kcal mol^{-1})	5.6±0.1	8.2±0.2
$\Delta G_{D\text{-}N}^{SF}$ (kcal mol^{-1})	4.9±0.4	8.7±0.4
k_f (s^{-1})	13 000±6 000	519 000±213 000
k_u (s^{-1})	3±1	0.20±0.10
m_f (kcal mol^{-1} M^{-1})	1.20±0.10	1.13±0.05
m_u (kcal mol^{-1} M^{-1})	0.23±0.03	0.30±0.04
β_T	0.84±0.02	0.79±0.09

* WT equilibrium thermodynamic quantities from Ries 2014; WT kinetic quantities from Heiby 2017.

According to Kubelka and colleagues, a protein is predicted to fold at the speed of $\tau_f = N/100$ µs, with N being the number of amino acid residues (Kubelka et al. 2004). The NTD has 137 amino acid residues, which results in its theoretical folding speed limit of ~1.4 µs. Our calculations project the L6-NTD close to this theoretical speed limit of folding with 1.9±0.8 µs ($1/k_f$). Together with its relatively small size, the L6-NTD constitutes a highly interesting example for

computational biology, since current computational methods can only address small proteins with up to ~100 residues with folding times < 1 ms. The faster the folding and smaller in size the molecule is, the better it is for molecular simulations (Gelman and Gruebele 2014). In turn, most naturally occurring proteins of interest are bigger, making them inapt to serve as a basis for computational models.

The drastic increase in folding speed of the L6-NTD could either be explained by structural changes or the interaction of the protein in the transition state between unfolded / folded protein. The respective NMR spectra showed that the structure of L6-NTD did not change, compared to the wild type (further below, Fig. 4. 4. 13). Therefore, the Met, which tends to be more adhesive than other amino acid residues at van der Waals contact, could have stronger interactions in the denatured state, which in turn would reduce its free energy. By mutating Met to Leu, these interactions are then weakened, reducing the free energy difference between unfolded and transitional state, which results in an increased folding speed rate. In addition, Leu showed the ability to stabilise the folded protein and to decrease the unfolding speed. We have also observed this tendency in the thermal and chemical denaturation experiments. Since the amino acid side chain of Leu is branched, it could potentially contribute to the stabilising effect by enhanced van der Waals contact to the surrounding amino acids, in comparison to Met.

Despite these findings, it was still unclear how other amino acids would contribute in the hydrophobic core. Met, Leu and Iso often occurred conserved at the same position, but in some homologous sequences, also valine, alanine and phenylalanine residues could be found. The amino acid residues leucine, isoleucine and valine are highly hydrophobic with 3.8, 4.5, 4.2 on the Kyte and Doolittle scale, respectively, compared to Met with 1.9 (the hydropathy scale spans from -4.5 being strongly hydrophilic to +4.5 for strong hydrophobic amino acid residues, Kyte and Doolittle 1982). Alanine and phenylalanine residues are also hydrophobic with 1.8 and 2.8, respectively. These amino acids are typically found in the core of proteins. The common approach to analyse the function of the different amino acid residues, is deletion mutation, i.e. exchanging a particular

residue by alanine. However, this mutation often caused poor or a lack of protein expression.

Since the mutation to M77L showed the strongest effect regarding the gain in folding stability (+7.2 K; compare table 4. 4. 5), together with Sarah Buchmann, we point-mutated this particular position in the wild-type NTD also to alanine, isoleucine, phenylalanine and valine, within the scope of her bachelor thesis, supervised by myself. As expected, no protein could be synthesised for M77A. Valine, on the other hand, showed no change in stability compared to the wild type. The exchange to phenylalanine yielded low concentration in protein material and showed destabilised T_m value (− 6.6 K), as could be expected from an amino acid of this size (Buchmann 2018). But the most interesting question remained, namely if the sterically close isoleucine, would cause different results than leucine. Thermal denaturation experiments showed that the stabilising effect was similar, albeit slightly weaker, to the one generated by leucine (+ 5 K). Therefore, one could consider them to have similar roles in NTD proteins.

4.4.2. Core methionine – association of the NTD

To probe the association of the L6-NTD, we measured fluorescence spectra under monomeric and dimeric conditions. Under monomeric conditions (pH 7) the tryptophan (Trp) is buried in core position, leading to a strong fluorescent signal, while under dimeric conditions (pH 6), the Trp gets solvent-exposed and therefore the signal quenched and red-shifted. Accordingly, the wild-type's Trp got quenched by 40 % and the maximal emission wavelength shifted from 340 to 329 nm, showing an evident change of the Trp surrounding conditions. For the L6-NTD, however, the signature got weakened with only 20 % quenching and no characteristic red-shift was observable (Fig. 4. 4. 7 a). The analysis of the cumulative mutation series showed that the loss of the quenching effect is gradually enforced by the number of exchanged Met (Fig. 4. 4. 7 b). Also, the wavelength shift disappears with increasing number of Leu.

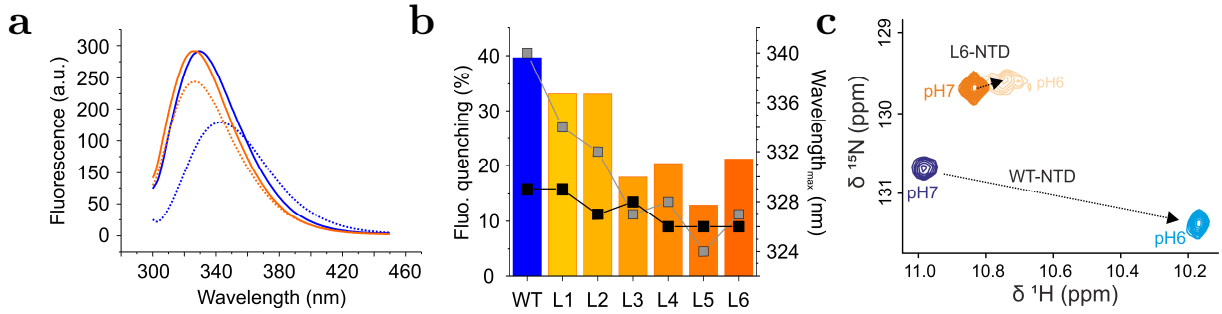


Figure 4. 4. 7: Monomeric and dimeric characterisation of Ea WT and L6-NTD. (**a**) Tryptophan fluorescence spectra (excitation at 280 nm) of WT (blue) and L6-NTD (orange) at pH 7 (solid lines) and pH 6 (dashed lines). Data were normalised to the fluorescence maximum of the wild type. (**b**) Left axis: quenching of Trp fluorescence of WT (blue) and cumulative mutation series L1 to L6-NTD (shades of orange) upon change of buffer condition. Right axis: wavelengths of maximum fluorescence emission at pH 7 (grey squares) and pH 6 (black squares) for each mutant. (**c**) ^{1}H, ^{15}N-HSQC NMR spectra of the amide side chain signal from tryptophan W10 of WT-NTD and L6-NTD at pH 7 (blue and orange, respectively) and pH 6 (light blue and light orange, respectively). ((c) measured and created by BG; Heiby et al. 2019)

The slight increase in fluorescence of M48L (L4) in Fig. 4. 4. 7 b can be explained by the photophysical interactions of Met and Trp. The Met at this position is in direct contact with Trp, which is known to lead to the quenching of Trp (Yuan et al. 1998). The replacement of Met by Leu thus leads to a rise in fluorescence. Interestingly, the fluorescence raw data showed that the Trp of L6-NTD is more fluorescent in pH 7 conditions than the WT's Trp. Still, the relative decrease in fluorescence, as shown in Fig. 4. 4. 7 b, is valid, as the quenching is defined by the relation of intensities at pH 7 and pH 6. Furthermore, these results were underlined by NMR spectroscopy, which clearly showed that the L6-NTD's Trp lost its mobility to undergo conformational change and, therefore, move to the outside of the protein core. The fact that the Trp remained buried in the NTD hinted towards a rigidified core, as detailed further below.

We employed size-exclusion chromatography in combination with multi-angle light scattering spectroscopy (SEC-MALS) to further characterise the association quantitatively. As expected, WT and L6-NTD eluted as a monomer at pH 7

(200 mM *Is*), with a molecular weight (MW) of 14 kDa. At pH 6 (with 60 mM *Is*), the NTD is expected to be dimeric and should, therefore, have a molecular weight of approximately 28 kDa, as did the WT-NTD (~ 26 kDa). However, we determined the average molecular mass of L6-NTD with SEC-MALS to be approximately 20 kDa, i.e. between a monomer and a dimer (Fig. 4. 4. 8 a). We realised that the *Is* of the used buffer was sufficient to weaken the dimer of L6-NTD so that an intermediate MW was measured. Ions in solution can shield electrostatics of a protein and therefore induce dissociation of the NTD (Gaines et al. 2010; Hagn et al. 2011; Heiby et al. 2017). The reduction of *Is* to 8 mM led to a MW for L6-NTD closer to the expected value. However, these experiments already indicated that the equilibrium dissociation constant for L6-NTD had to be increased compared to WT. Augmenting the protein concentration of L6-NTD showed that the MW increased in SEC-MALS experiments (Fig. 4. 4. 8 b), but due to sensitivity limitations of the setup, no K_D value could be measured. Therefore, we performed size-exclusion chromatography (SEC) at high resolution in combination with high-performance liquid chromatography (HPCL) and fluorescence detection, which achieved sub-µM sensitivity. Serial dilutions of NTD wild-type and L1 to L6 mutants were eluted from the Superdex75 column in pH 6 buffer containing only 8 mM *Is*, to probe the association capacity (Fig. 4. 4. 8 c-d, Fig. 4. 4. 9). Hence, the dimer dissociated into monomers with decreasing protein concentration (Fig. 4. 4. 8 c). Plotting the elution volume V_E against the protein concentration allowed the determination of the K_D value (equilibrium dissociation constant) at the mid-point equilibrium state. The monomer of L6-NTD eluted at approximately 10.7 ml, with a MW of 14 kDa, and the dimer at 12.5 ml with presumably doubled molecular weight (Fig. 4. 4. 8 d, in orange spheres). The wild type eluted as a dimer at 11.6 ml and could not be analysed under monomeric conditions, as the resolution limit was reached before the dimer dissociated into monomers (Fig. 4. 4. 8 b, in blue squares). Nevertheless, we could observe the onset of this process at low protein concentration (< 10 nM). The K_D value of the wild-type NTD was previously determined by stopped-flow measurements, as described in chapter 4. 1., to be in the low nM range (Heiby et al. 2017). The diverging elution volumes of both proteins under dimeric conditions further indicated a loosened dimer of L6-NTD, resulting in a mutant overall bigger.

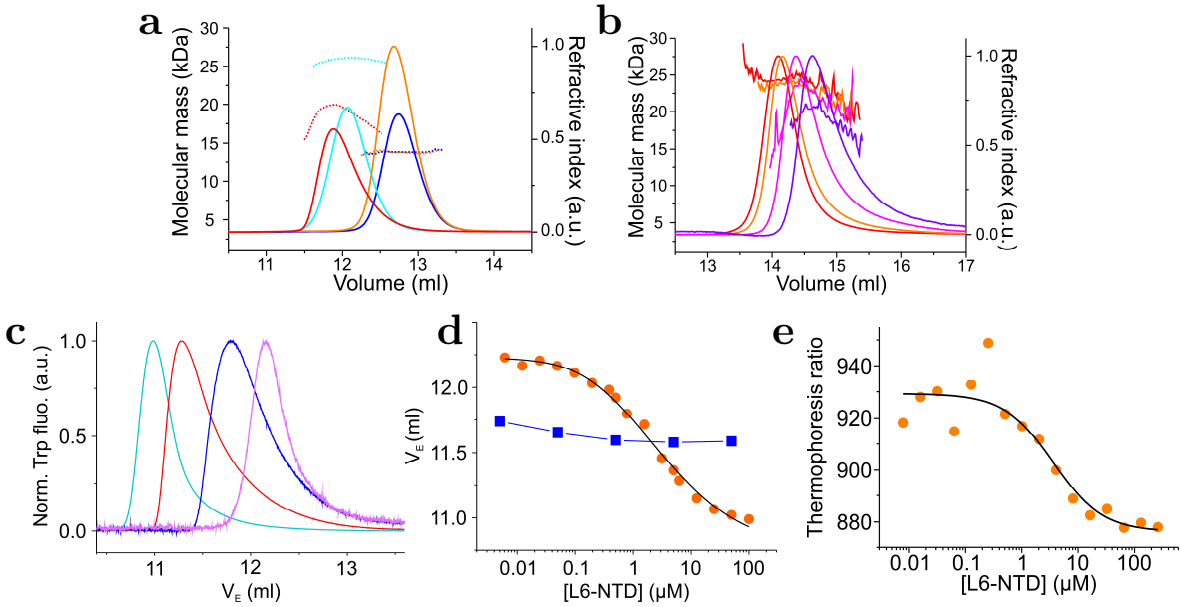


Figure 4. 4. 8: Dimerisation of Ea WT and L6-NTD. (**a**) SEC-MALS chromatograms (solid lines) of monomeric (pH 7, 200 mM *Is*, blue and orange) and dimeric (pH 6, 60 mM *Is*, red and cyan) Ea WT and L6-NTD, respectively, with calculated molecular mass indicated by horizontal, dashed lines. (**b**) SEC-MALS elution bands of L6-NTD with different protein concentration (12 µM in red, 4 µM in orange, 1.2 µM in magenta and 0.4 µM in violet) and calculated molecular mass indicated by horizontal lines, measured at pH 6 (8 mM *Is*). (**c**) Representative, normalised elution bands of L6-NTD detected by Trp fluorescence with high-resolution SEC with 100 µM (cyan), 6 µM (red), 0.8 µM (blue), and 0.05 µM (magenta) protein concentration at pH 6 conditions. (**d**) Elution volume measured by using various protein concentrations with L6-NTD (orange) and late onset dissociation of Ea WT (blue) at pH 6. Data points were fitted by a thermodynamic model for dimerisation equilibrium (black line). In the case of Ea WT, lines connecting the data points are guide to the eye. (**e**) Thermophoresis of L6-NTD, labelled with AttoOxa11, under pH 6 conditions, with titration of unlabelled L6-NTD. Data points were fitted with a binding isotherm model (black line). ((a-b) measured and created by CJ; Heiby et al. 2019)

To validate these results, we performed thermophoresis with L6-NTD (Fig. 4. 4. 8 e). In principle, labelled, dimeric proteins should dissociate and re-associate with unlabelled protein through competitive displacement under equilibrium conditions, leading to a fluorescence decrease. To serve this purpose, we labelled cysteine modified L6-Q50C-NTD at the dimerisation interface with AttoOxa11. By the gradual addition of unlabelled protein material, we could

measure a significant decrease in fluorescence. The resulting K_D values, both from thermophoresis (3.6±1.8 µM, Fig. 4. 4. 8 e) and high-resolution SEC measurements (1.1±0.2 µM, Fig. 4. 4. 8 d), were consistently three orders of magnitude higher than the wild-type K_D value (low nM range). In comparison, the wild-type protein had an equilibrium dissociation rate constant below the detection limit of these methods. Nevertheless, the detectable onset of dissociation of the wild type shown in Fig. 4. 4. 8 d indicated a stronger binding.

Additionally, we analysed the cumulative series of L1-L5-NTD mutants and could show that the effect of a weakened K_D value increased by the number of introduced mutation (Fig. 4. 4. 9).

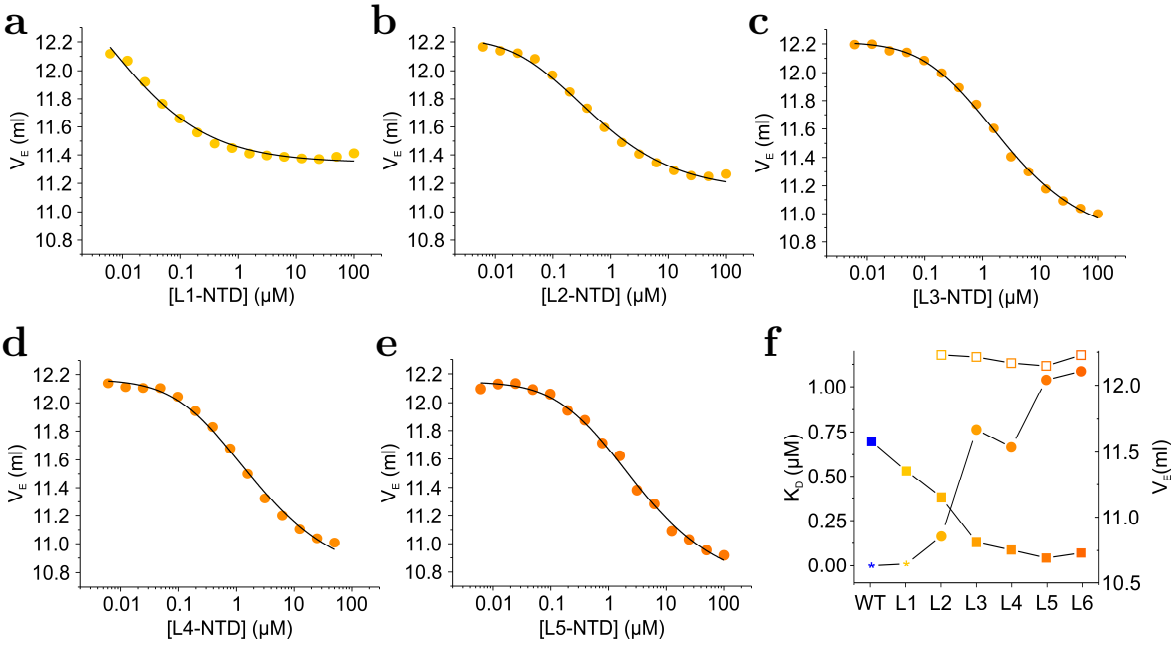


Figure 4. 4. 9: Dimerisation of cumulative L6-NTD mutation series using high-resolution SEC. (**a-e**) Elution volumes of increasing protein concentrations in L1 to L5-NTD mutant. Data points were fitted by a model for equilibrium dimerisation (black lines). (**f**) Calculated K_D values (from a-e) are plotted by mutation (closed spheres) from Ea WT (blue) to L6-NTD (shades of orange). * indicates WT K_D determined from stopped-flow experiments and for L1 estimated K_D (a) as both values are lower than the detection limit. Elution volume of monomer (open squares) and dimer (closed squares) changing with the number of mutations. WT and L1 monomer elution volume could not be determined due to the signal detection limit. Solid lines are guide to the eye. (Heiby et al. 2019)

The elution volume of the dimeric mutants at high concentration (100 µM) shifted systematically from the wild type to L6-NTD dimer elution volume, indicating gradually loosened dimers with every extra introduced Leu. The monomeric V_E was constant for L2 to L6-NTD, but could not be quantified reliably for wild type and L1-NTD (marked with * in Fig. 4. 4. 9 f). The K_D value gradually grew with every introduced leucine residue, with L4-NTD being an exception, which showed a similar K_D value as L3-NTD within error margins (Fig. 4. 4. 9, table 4. 4. 3).

Table 4. 4. 3: K_D values measured by high-resolution SEC.

NTD	K_D value
WT*	1.1±0.1 nM
L1	14.0±2.0 nM
L2	0.16±0.03 µM
L3	0.76±0.07 µM
L4	0.66±0.09 µM
L5	1.04±0.13 µM
L6	1.09±0.16 µM

*WT measured by stopped-flow spectroscopy. (Heiby et al. 2017)

As listed in table 4. 4. 3, all K_D values from L3 onwards are in a similar range of 0.7 - 1 µM. It might be, that the highest attenuation of the K_D value has been reached with L3 and the remaining interaction (of 1µM) is sustained by electrostatic forces (Gaines et al. 2010; Hagn et al. 2010).

The data set presented in this chapter demonstrates that the dimerisation of the methionine depleted NTD core in mutants was impaired gradually and substantially. Therefore, we deduced that methionine residues play a crucial role in the malleability of the NTD, which is needed for the tight association of the dimer.

4.4.3. Interface methionine

To complete the picture from the previous chapters, we then mutated the methionine residues located on the surface of the NTD. More specifically, the three mutated Met were situated on the dimeric interface of the NTD (this construct is subsequently called surface-L3-NTD).

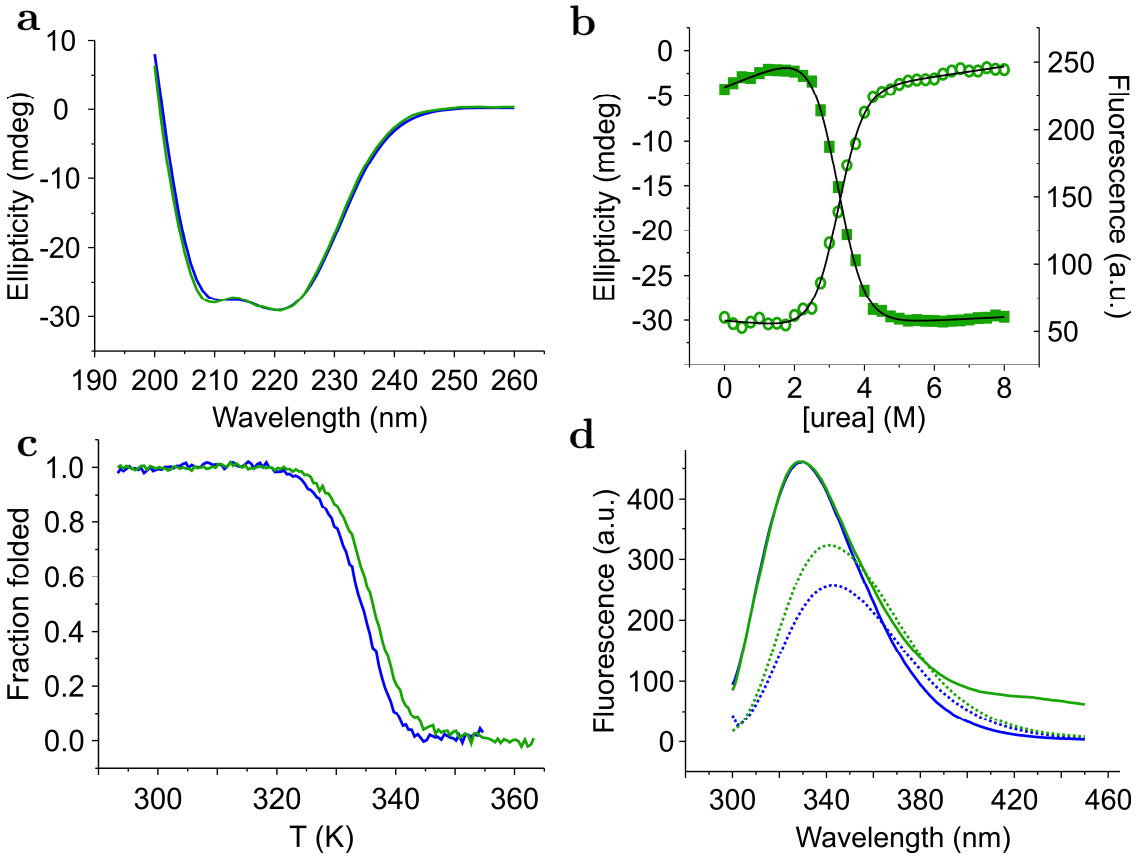


Figure 4. 4. 10: Secondary structure, equilibrium denaturation and conformational change of surface-L3-NTD. (**a**) Far-UV CD spectra of Ea WT (blue) and surface-L3-NTD (green). (**b**) Chemical denaturation of surface-L3-NTD. Titration of NTD in urea in pH 7 buffered solution (200 mM *Is*) and measured at 222 nm with CD (light open circles), and tryptophan fluorescence at 325 nm (closed squares). Data points were fitted with a model for a two-state equilibrium (solid lines). (**c**) Thermal denaturation of Ea WT and surface-L3-NTD with CD under pH 7 (200 mM *Is*) condition at 222 nm. Data were normalised to the fraction of folded protein. (**d**) Tryptophan fluorescence spectra of WT (blue) and surface-L3-NTD (green) at pH 7 (solid lines) and pH 6 (dashed lines). Data were normalised to the fluorescence maximum of the wild type. (Heiby 2015)

Our results showed no change in the secondary structure measured by CD spectroscopy, consistent with the L6-NTD, but a slight destabilising effect of −0.3 - 0.6 kcal/mol in chemical denaturation experiments (CD and F), compared to the wild type (Fig. 4. 4. 10, table 4. 4. 4). On the other hand, thermal denaturation showed a minor stabilising effect with + 1.5 K and was irreversibly unfolded at pH 7.0 conditions. To further investigate the dimerisation process, we measured tryptophan fluorescence spectra under pH 7 and 6 conditions (Fig. 4. 4. 10 d). By modulating the interface of the proteins dimerisation binding site, we expected weakened binding and therefore no change in the Trp fluorescence signal. The results showed less quenching between pH 6 and 7 (30 %) in comparison to the WT (40 %), signifying that the Trp did not get as solvent-exposed in the surface-L3-NTD as in the wild type under dimeric conditions (pH 6). However, the characteristic spectral redshift was still preserved.

Table 4. 4. 4: Equilibrium denaturation characterised by far-UV CD and fluorescence (F) spectroscopy. Equilibrium $m_{D\text{-}N}$-values, mid-point concentration of urea $[urea]_{50\%}$ and the difference in free energy between native and denatured state $\Delta G_{D\text{-}N}$ of Ea WT and surface-L3-NTD from chemical denaturation. Mid-point transition temperature T_m of thermal denaturation for wild type and surface-L3-NTD. (Heiby 2015)

	WT-NTD*	Surface-L3-NTD
$m_{D\text{-}N}^{CD}$ (kcal mol^{-1} M^{-1})	1.69±0.03	1.62±0.07
$m_{D\text{-}N}^{F}$ (kcal mol^{-1} M^{-1})	1.69±0.02	1.56±0.04
$[urea]_{50\%}^{CD}$ (M)	3.30±0.01	3.24±0.02
$[urea]_{50\%}^{F}$ (M)	3.31±0.01	3.23±0.01
$\Delta G_{D\text{-}N}^{CD}$ (kcal mol^{-1})	5.6±0.1	5.3±0.3
$\Delta G_{D\text{-}N}^{F}$ (kcal mol^{-1})	5.6±0.1	5.0±0.1
T_m (K)	334.5±0.1	336.0±0.1

*Wild type equilibrium chemical denaturation quantities from Ries et al. 2014.

These results imply that the impact methionine has on the protein properties is considerably stronger, if the amino acids are in core positions or tertiary

interactions with the core. Hence, the leucine residues in the core positions stabilise, while leucine residues on the surface of the dimerisation interface seem to rigidify the protein and thereby potentially impeding the association slightly. In turn, methionine can shift the proteins' surface ductility and can, therefore, facilitate binding (Gellman 1991; Bernstein et al. 1989; O'Neil and DeGrado 1990).

4.4.4. Position dependency of the Met-to-Leu exchange

To examine the position-dependent effects of the Met-to-Leu exchange in the core, we analysed single-point mutants with the above described methods. In terms of stability, we observed a distance-dependent effect, where the mutation positions near the core centre of the protein exhibited a more significant impact on stability gain, then ones being farther away (Fig. 4. 4. 11).

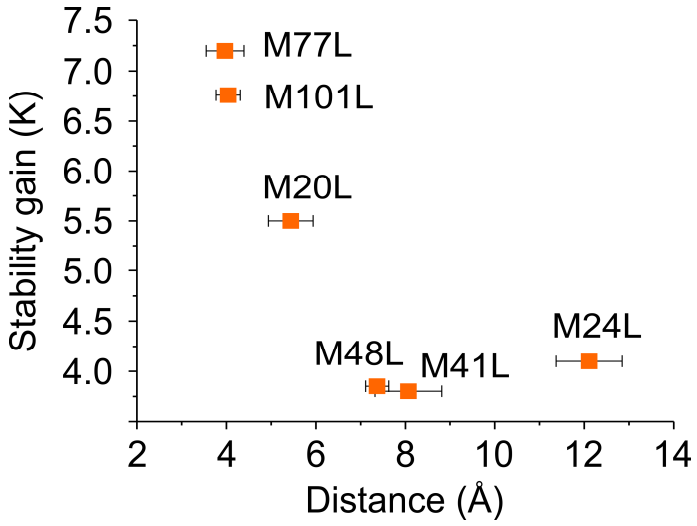


Figure 4. 4. 11: Distance dependency of the amino acid residue in the hydrophobic core in regard to the stabilising effect of the mutant. Distances are measured by PyMOL from structure Ea WT PDB-code 2lpj. Errors originate from the standard deviation of 20 deposited structures. The relative stability gain in thermal denaturation experiments is defined by the difference between WT and the single point mutants L1-NTDs.

Consequently, the single-point mutant M77L had the greatest effect on protein stability with a gain of 7.2 K and appeared to be nearest to the core centre (< 4 Å; determined by PyMOL, centre of mass). Strikingly, the stabilising effect of the Met-to-Leu exchange stagnates when exceeding a certain distance from the

protein core (> 6 Å). Single point mutants located on the interface of the NTD are not shown, as their effect was negligible in comparison.

Together with the K_D values of the single point mutants measured by high-resolution SEC, these findings demonstrate the significance of the Met's position relative to the protein core for the stabilising effect from the Met to Leu exchange. Consequently, the exchange of the amino acid farthest from the core (> 12 Å) at position 24, results only in a minor effect on the dimerisation ability of the NTD. The other measured L1 mutants yielded a similar K_D value in the low nM range (table 4. 4. 5). The K_D values of M24L, M41L and M77L should be considered semi-quantitative, as only 6 data points were collected with high-resolution SEC.

Table 4. 4. 5: K_D values obtained from dimerisation experiments with high-resolution SEC, T_m thermal denaturation mid-points measured by CD spectroscopy and the relative stability gain of the L1 series in relation to wild-type NTD.

Ea-NTD	K_D value (nM) (number# of data points)		T_m (K)	Stability gain (K) (compared to WT)
WT	*1.1±0.1		334.5±0.10	--
M20L	14±2	(#15)	340.0±0.04	+ 5.5
M24L	4±2	(#6)	338.6±0.05	+ 4.1
M41L	67±17	(#6)	338.3±0.04	+ 3.8
M48L	---		338.3±0.06	+ 3.8
M77L	30±16	(#6)	341.7±0.02	+ 7.2
M101L	---		341.2±0.10	+ 6.7

* Ea WT K_D value from stopped-flow measurements. (Heiby et al. 2017)

Overall, the four analysed L1 mutants showed a consistent stabilising effect regarding melting temperatures T_m and comparable reduction of their K_D value. Interestingly, the sum of the stability gain from the individual point mutations exceeds the additive gain in the L6-NTD. Furthermore, the closer the position of the mutation was to the core of the protein, thereby generating more interactions with surrounding amino acid residues, the bigger the impact was generated on the protein folding properties.

4.4.5. Dynamics of the NTD

To assess the dynamics of the L6-NTD mutant, we applied fluorescence correlation spectroscopy, utilising photoinduced electron transfer (PET-FCS) (Fig. 4. 4. 12). To that end, we labelled the cysteine modified wild-type WT-G3C-NTD and L6-G3C-NTD with AttoOxa11, by maleimide derivate conjugation.

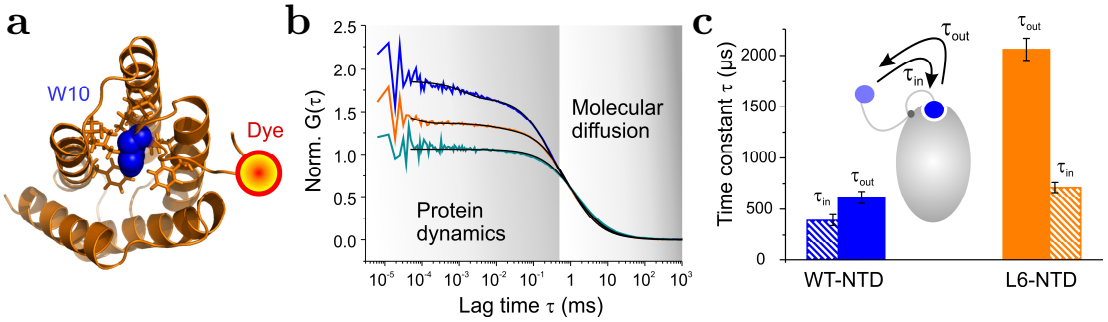


Figure 4. 4. 12: Dynamics of Ea WT and L6-NTD probed by PET-FCS (**a**) NTD cartoon depicting the reporter design with tryptophan W10 (blue) in the centre of the protein and the labelling position (G3C; red-yellow sphere). (**b**) PET-FCS kinetics with normalised G(**τ**) autocorrelation functions from Ea-G3C-WT (blue), L6-G3C-NTD (orange) and reference molecule of L6-Q50C-NTD (cyan). All proteins were labelled with AttoOxa11 dye and measured at a concentration of 1 nM in pH 7 buffer. Data points were fitted by a model for diffusion of a globule containing two single-exponential relaxations, and the control sample was fitted by a model for diffusion only (black lines). (**c**) Calculated time constants $\boldsymbol{\tau}_{in}$ (patterned bars) and $\boldsymbol{\tau}_{out}$ (solid bars) from autocorrelation data for Ea WT (blue) and L6-NTD (orange), of Trp movements in and out of the protein core. Error bars are propagated s.e. from regression analysis. Cartoon portrays Trp W10 (blue) movements in the NTD (grey ellipsoid). (Heiby et al. 2019)

By monitoring the Trp fluctuation in and out of the hydrophobic core of the protein, we could confirm the increased rigidity of the L6-NTD, observed by equilibrium measurements (decreased m-values for mutant protein, see table 4. 4. 2) and Trp fluorescence quenching measurements upon dimerisation (see Fig. 4. 4. 7). Fig. 4. 4. 12 a schematically visualises the distance between the fluorescent dye and intrinsic Trp. When moving out of the NTD core to a more solvent-exposed position, the Trp quenches the dye upon contact via PET. If the Trp is buried, no interaction takes place. The observed PET fluctuation was strong and had several additional decays in the wild type FCS autocorrelation

function (blue curve in Fig. 4. 4. 12 b, table 4. 4. 6). The main decay originated from the free diffusion through the detection focus, which was on the millisecond time scale (τ_{Dif}). At equilibrium state, the Trp toggles in and out of the core. Therefore, two additional decays are caused by PET fluorescence fluctuation (τ_{R1} and τ_{R2}). These conformational changes occur on the sub-ms time scale. The relaxation time constants increased for the L6-NTD, compared to the WT, while the amplitudes a_1 and a_2 were reduced in the mutant.

Table 4. 4. 6: PET-FCS diffusion time, amplitudes and relaxation time constants of Ea WT and L6-NTD. (Heiby et al. 2019)

Construct	τ_{Dif} (ms)	a_1	τ_{R1} (µs)	a_2	τ_{R2} (µs)
WT-NTD	1.43 ± 0.08	0.64 ± 0.03	240 ± 17	0.19 ± 0.04	2.00 ± 1.50
L6-NTD	1.37 ± 0.03	0.35 ± 0.01	529 ± 19	0.05 ± 0.01	3.50 ± 1.00

The microscopic time constants of the Trp conformational change were calculated from the observed relaxation time constant τ_{R1} and amplitude a_1 (Fig. 4. 4. 12 c). We found that the Trp time constant to 'wiggle out' of the NTD was substantially longer for the mutant with $\tau_{out}^{L6} = 2.1 \pm 0.1$ ms, compared to the WT with $\tau_{out}^{L6} = 0.61 \pm 0.06$ ms.

From this follows that the methionine residues ensure a highly dynamical range of the N-terminal domain on a secondary and tertiary structure level. We further reckoned that not only the Trp seemed to stay more buried in the protein core for the L6-NTD, but that the mutant probably was overall less dynamic, as supported by the following NMR hydrogen-exchange experiments (chapter 4.4.6.).

4.4.6. Structure of the L6-NTD

To conclude the characterisation of the L6-NTD, we examined the protein structure with NMR spectroscopy. For this purpose, isotopically labelled proteins with ^{15}N and ^{13}C heavy atoms were synthesised. NMR measurements and analyses were performed by Benedikt Goretzki (University of Mainz).

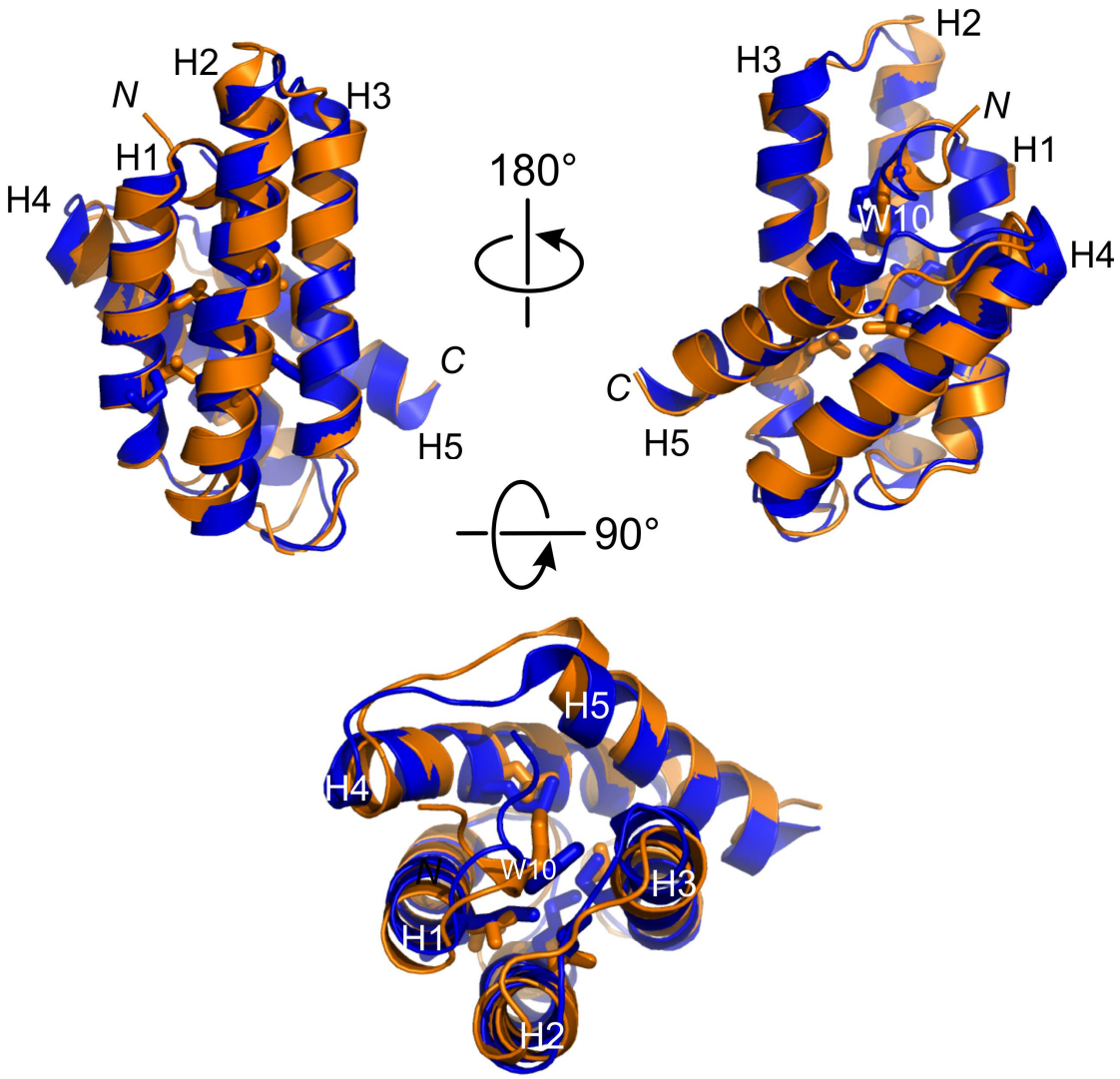


Figure 4. 4. 13: Superimposed NMR structures of Ea WT (blue, PDB-code: 2lpj) and L6-NTD (orange, PDB-code: 6qjy). Side view, tiled side view and top view with zoom on mutated side chains and conserved Trp W10. Helices are numbered H1 to H5, N- and C-terminal domains of the protein are indicated. (Fig. created by BG, HN and JH; Heiby et al. 2019)

Remarkably, the HSQC data showed that the wild type and L6-NTD mutant overlaid surprisingly well under monomeric conditions (Fig. 4. 4. 13). Therefore, the structure seemed unchanged. These results validate the various findings described above since an intact protein is essential for intrinsic fluorescence and dimerisation measurements. Moreover, the protein structure indicated that our modulation of the protein core was well tolerated by the NTD. Finally, the NMR data supported the hypothesis that the environment of the Trp seemed immobile upon dimerisation of the L6-NTD while changing significantly for the WT. Spectra for the dimeric NTD were only partially assigned, due to line broadening. Thus, the data backed the reduced dimerisation capacity of the mutant, as seen in high-resolution SEC and SEC-MALS experiments. The NTD seemed overall rigidified, having implications for both the dynamics and functionality of the protein. Proton-deuterium exchange experiments showed that on a per-residue level, the solvent-accessibility was ~10-fold slower for the mutant, compared to WT, demonstrating that the L6-NTD was less dynamic (Heiby et al. 2019).

The fusion of our above reported findings with the NMR structure of L6-NTD, lead us to the conclusion that the NTD had probably evolved from a leucine and isoleucine rich core to a protein, abundantly containing methionine. The original protein had to be stable, capable to endorse destabilisation in order to evolve higher functionality. The methionine enriched NTD, therefore, enables the dynamics and functionality of the protein, leading to a tightly bound dimer. Phylogenetic studies have shown that Met is one of the few amino acids which has gained in frequency in proteins over the past (Jordan et al. 2005). Since the three amino acids (Met, Leu and Ile) only vary by one nucleotide, a mutation over time seems feasible, if not even likely.

4.5. The role of leucine in the NTD

To test our theory, that methionine induces malleability in the NTD, we introduced new methionines on top of the methionine depleted L6-NTD mutant. The original idea of introducing Leu into the wild-type NTD was only implemented as a single point mutant, since a possible enhancement of the protein to an even tighter dimerisation, would not match the temporal resolution or signal sensitivity of our methods. Therefore, we exchanged the native leucines present in the hydrophobic core of the L6-NTD to Met, generating a 12-point mutant L6+M6-NTD (for mutated amino acid positions see table 4. 5. 1). In total, the original balance of Met and Leu was restored to native conditions, but with changed amino acid positions. We speculated to restore the physical properties of the NTD, which would answer the question if the amino acids Met and Leu had a position-dependent role in the hydrophobic core.

The expression of this protein and the gradual mutants (L6 + M1 to M6 NTD) was still very efficient. The CD measurements revealed that their secondary structure was still intact (Fig. 4. 5. 1 a). In thermal denaturation experiments, we observed, as expected, a gradual decrease in protein stability with every newly introduced Met (Fig. 4. 5. 1 b). Yet, the L6+M6 NTD's thermal stability was still greater compared to the wild-type NTD (by 7 K). However, the tryptophan spectra revealed that the Trp mobility did not improve by additional Met in the core (Fig. 4. 5. 1 c). Moreover, the Trp quenching observed in L6+M6 NTD (10 %) was significantly lower than for the L6-NTD (20 %), indicating that the dimerisation efficiency upon pH change did not improve. For all L6+ M1-M6 NTDs, the Trp quantum yield stayed steady (Fig. 4. 5. 1 d). Interestingly, the Trp fluorescence maximum wavelength shifted to higher values with every additional Met but did also not reach wild-type conditions in the L6+M6 NTD.

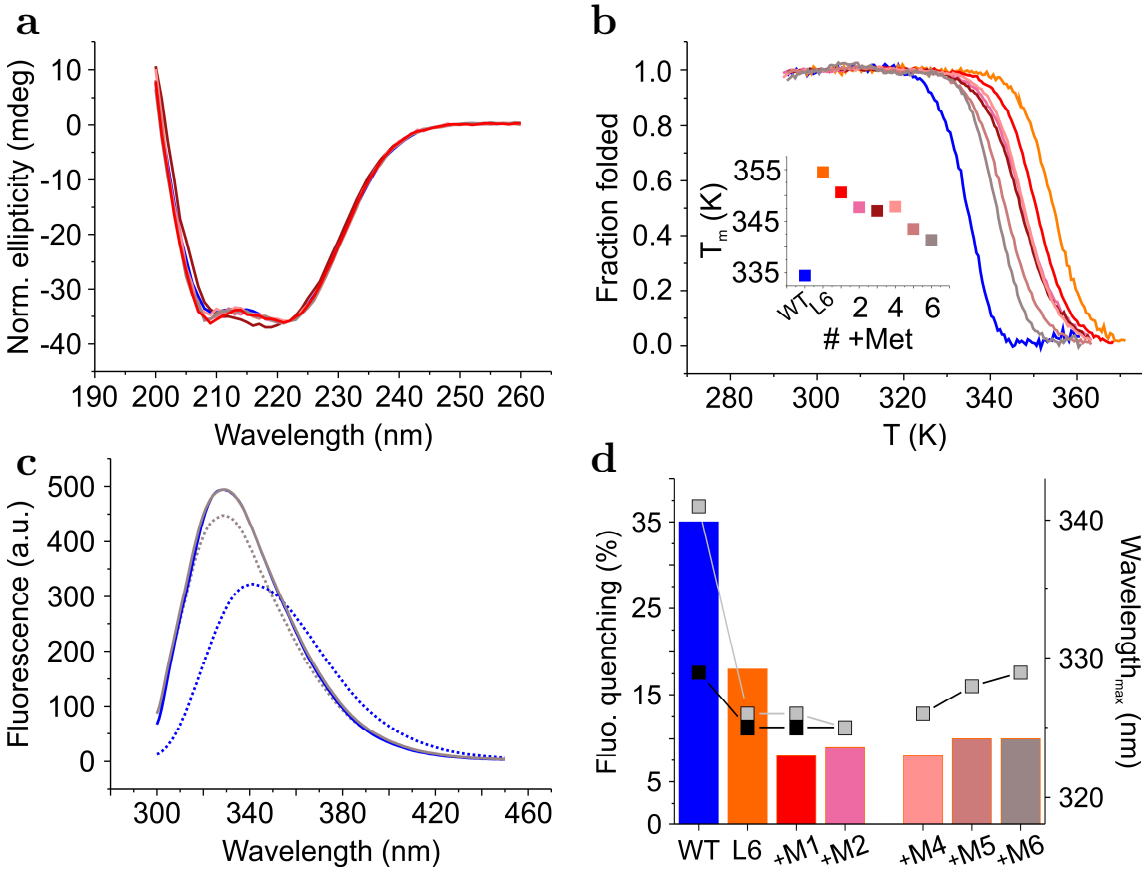


Figure 4. 5. 1: Secondary structure, equilibrium denaturation and conformational change of L6 +M1-M6 NTD. (**a**) Superimposed and normalised far-UV CD spectra of Ea WT (blue), L6-NTD (orange) and series of L6 +M1-M6 NTDs (shades of red: L6+M1 red, L6+M2 pink, L6+M3 brown, L6+M4 light salmon, L6+M5 dark salmon, L6+M6 grey). (**b**) Thermal denaturation of Ea WT, L6-NTD and series of L6 +M1-M6 NTDs measured with CD spectroscopy under pH 7 (200 mM *Is*) condition at 222 nm (same colour code as in (a) applies). Data were normalised to the fraction of folded protein. Graphic inset with transition mid-point melting temperatures T_m by number of exchanged Leu. (**c**) Tryptophan fluorescence spectra at pH 7 (solid lines) to pH 6 (dashed lines) of WT (blue) and L6+M6-NTD (grey). Data were normalised to the fluorescence maximum of the wild type. (**d**) Quenching of Trp fluorescence of WT (blue), L6-NTD (orange) and cumulative mutation series L6+M1-M6 NTD (shades of red) upon change of buffer condition. Wavelengths of maximal fluorescence emission at pH 7 (grey squares) and pH 6 (black squares) for each mutant. Maximal wavelengths for M2-M6 are concurrent for pH 6 and pH 7. L6-M3 was not included. Same colour code as in (a) applies.

We could not measure a reliable fluorescence spectrum for the L6+M3 mutant, which was therefore not included in Fig. 4. 5. Also, the CD spectrum of this mutant exhibited slight differences compared to the others (Fig. 4. 5. 1 a, L6+M3 in brown). Furthermore, this mutant showed no sign of stability loss, which could be explained for the reason that the amino acid is only adjacent to the core (> 16 Å to the centre of mass, see table 4. 5. 1). We demonstrated before, that the amino acid distance from the core centre is of importance for the effect on the protein (Fig. 4. 4. 11). The amino acid residue at position L55M (+M3) does therefore not have significant interactions with the hydrophobic core. Another possible explanation of why, compared to the L6-NTD, the quenching keeps increasing, is that Met is a Trp quencher. Assuming that L6+M1 is close to the Trp would explain why no initial improvement in fluorescence is detectable. However, this should have been compensated by the exchange of six Met in L6+M6, which was not the case. The malleability was not restored. There must, therefore, be an amino acid position dependency in the protein core.

Table 4. 5. 1: Relationship between amino acid position and distance to the core (determined by PyMOL, referring to WT-NTD PDB-code 2lpj) with melting temperatures T_m of the L6-NTD and the cumulative series L6+M1-M6 NTD.

Leu-Position	Distance (Å)	T_m (K)	T_m loss (K) relative to L6-NTD
L6-NTD*	(see table 4.4.1)	354.5±0.20	--
27 (+M1)	9.5±0.3	350.6±0.09	-3.9
38 (+M2)	13.4±0.5	347.7±0.16	-6.8
55 (+M3)	16.3±0.3	347.0±0.16	-7.5
66 (+M4)	12.7±0.3	347.8±0.13	-6.7
69 (+M5)	10.1±0.1	343.5±0.09	-11.0
90 (+M6)	14.3±0.3	341.4±0.18	-13.1

*L6-NTD T_m melting temperature from Heiby et al. 2019.

The combination of L6+M4 showed a stabilising effect with the additional mutation +M4 (+0.82 K), whereas the single point mutant of the same position L66M showed a slightly destabilising effect (-0.35 K), as would be expected. The difference being that in the polymutant (=10 mutations) the protein was

presumably stiff (as was the L6-NTD, shown in chapter 4.4.), while the new Met (+M4) interacted with Trp and had a stabilising effect. Direct contact between Trp and Met is stabilised by the interaction of the sulphur atom and the NH group of Trp, which can form a hydrogen bond when the sulphur atom is in the right angle. The proton acts as an electrophile and the divalent sulphur as a nucleophile (Pal and Chakrabarti 2001). In contrast, the NTD of the point mutant is very dynamic (confirmed by PET-FCS data Fig. 4. 5. 2, table 4. 5. 3), and the mutation has a slightly destabilising effect, as is the case with the other +Met mutations (table 4. 5. 1). Amino acid residues in long distance from the core appear to be at the edge of the NTD, and therefore have fewer interactions with other amino acids. This, in turn, may explain why the destabilising effect of L66M was small (compared to M→L stabilising effect).

Next, we compared K_D values of the L6+M1-M6 series, measured by high-resolution SEC (table 4. 5. 2). The K_D value did not regain wild-type values, as was anticipated by the previous experiments (Fig. 4. 5. 1).

Table 4. 5. 2: K_D-values of Ea WT, L6-NTD and cumulative series of L6+M1-M6 NTD measured by high-resolution SEC.

NTD (number of measurements)		K_D-value
WT*		1 nM ± 0.1 nM
L6*	(#17)	1.09 ± 0.16 µM
L6+M1	(#6)	1.02 ± 0.33 µM
L6+M2	(#5)	0.93 ± 0.45 µM
L6+M3	(#5)	0.96 ± 0.31 µM
L6+M4	(#5)	1.00 ± 0.24 µM
L6+M5	(#5)	0.57 ± 0.17 µM
L6+M6	(#13)	1.78 ± 0.27 µM

*WT measured by stopped-flow spectroscopy from Heiby et al. 2017; L6-NTD from Heiby et al. 2019.

Neither the Trp (Fig. 4. 5. 1 c, d) nor the whole protein did regain the mobility it had in wild-type condition (table 4. 5. 2). The protein core seemed to remain rigid.

The question remained, why the reverse mutation L6+M6 did not result in a better K_D value. Presumably, all L6+M1-M6-NTD dimers were dominated by strong electrostatic forces, which remained unchanged here, since the protein could be mutated so intensively and still showed dimerisation.

The thermal folding T_m destabilisation (Fig. 4.5.1 b) could signify that the protein becomes more malleable, or in turn, that we gradually damage the protein. However, all proteins expressed very well, indicating otherwise. A final analysis of this intriguing, yet thus far not completely understood issue could be the subject of future work.

Following these observations, it seemed crucial to examine the role of leucine in the NTD in the context of the wild-type protein. Consequently, we analysed the single point mutant L66M, where the wild type's Leu at sequence position 66 was replaced by Met. PET-FCS measurements indicated that that the newly introduced Met made the Trp in the NTD core more mobile, compared to the wild type (Fig. 4.5.2).

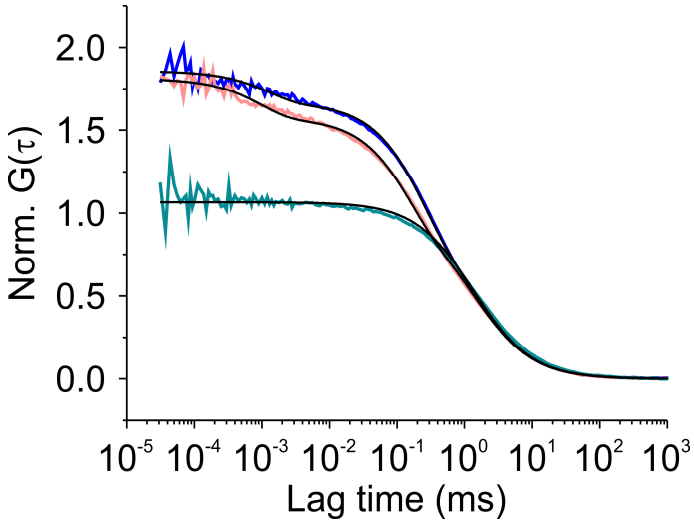


Figure 4.5.2: Dynamics of Ea WT and L66M-NTD probed by PET-FCS. The transients were normalised from Ea-G3C-WT (blue), L66M-G3C-NTD (light rose) and reference molecule of L6-Q50C-NTD (cyan). All proteins were labelled with AttoOxa11 dye and measured at a concentration of 1 nM in monomeric pH 7 buffer. Data were fitted by a model for diffusion of a globule containing two single-exponential relaxations, and the control sample was fitted by a model for diffusion only (black lines).

L6-Q50C-NTD was chosen as a reference, as the labelling site did not interact with the intrinsic Trp and did therefore only show diffusion. In comparison to the reference curve, the *wild-type* G3C-WT showed strong Trp dynamics, signified by additional PET kinetics with a first relaxation time constant τ_{R1} of 240 µs. This time constant was even faster for the L66M-NTD with 145 µs. The second kinetic τ_{R2} also showed faster behaviour by a factor of two (table 4. 5. 3). This observation indicates that by exchanging Leu to Met, the Trp's dynamical range can be improved.

Table 4. 5. 3: PET-FCS diffusion time, amplitudes and relaxation times of Ea G3C-WT and L66M-G3C-NTD.

NTD	τ_{Dif} (ms)	a_1	τ_{R1} (µs)	a_2	τ_{R2} (µs)
G3C-WT	1.43±0.08	0.64±0.03	240±17	0.19±0.04	2.00±1.50
G3C-L66M	1.39±0.01	0.58±0.02	145±6	0.23±0.01	1.00±0.60

At the time of completion of this thesis, the K_D-value of the single point mutant L66M has not been determined yet. According to our current theory, the K_D value would be smaller for the mutant than the wild type's. In order to confirm this hypothesis, k_{ass} and k_{diss} of the L66M should be determined with the stopped-flow technique, consistent with the previously described measurements.

Ultimately, the wild-type characteristics could not be restored by interchanging the position of the residues. Yet, this is not completely surprising, since we established that the residue's distance from the protein core is crucial for the effect on protein properties. The reintroduced Mets in the L6-NTD were at positions significantly farther from the core than their native counterparts (compare table 4. 5. 1.). The residual effect, observed in the L6+M6-NTD, could be a result of this distance mismatch. However, a higher dynamical range could be generated as a result of exchanging only a single leucine with methionine. We could demonstrate conclusively, that methionine has a destabilising effect when introduced at the expense of leucine, which is probably due to its side-chain flexibility, as the thioether bond promotes rotation, but with this, adaptability (Gellman 1991).

4.6. The methionine-effect in NTD homologs

To transfer the concept of methionine residues inducing structural plasticity and leucine residues rigidity and stability, we reproduced the mutations, analysed for the *E. australis* NTD, in homologous NTDs. Therefore, we chose to transform the *N. clavipes* and *L. hesperus* NTD, each having five methionine residues in core positions or being involved in tertiary interactions (amino acid positions for Nc: 24, 41, 52, 55, 77; Lh: 38, 41, 49, 52, 55). After exchanging the Met to Leu residues accordingly, the expression of the mutant proteins worked well. The analysis of their secondary structure by circular dichroism spectroscopy showed no significant change, compared to their respective wild-type form (Fig. 4. 6. 1). Thus, the L5 mutants (L5-NTD^{Lh} and L5-NTD^{Nc}) appeared to be structurally preserved.

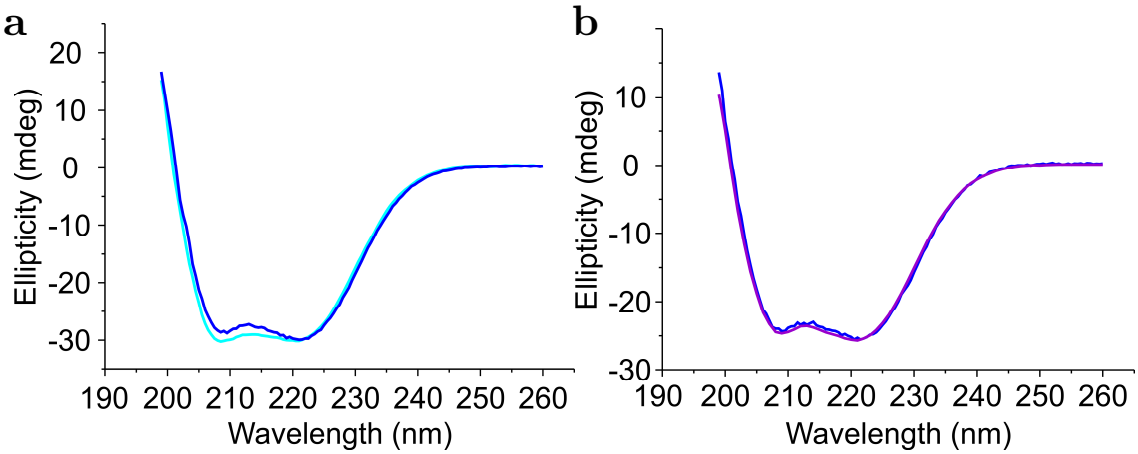


Figure 4. 6. 1: Analysis of the secondary structure of L5-NTD^{Lh} and L5-NTD^{Nc}. (**a, b**) Far-UV CD spectra of WT-NTD^{Nc} and WT-NTD^{Lh} (blue) in comparison to spectra recorded from L5-NTD^{Nc} (cyan, (a)) and L5-NTD^{Lh} (violet, (b)), respectively. Data were normalised to the according WT ellipticity.

To verify if the stabilising effect found for L6-NTD of Ea was also apparent in the homologs, we performed chemical and thermal denaturation experiments with CD and fluorescence spectroscopy (Fig. 4. 6. 2). The result showed stabilising effects in both homologs, albeit slightly weaker than in the *E. australis* L6-NTD. Relative to their wild types, the L5-NTD^{Nc} gained 2.0 kcal/mol in folding stability, L5-NTD^{Lh} gained 0.7 kcal/mol, in comparison to 2.2 kcal/mol for Ea L6-NTD.

Consistently, the thermal denaturation showed for L5-NTD^{Nc} a folding stability gain of 13 K and for L5-NTD^{Lh} 8 K, compared to 17.2 K for L5-NTD^{Ea}, in regard to their wild-type from.

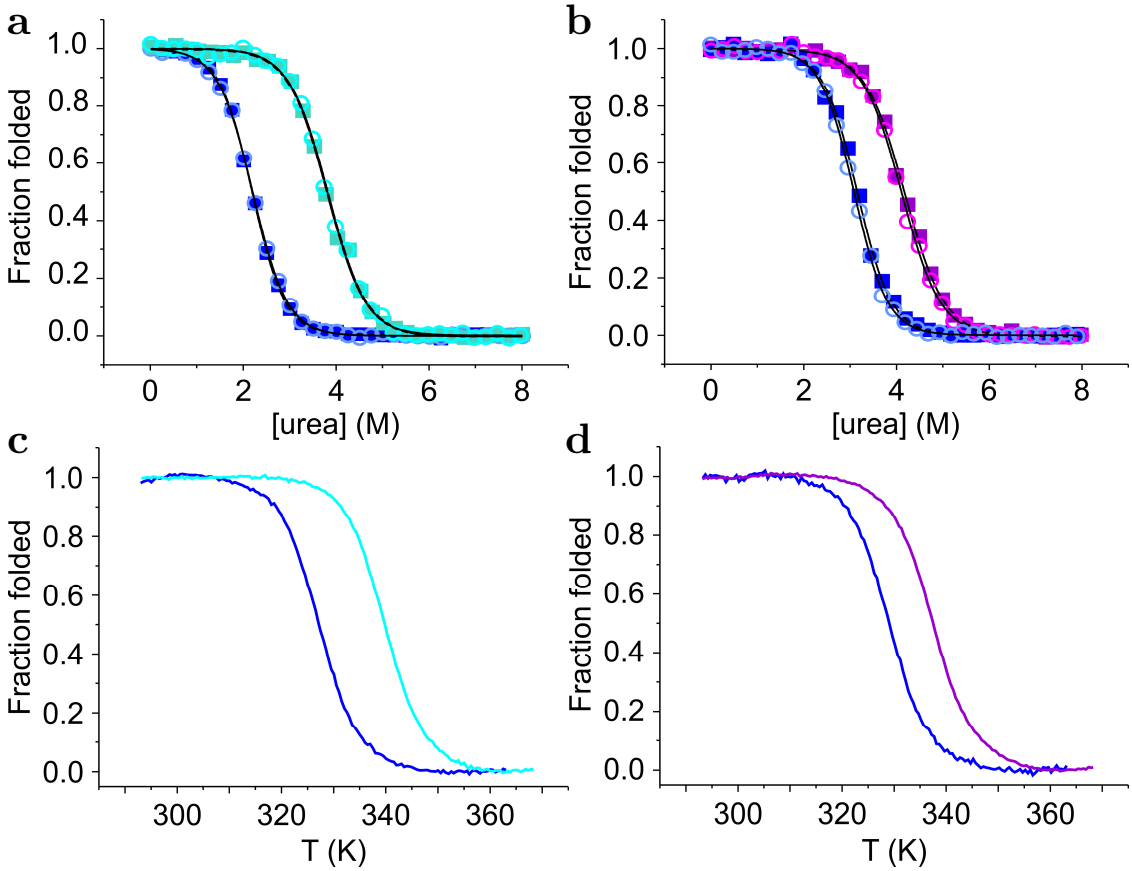


Figure 4. 6. 2: Denaturation of L5-NTD^{Lh} and L5-NTD^{Nc} under equilibrium conditions measured by CD and fluorescence spectroscopy. (**a, b**) Chemical denaturation of L5-NTD^{Nc} (cyan, (a)) and L5-NTD^{Lh} (violet, (b)) with respective wild-type NTD (blue). Denaturation was measured using far-UV CD spectroscopy at 222 nm (open circles) and Trp fluorescence at 325 nm (closed squares). Data recorded from the corresponding wild-type proteins are shown in blue. Data were fitted by a model of two-state equilibrium (solid lines). (**c, d**) Thermal denaturation monitored by far-UV CD spectroscopy at 222 nm with L5-NTD^{Nc} (cyan, (c)) and L5-NTD^{Lh} (violet, (d)) with respective wild-type NTD. All data were normalised to the fraction of folded proteins.

While the Met-to-Leu mutations seem to have a similar impact on Nc and Ea, the measured effect was smaller for Lh. In order to elucidate the reason for this reduced stabilisation, the composition of the cores of Ea, Nc and Lh were examined. Even though the core of Ea is packed with Met, it is also packed with

Leu, Ile and Phe. Phe is also a hydrophobic amino acid residue, typically found in the core of proteins (Villar and Koehler 2000). This phenomenon seems to be true for all three NTDs. Ea WT is slightly more stable than the other two homologs when comparing mid-point melting temperatures T_m (see chapter 4.1.). The reason for this might be, that Met-Phe has a stabilising effect when oriented in the correct angle (Pal and Chakrabarti 2001), while Met-Phe in combination with Leu/Ile in the core probably also leads to higher stability (=Ea WT). In case of the other two NTD homologs, the missing or smaller effect of Met-Phe stabilisation seems to be compensated by a Leu/Ile richer core (both effects for Nc, but with less Met).

There are several relevant differences regarding the methionine content of the three proteins: Ea has the most Met in core position, followed by Nc, which also has several Met in core position, explaining the similarity to Ea; Lh has its Met almost all in tertiary interactions, thus not as central as in case of the other homologs.

Therefore, the question arose, why the Lh-NTD does not have Met in core positions, as they seem to be essential for the NTD's malleability. Strikingly, the *L. hesperus* and *N. clavipes* exhibits Leu and Ile residues at positions where *E. australis* has Met. Hence, the core of those NTDs must be already stabilised. As a result, we could only measure a small stabilising effect on Lh, since the newly introduced Leu residues were not very close to the core centre, and there are already Leu's in the core (which was not the case for Ea).

The comparison of the homolog's Trp fluorescence spectra revealed that the quenching effect varied from the one previously observed for the Ea-NTD. The characteristic signature of Trp fluorescence, seen for the Ea L6-NTD, disappeared for both homologs (Fig. 4. 6. 3). The Trp residue of L5-NTD^{Nc} still seems to face a change in surrounding conditions, but to a lesser extent than for its WT. There was near to no observable quenching of L5-NTD^{Lh} fluorescence upon pH change. This suggests, that either the protein had lost its ability to dimerise entirely or that the Trp got immobilised in core position.

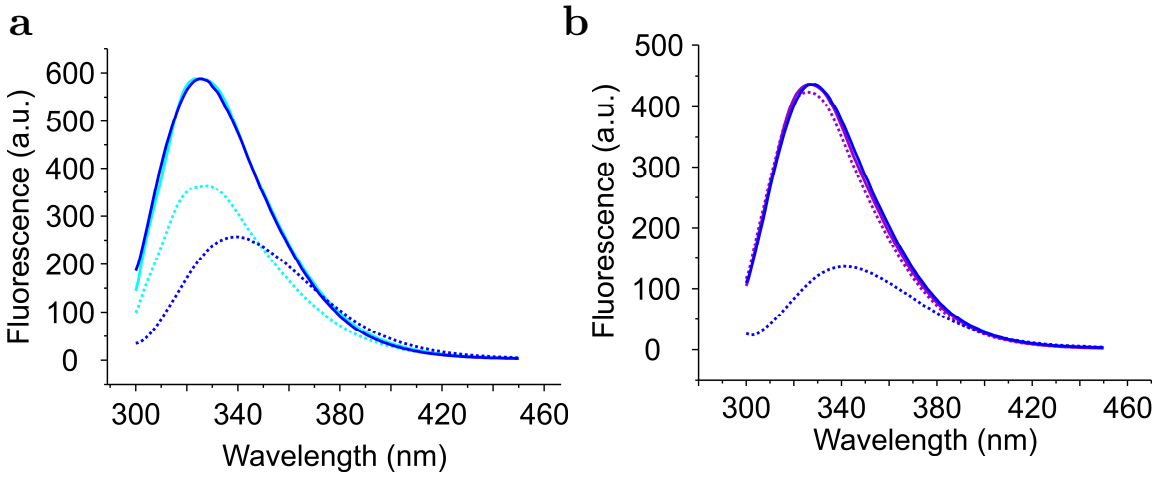


Figure 4. 6. 3: Dimerisation of L5-NTD^{Lh} and L5-NTD^{Nc}. (**a**, **b**) Trp fluorescence spectra of L5-NTD^{Nc} (cyan, (a)) and L5-NTD^{Lh} (violet, (b)) measured at pH 7 (200 mM *Is*; solid lines) and pH 6 (8 mM *Is*; dashed lines), with respective wild-type NTD (blue). Data were normalised to the maximum fluorescence of their relative wild type at pH 7.

Yet, since Met acts as a Trp quencher, the question remained, if the reported measurements are a reliable metric for the dimerisation of the mutants, or if the replacement of said Met by Leu caused a significant increase in fluorescence. Based on the Trp spectra of Nc, we reckoned, that no mutated Met is in direct contact with the Trp (verified in PyMOL, PDB-code 5iz2). Therefore, there can be no attenuation of quenching by the mutation. Furthermore, the non-quenching of Lh's Trp is probably based on other effects. Here, all mutations were performed on only one of the five helices. Yet, this particular helix appears to be involved in the dimerisation interface. If the helix is rigidified upon mutation, the dimerisation would be constrained or the Trp blocked.

Notably, the high-resolution SEC measurements clearly showed that dimerisation of the mutated proteins was still possible (Fig. 4. 6. 4). Consistent with the Ea-L6-NTD, the K_D value increased for both homologous L5 mutants, albeit to a lesser extent. As demonstrated in Fig. 4. 6. 4, we could observe a clear change in the dimerisation characteristics of the mutants. There was an onset of dimer dissociation and an earlier elution volume of the dimer, in contrast to their wild-type form. The K_D value seemed to be increased as well, but no quantitative values could be calculated, as the detection limit was reached before monomer

conditions were attained (concentration in low nM). We estimated from the late onset of association that the K_D had to be significantly higher for both mutants, relative to their wild types.

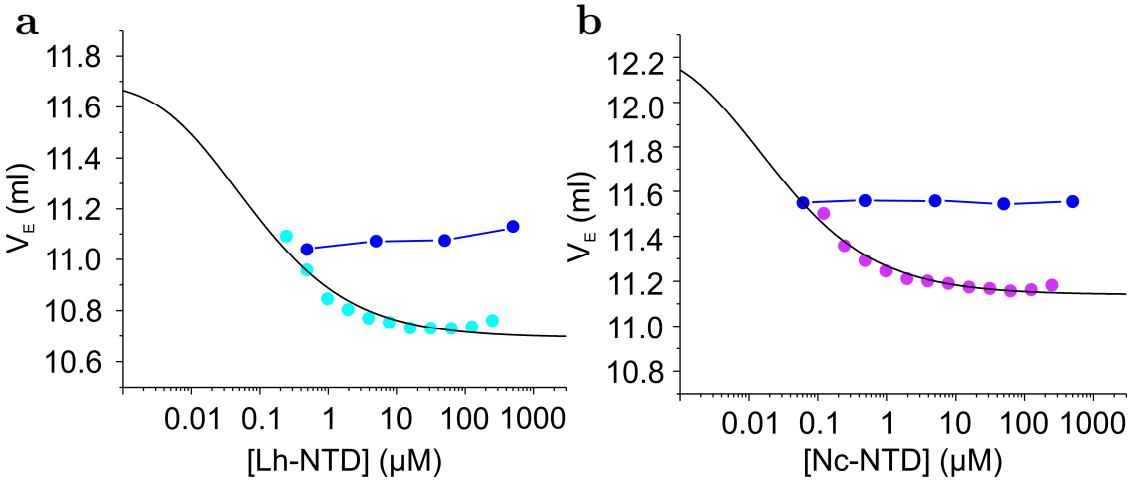


Figure 4. 6. 4: (**a, b**) Characterisation of dimerisation by titration of L5-NTD^{Lh} (cyan, (a)), L5-NTD^{Nc} (violet, (b)) and respective wild-type NTDs (blue) using high-resolution SEC at pH 6. Elution volumes V_E were plotted against the protein concentration of L5-NTD^{Lh} and L5-NTD^{Nc}. Data points were fitted by a model for dimerisation equilibrium (black lines) and extrapolated to the x-axis, with fixed V_E for the monomer. WT V_E connecting blue lines are guide to the eye.

The titration curve of the equilibrium dissociation constant indicated that the wild-type proteins dissociated far later as the mutants, and therefore had to be associated significantly stronger. The similarity to the L6-NTD^{Ea} measurements is striking. Yet, the estimated K_D values for the mutants correlate with the K_D values measured by stopped-flow experiments for the wild types. K_D for Nc WT was determined to 0.73 nM and Lh WT with 0.14 nM, being below the detection limit for our high-resolution SEC technique. If the WT had a late onset of dimer dissociation, the K_D value would be smaller than the mutants. Therefore, the concentration detection for those measurements stays uncertain. However, the apparently existing association can be attributed to electrostatics, as previously described for the L6-NTD^{Ea} (Heiby et al. 2019). Here again, electrostatic forces are probably an important supporting agent enabling the protein association.

Although less Met residues appear in the core of Lh-NTD than in Ea-NTD, the amino acid likely fulfils the same functions. The malleability factor of Met may

not be position bound, in the sense that closer to the core centre positions do not generate a more significant effect, unlike the Leu position dependency (as seen in Fig. 4. 4. 11). Thus, the effect generated in the Lh-NTD is slightly weaker than for Ea-NTD. In this regard, the Ea inherits a truly special NTD.

Finally, when introducing Leu to the homologous proteins, a qualitatively comparable impact could be observed on the protein's characteristics, as was for the Ea-NTD.

4.7. Transfer of the methionine hypothesis to Hsp90

Intending to test, if our methionine hypothesis applies to a more general range of proteins, we chose to examine the protein Hsp90. It is a molecular chaperone that is of high importance in an extensive set of organisms and necessary to fold new proteins correctly. Hsp90 correlates with many human diseases, such as cancer or diabetes (Csermely et al. 1998). Therefore, it constitutes an attractive target to modulate its functionality and stability. Compared to the NTD of spidroins, the Hsp90 protein is bigger, with approximately 90 kDa. It is composed of three major regions (Fig. 4. 7. 1): The N-terminal domain (NTD^{Hsp90}), an ATP/ADP binding site, a highly charged middle region and the C-terminal domain (CTD^{Hsp90}), which is essential for the homo-dimerisation of the monomers (Csermely et al. 1998). We chose to modulate the CTD^{Hsp90}, as it has a critical role in dimerisation, similar to the spider silk NTD. For the spider silk NTD, we found Met to be essential for the protein association. Therefore, we introduced Met in Hsp90 intending to promote structural dynamics and potentially enhance the interaction of the homo-dimer.

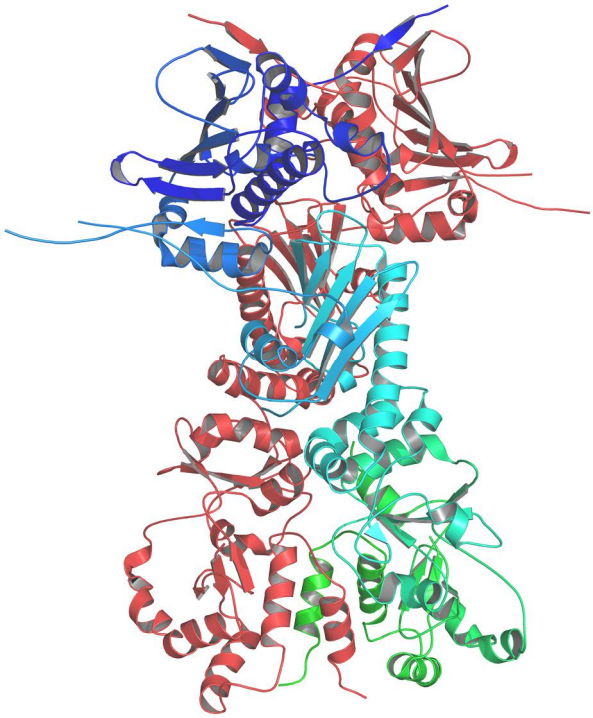


Figure 4. 7. 1: Hsp90 dimer ribbon diagram with one monomer in red, the other in shades of blue to green, PDB-code: 2cg9; rendered by PyMOL. From top to bottom: NTD, middle domain and CTD of the protein.

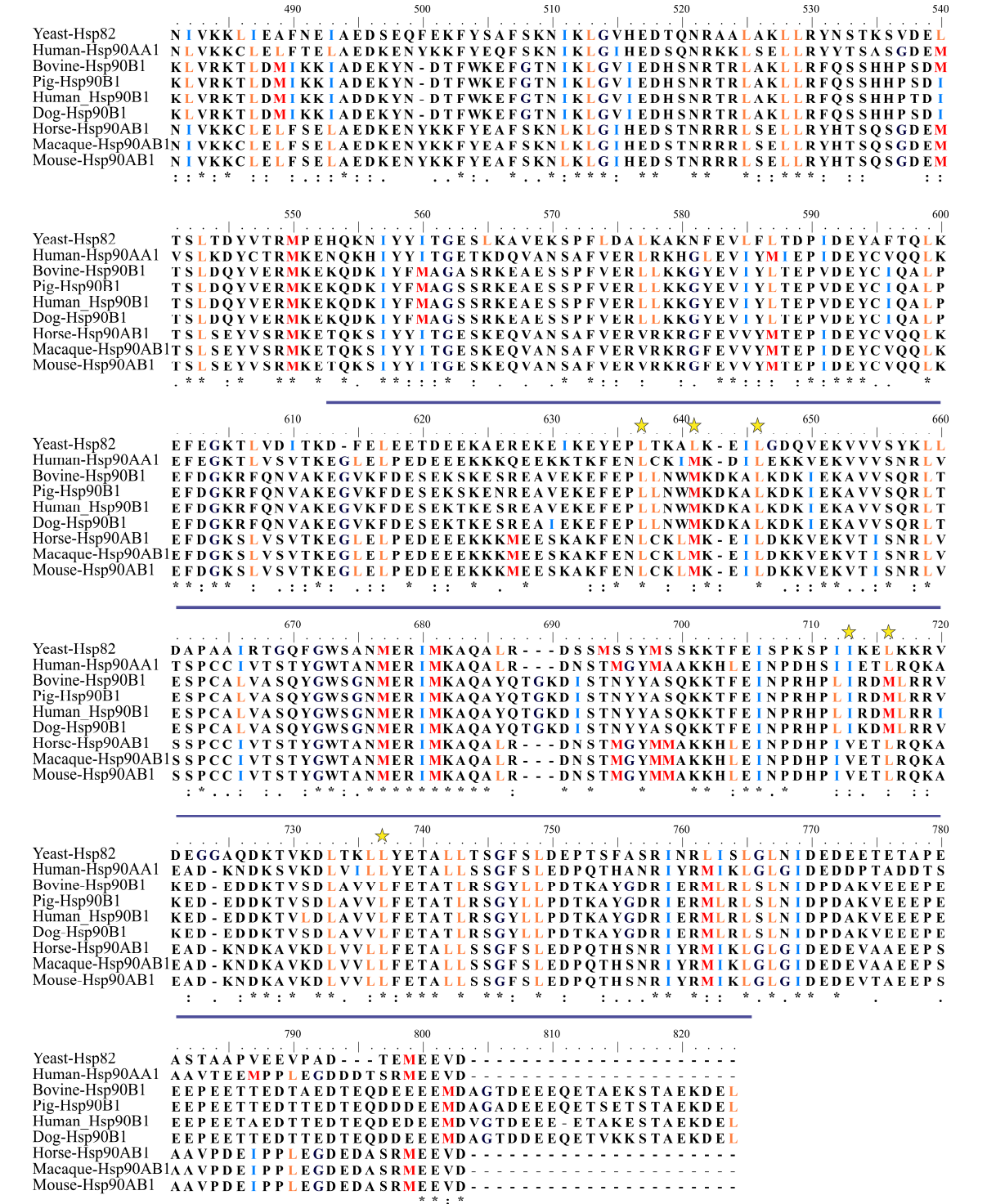


Figure 4. 7. 2: Sequence alignment of Hsp90 homologues, originating from different species. Conservation of the amino acid residues methionine (M), leucine (L) and isoleucine (I) are highlighted in red, orange and blue, respectively. Indication of encoded CTD by a blue bar and mutated positions of the Yeast-Hsp82 marked by yellow star, on top of the alignment. Amino acid conservation are denoted identical (*), very similar (:) and similar (.) with Clustal Omega.

The CTD^{Hsp90} is highly conserved in all kind of organism, ranging from Arabidopsis thaliana (plant) to the human Hsp90 (Fig. 4. 7. 2). The alignment of the Hsp90 homologs shows that especially leucine is highly conserved between species (sequences from Chen et al. 2006). Interestingly, several amino acid positions have methionine or isoleucine instead of leucine. The phenomenon of the three amino acid (MIL) sharing the same sequence position seems to be conserved in Hsp90 as well. Here, only a partial sequence is shown, as the full-length sequence has over 700 amino acids and the focus lies on the C-terminal part of the Hsp90 (CTD^{Hsp90} highlighted with blue bar in sequence alignment). Nevertheless, all experiments were conducted on the full-length Hsp82 of yeast (*Saccharomyces cerevisiae*, referred here as Hsp90). According to our theory, we should induce malleability into the CTD^{Hsp90} of the protein, by changing Leu or Ile to Met, resulting at best into a tighter homo-dimer of Hsp90.

Our theory suggests that the effect of methionine is strong in core positions of proteins. Therefore, we only considered mutating leucine residues in α-helical domains contained in a hydrophobic pocket. In the CTD^{Hsp90}, we found several Leu being in said positions. We chose five Leu and one Ile in core positions and gradually mutated them by site-directed mutagenesis. Leucine positions 637, 641, 646, 716, 737 and isoleucine at position 713 were mutated to methionine, generating Hsp90-M1 (L641M), -M3 (637, 641, 646), -M4 (637, 641, 646, 716), -M5 (637, 641, 646, 716, 737) and –M6 (637, 641, 646, 713, 716, 737)(nomenclature referring to sequence alignment in Fig. 4. 7. 2 from sequence PDB-code: 2cg9, mutations highlighted by stars). Especially the Met amino acid at positions 641 and 716 were of special interest, as they seem to be conserved in other homologs.

Under normal conditions, Hsp90 hydrolyses ATP (adenosine triphosphate, storage of molecular energy) in order to function. This induces a clamp movement of its N-terminal domain, closing and opening in response to the binding and hydrolysis of ATP (Schulze et al. 2016, Fig. 4. 7. 3). A classical approach to test the functionality of Hsp90 is by probing the ATPase rate.

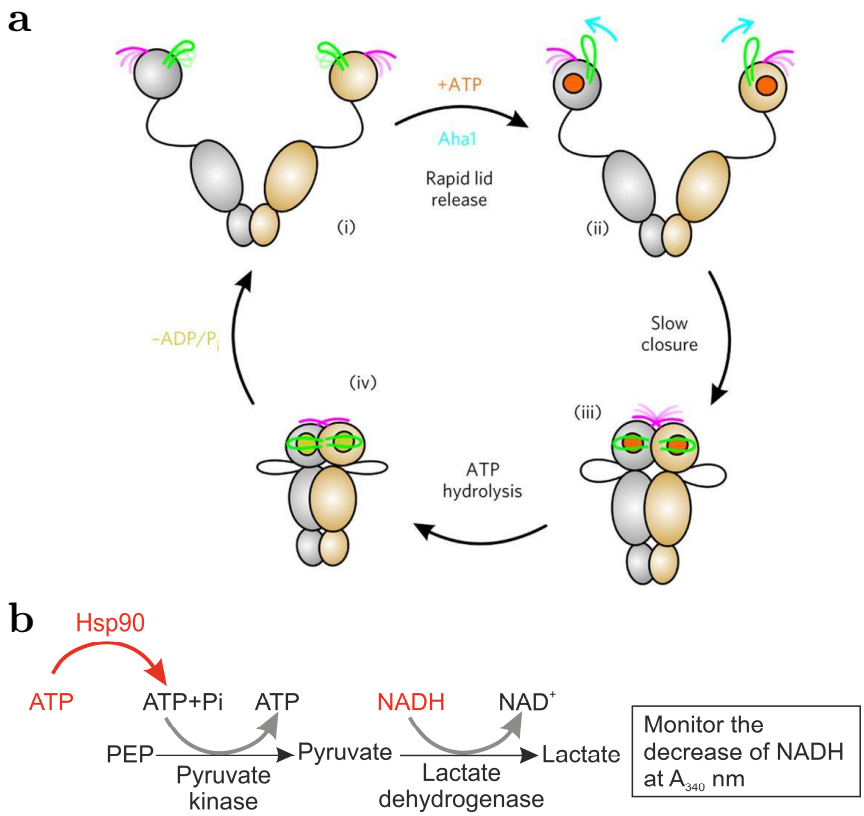


Figure 4. 7. 3: Model of the bacterial Hsp90 ATPase cycle and assay. (**a**) ATPase cycle showing Hsp90 "clamp" closure of the Hsp90-NTD, ATP hydrolysis and release. (**b**) Detailed molecular ATP hydrolysis assay to monitor the ATPase rate by addition of PEP (phosphoenolpyruvate), pyruvate kinase, NADH and Lactate dehydrogenase. (with permission from both Schulze et al. 2016 and Hoskins et al. 2018)

The wild-type ATPase rate was determined to be 1.81±0.14 ATP/min at 310 K, which was in good agreement with previously reported values (1.60±0.04 ATP/min, Schulze et al. 2016). The ATPase assay showed a 30 % enhanced ATPase rate for the triple M3, quadruple M4 and quintuple M5 Hsp90 mutant (table 4. 7. 1). For the Hsp90-M6 mutant, the ATPase rate had almost dropped to zero. All measurements were conducted at 310 K, a typical temperature condition present in many mesophilic organisms (Wiegel 1990).

Table 4. 7. 1: ATPase rate constants of Hsp90 WT and cumulative Hsp90 M1, M3-M6 mutants. Errors are stated as standard deviation of the mean (# number of measurements indicated in brackets for each construct).

Hsp90	ATPase rate (ATP/min)
WT (#3)	1.81 ± 0.14
M1 (#4)	1.64 ± 0.04
M3 (#3)	2.59 ± 0.07
M4 (#3)	2.43 ± 0.17
M5 (#3)	2.52 ± 0.28
M6 (#3)	0.002 ± 0.0008

Based on the newly gained insights of spidroin NTD homologs, Leu ought to be involved in the stability of the protein, and therefore, we should have strongly destabilised the Hsp90 protein. By adding Met, the malleability was increased, which lead to a slightly improved ATPase rate. To compensate for the stability loss, Met would require interaction with Phe. Only one of the mutated Met exhibited a possible contact with Phe, namely the M2 mutant. Unfortunately, this particular mutant was not expressed and therefore not examined separately.

The mutations affected the K_D value of the Hsp90 dimer, as expected. The homo-dimerisation K_D value of wild-type Hsp90 is 60±12 nM (measured with SEC-HPLC, Richter et al. 2001). To determine the K_D values of the mutants, we performed FCS measurements. The diffusion time of the proteins was tracked while titrating increasing amounts of unlabelled Hsp90 protein to 0.1 nM labelled Hsp90 (NHS-ester labelling, SeTau647, Fig. 4. 7. 4). When the Hsp90 dimer associated, the diffusion time increased as the effective molecular weight augmented. The wild-type K_D was in good agreement with the literature value of 33±9 nM. Hsp90-M1 had a similar K_D of 55±19 nM, being almost equal to wild type. With further Met mutations, the K_D value increased. Hsp90-M3 showed an equilibrium dissociation constant of 82±24 nM, M4 of 1.5±0.7 µM, while M5 and M6 were both not quantifiable.

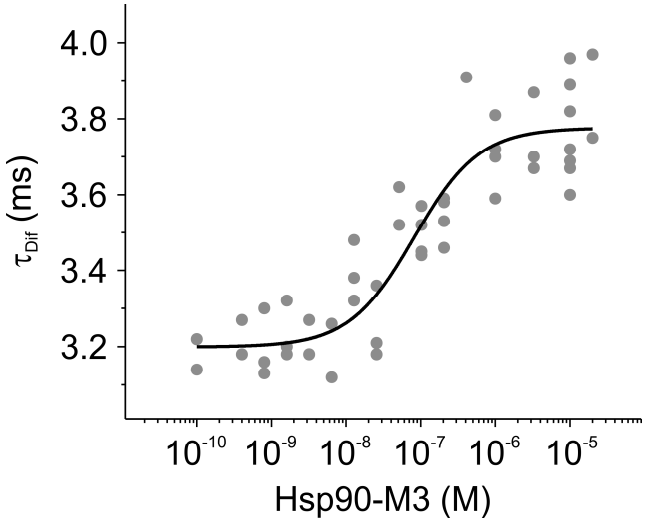


Figure 4. 7. 4: Exemplary dimerisation analysis of Hsp90-M3, based on FCS measurements. Titration of unlabelled Hsp90-M3 to labelled Hsp90-M3 protein (0.1 nM, with SeTau647) and the measured diffusion time constant (τ_{Dif}) concentration-dependently (grey spheres). Data points were fitted with a ligand-receptor model (black line).

Intriguingly, the results for M1 showed little to no effect in comparison to the wild type. Nevertheless, the M1 mutation seemed to be well tolerated by Hsp90. The triple mutant M3-Hsp90 revealed a tendency to higher ATPase rates (30 % increase). However, inserting the fourth methionine residue into the protein caused a significant jump of almost two orders of magnitude in the observed equilibrium dissociation constant. Furthermore, in case of six Leu-to-Met mutations, the protein seemed to have dissociated entirely, which is consistent with its effectively reduced ATPase rate to 0 ATP/min (table 4. 7. 1). Potentially, there is a certain Leu-to-Met threshold level that has to be overcome in order to achieve an effective change in the dimerisation capacity of Hsp90. If the K_D value increases with every additional Leu-to-Met mutation, it would result in K_D values beyond the detection limit of our method (as apparent for the M5 and M6 mutants) and finally in a completely destabilised protein (as seemingly observed for the M6 mutant).

Nevertheless, the effect of the heightened ATPase rate was not cumulative with increasing number of Met-to-Leu exchanges and rather small when compared to other mutations made on Hsp90 (Richter et al. 2001; Schulze et al. 2016).

Also, the K_D value did not decrease as anticipated and thus did not cause a stronger binding of the dimer. But with three mutations, the association remained relatively close to the wild type condition and showed tentative signs of improved functionality. Surprisingly, although the mutations were conducted in the CTD^{HSP90}, we observed quantifiable effects on the opposing NTD^{HSP90}. This seems even more impressive when considering that the K_D value increased only slightly with three mutations, and the protein was still active. This could mean, that the malleability and therefore an enhanced functionality is possible if the protein itself is stable enough.

Consequently, one should seek other model proteins to validate these results found for the spider silk NTD and Hsp90. Optimally, the protein should be α-helical with a hydrophobic core, in which leucine and methionine residues are located. The ideal protein would be a homo-dimer, like the spidroin NTD, and examinable by the same methods used for this work. Then, one could aim to systematically alter specific protein properties, like kinetic parameters, stability or functionality, by introducing methionine and/or leucine at relevant protein sequence positions.

5. Conclusion

Altogether, this thesis showed that the terminal domains of the spidroins are conserved between species, not only in their amino acid sequence but also in their energetics and kinetics of self-assembly. For this, I analysed four NTD homologs, all exhibiting remarkably fast kinetics of folding. These findings underline that this domain could be a promising object for future studies, which combine experimental and computational approaches, as they access overlapping time scales. The NTD folding appeared cooperative and could be described by a two-state transition model. We found similarities between homologs, regarding the dimerisation signatures, including pH and salt dependency, high-speed self-association and low K_D values. Also, the consistent red-shift and intensity-loss of tryptophan fluorescence emission indicated how closely related the proteins are to each other and confirmed their equal roles in spider silk association while sharing the same mechanisms of dimerisation. After having elucidated the folding mechanism of the N-terminal domain in detail, the CTD was surprisingly different. The fold turned out to belong to a class called knotted proteins. The CTD proteins are linked in a homo-dimeric way from early on, by a natural di-sulphide bridge. The different folding steps included a fast folding step and dimerisation of the central helix, which was followed by a slow domain swap. The C-terminal helix was then threaded in the dimer to form a molecular clamp that provided a scaffold for the attachment of the labile N-terminal helices. We deduced from our findings that the CTD could have a significant role in the elasticity of spider silk.

In the near future, the synthesis of spidroins or miniaturised spidroin material will hopefully be possible, as the count of published protocols increases. It will then be possible to analyse the effects of diverse mutations, observed for the terminal domains and to compare gained insight from the isolated terminal domains with full-length characteristics. For now, preliminary data indicates that our results from isolated silk domains can be considered valid for the full-length

spidroin. In this context, it would be interesting to see what impact Met-depleted NTDs have on whole spidroins. When considering the evolutionary trend, indicating that Leu was predominant in protein sequences, while only later Met was introduced (Jordan et al. 2005), one could imagine that the silk threads were once less elastic but therefore more resistant. Different environmental conditions might have led to different demands of protein characteristics. For example, thermostable proteins could have been subject to an optimisation process during evolution, responding to heat-stress with structural rearrangement of their energetic interactions (Miotto et al. 2019). To answer the question if a Met-depleted core has implications for the overall characteristics of the silk, one would need to mutate the core methionine residues of the NTD and synthesise the spidroins. But to this day, the production of synthetic spidroins is an ongoing challenge.

The effect methionine residues have on the NTD is significant, as has been described here. They are crucial to render the domain malleable, dynamic and adaptable. One could even speculate that Met has a positive impact on the elasticity of the NTD; a feature which could be conveyable to the whole silk. By exchanging those core-methionine residues to leucine, the domain was substantially stabilised, but at the cost of its dimerisation capacity, since leucine residues rigidified the NTD. Remarkably, six cumulative point mutations were necessary to gather enough evidence to comprehend their role and functionality. Nevertheless, the modification of proteins, by either introducing Met or Leu, should be used lightly. Therefore, to use this approach as a protein engineering tool, two mutations should be considered as a reasonable starting point. In order to modify other proteins, I would suggest to first examine the alignment of several homologous sequences from the protein of interest, to compare the MIL content (Met/Ile/Leu) and then decide for specific mutations. The interesting amino acid positions should at best have the trio of Met/Ile/Leu conserved and should relate to the protein core. New Met should be introduced as a substitute for Leu or Ile, which are in close contact with Phe. Furthermore, core centred positions should be favoured for the introduction of new Leu instead of Met.

When trying to restore the balance of Met and Leu in a Met-depleted core, we could not regenerate wild-type conditions, which would have been indicated by a tightly bound NTD dimer. Surprisingly, the ultra-mutation with twelve-point mutants was still synthesised with high yield and folded seemingly correctly, judging by the secondary structure measured with circular dichroism. Yet, association essays with Trp fluorescence spectroscopy and high-resolution SEC experiments showed that dynamics could not be restored. In turn, a higher dynamical range of the NTD could be induced by introducing only a single Met into the wild type, as shown by PET-FCS measurements. This experiment enforced the conjecture that Met can be used to reshape the functionality of proteins.

Surprisingly, the exchange of the core Mets to Leu of two homologous NTDs showed a weaker effect in stability gain and association capacities, as was seen for the Ea L6-NTD or even the Ea L5-NTD. In short, the effects between homologs should not be compared quantitatively but rather qualitatively. As the numbers of Met in core centre positions vary between the three NTDs, position-dependent conclusions of the Met influence can be made. It seems probable that Met is independent of its position in the core. This seems likely because the Lh-NTD has the Met distributed rather in tertiary interaction with the core, than in the very centre. Of course, this affects newly introduced Leu, which would have greater influence when placed in the core centre. Nonetheless, the qualitative effects of Met enhancing tight binding was observable for all examined homologs.

In order to transfer the concept of methionine residues, engaging dynamically in the N-terminal domain of spider silk, to other proteins, we chose Hsp90 as a target. Consequently, we addressed core leucine residues and mutated them to methionine. Hence, we realised that in case of Hsp90, not many Leu to Met mutations at a time could be made. Changing too many amino acids simultaneously led to a critically unstable protein. Interestingly, we could show that the introduction of two to three Met led to a stably associated Hsp90 dimer with 30 % enhanced function. Even if this seemed to be a small effect, it could indicate that our hypothesis and general understanding of the function of Met (and Leu) in protein cores is not restricted to the spidroin's NTD.

Envisioning the transfer of the idea to malleabelise proteins so that their function is enhanced, may optimise proteins to work better. Proteins that are intrinsically unstable could be stabilised by just a few point mutations, while ones with low binding efficiency could turn to high performance. Ideally, this could be part of a universal applicable mechanism for all proteins and could mean a major step towards targeted protein engineering.

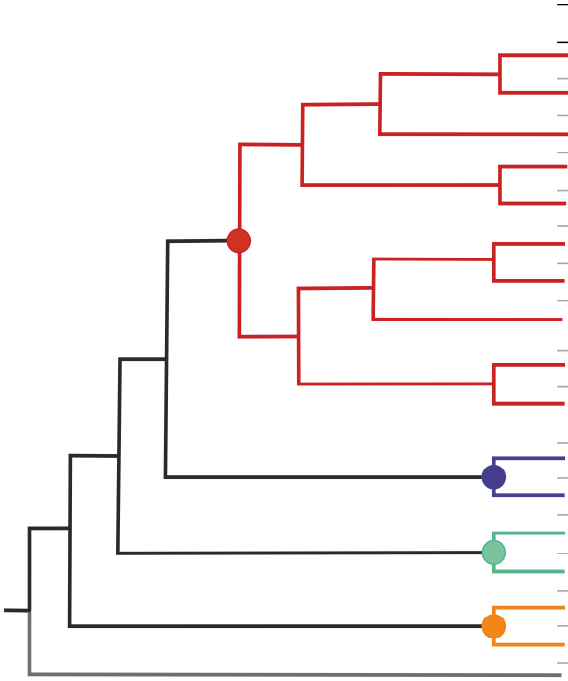
6.Outlook

6.1. Phylogeny

Tensile strength originated early in the history of spider silk evolution and extensibility only later (Garb et al. 2010). In light of our new findings, tensile strength could, therefore, be associated to leucine (Leu) and extensibility with methionine (Met) residues. So far, the repeats in the middle domain were known to be responsible for the strength, but also for extensibility if they contain many glycine-rich segments (Rising and Johansson 2015). However, the terminal domains may also contribute to extensibility (Rat et al. 2018), since artificial spider silk lacking the terminal domains, is less extensible (Rising and Johansson 2015).

During evolution, the frequency of specific amino acids seemed to have increased differently. First, high levels of Leu were present and only then did the number of Met increase in protein sequences (Jordan et al. 2005). Furthermore, the expansion of initially under-represented amino acids continues to this day. The fact that Met was only introduced later in evolution to presumably achieve higher functionality might be of significance for protein engineering today. In order to test this, one would have to introduce new Mets into proteins, as was exemplary shown with Hsp90 (in chapter 4. 7.). Too many mutations at once, however, are usually not tolerated by proteins (Jordan et al. 2005), which also occurred in our Hsp90 experiments. In contrast, the two or three-fold mutants of Hsp90 (M2 and M3) tolerated the mutations well and even became slightly more functional. Given the substantial difference in structure and function between Hsp90 and the NTD, this is a remarkable observation. The trend from Leu to Met corresponds to the evolutionary path that probably took place in spider silk over millions of years. Already one or two mutations should provide a slight advantage to the protein of interest. The proteins might therefore become more elastic or dynamic.

When comparing several homologous sequences (e.g. NTDs, Hsp90s, rhodopsins), it has become evident that the closer the sequences are to each other phylogenetically, the more similar the number of Met/Leu/Ile are (and Met in particular). This led to the idea that Met could possibly be used as a phylogenetic marker: The Met content could help to elucidate genealogy questions in homologous sequences. Consequently, similar "old proteins" should have a similar proportion of Met. New genealogical trees of, e.g. spider species are revised repeatedly based often on morphology and behaviour, DNA sequence or tensile behaviour of spider silk (Kuntner et al. 2008; Garb et al. 2010; Elices et al. 2009). New criteria constantly emerge after which the species relations to each other are classified. Of course, we already know that the more conserved homologous sequences are, the closer they appear in the genealogical tree. This is also valid for the number of amino acids. But the more similar the Met content in the sequences seemed, the more accurate the classification was. I determined the content of Met in relation to the sequences from Garb et al. phylogenetic tree for spidroins. A similar Met-content was corresponding to a close relationship (Fig. 6. 1).



Species	MaSp1	MaSp2	MaSp
Ab	-	4 %	-
Aa	-	3.4 %	-
At	-	3.8 %	-
Av	-	-	-
Mg	-	-	-
Nc	5.1 %	5.4 %	-
Nim	-	6.7 %	-
Ncl	-	-	-
Lg	5.3 %	4.7 %	-
Lh	5.3 %	4.7 %	-
Ud	-	-	-
Ds	-	7.7 %	-
Aap	-	-	6.7 %
Ea	7.3 %	-	-
Dc	-	-	6.2
Kh	4.6 %	-	-
Bc	-	-	-

Fig. 6. 1: Phylogenetic tree of spiders classified by their NTD sequences (adapted from Garb et al. 2010). Representation of methionine residue share (given in per cent) in major ampullate spidroin 1 and 2, NTD. The closer the spider species are related to each other, the stronger the similarity of the Met content, indicated in percentage. Two of the spider species might be incorrectly categorised, based on their Met share (circled in red). Gaps indicated by a line, are not yet sequenced MaSp for this species. Ab *Argiope brunennichi*, Aa *Argiope argentata*, At *Argiope trifasciata*, Av *Araneur ventricosus*, Mg *Metepeira grandiosa*, Nc *Nephila clavipes*, Nim *Nephila inaurata madagascariensis*, Ncl *Nephila clavata*, Lg *Latrodectus geometricus*, Lh *Latrodectur hesperus*, Ud *Uloborus diversus*, Ds *Deinopis spinosa*, Aap *Agelenopsis aperta*, Ea *Euprosthenops australis*, Dc *Diguetia canities*, Kh *Kukulkania hibernalis*, Bc *Bothriocyrtum californicum*.

The advantage of using this particular family tree as a template, is that only NTD sequences were compared. Yet, NTDs originating from different spider glands were mixed for the analysis. Therefore, it seems likely that this classification could change when only MaSp NTDs are compared. To date, not all MaSp are sequenced or all spider species characterised (Fig. 6. 1). Especially the N-terminal domains seem to be more difficult to sequence as the full-length spidroin cDNAs or genomic sequences have up to 15 kb (Garb et al. 2010). Since

it is likely that the sequences from all homologs will remain elusive, at least in the near future, other standard phylogenetic techniques are required. As it seems, two of the spider species in the phylogenetic tree of figure 6. 1 might be incorrectly categorised, when considering their Met share (circled in red). Species in the lower figure, branched in lilac, green and orange, appear to have an increased amount of Met (6.2 - 7.7 %). Yet, *Kukulkania hibernalis* only has a 4.6 % Met share, indicating that this species NTD might be misplaced. This phenomenon seems similar for the *Nephila inaurata madagascariensis*.

It remains to be seen if the validity of this new way of classification can be proven. However, Met can certainly assist in classification. It seems sensible to me to utilise such a categorisation approach since evolution has introduced Met as an essential feature in proteins by a very slow and steady process (Jordan et al. 2005). As anticipated, it would be necessary to analyse other proteins genealogy trees and their related Met content with relevant statistical conditions in order to formulate a general hypothesis.

Consistent with this thought, it is interesting to note that the Met content of rhodopsin proteins seemed to be constant within a category as well. Admittedly, I did not compare many sequences, because this would have gone beyond the scope of this work. Yet, within the four chosen rhodopsin sequences from algae (PDB-codes 6EID, 3UG9, and GenBank numbers AF385748.1, AAM15777.1) and *E. coli* (PDB-codes 1GU8, 5AZD, 4HYJ, 4TL3) the content of Met residues was astonishingly constant (both approximately 2.8 % Met content). For the group of proteins with only few Met (content of 2.5 % or less), this approach of classification is probably not as suited. Nevertheless, there are many other rhodopsin classes that contain an elevated amount of Met, which could be devoted to this study (see further below).

6.2. Methionine residues in other proteins

In an attempt to find more Met-rich proteins, a new article about an algorithm called FoldIt has been published, which predicts the structure of proteins based on amino acid sequences (Koepnick et al. 2019). The program is used by online users, the citizen scientists, to playfully create folded proteins from scratch. There

are fundamental physical and chemical rules implemented in the program to ensure that the new sequences are as similar as possible to real proteins. The only restriction is to create proteins that do not correspond to an already known sequence. The game was well received and has led to further analysis of the *de novo* sequences. The protein structures were resolved by X-ray crystallography and NMR spectroscopy and basic stability measurements were made with the synthesised proteins. The novel sequences can be found in annexe D, table 8.5.

Considering the scope and the results of this thesis, especially regarding the Met-hypothesis, my interest in these newly designed proteins was in the comparison of the sequences and their stability. Interestingly, there is a high proportion of leucine residues in these proteins, a property that has probably simplified the creation of proteins according to the physical laws of the game. The developers argued that in the game, new proteins must have a hydrophobic core, just like most natural proteins (Koepnick et al. 2019). Strikingly, the new proteins, which have many leucine residues, exhibit a higher free enthalpy ΔG, consistent with higher stability. The proteins with fewer leucine residues are less stable, but appear to have been compensated with methionine residues. The sequences for the proteins Foldit1 and Foldit3 are relatively short, while the share of Leu and Ile+Leu is 10 % and 20 %, respectively. Accordingly, these are the two most stable proteins, with ΔG of 12-14 kcal/mol. The two other newly created proteins, Peak6 and F.-Diesel, both have lower Leu/Ile shares with 4-5 % and 2 % Met content, where the latter is, however, not located in their hydrophobic cores (PDB-codes for Foldit1: 6MRR; Foldit3: 6MSP; Peak6: 6MRS; Ferredog-Diesel: 6NUK).

In case of the four FoldIt proteins, the more stable exponents have a leucine content of approximately 10 % in the core. In comparison, the Leu content of Ea WT NTD and Ea L6 NTD is in the same range with 7 % and 11 %, respectively. Judging by the stability data and amino acid percentage of Met/Leu/Ile in other proteins, the bovine rhodopsin (PDB code: 1F88) is very similar to the spider silk NTD (4.1 kcal/mol, 5 % Met, 8 % Leu for BovineRd from Tastan et al. 2007; 5.6 kcal/mol, 7 % Met, 7 % Leu for Ea NTD from Heiby et al. 2017). Rhodopsins belong to the category of photoreceptor proteins, present in visual pigment cells. They are activated when the 11-cis bound retinal isomerises into all-trans upon

light excitation, which allows the stimulation of the optical nerve (Okada et al. 2004). When comparing different rhodopsin species, it is conspicuous that their Met content varies greatly. Human rhodopsin (beta-adrenoreceptor, PDB code: 2RH1) has three Met in core positions and at least three Leu. Channelrhodopsin, from *Chlamydomonas reinhardtii* (single-cell green alga), has one Met in the core in the closed state and only a few leucine residues (PDB code: 3UG9). Another rhodopsin with a high Met content (almost 6 %) occurs in the Japanese flying squid (*Todarodes pacificus*, PDB-code 2Z73), with 15 Mets, many of which are in the core. This makes the rhodopsin a fascinating, Met related, research subject. Both, the bovine and the squid rhodopsin mentioned here, are homo-dimers consisting of α-helix bundles with several Met in the core, like the spider silk NTD. The bovine rhodopsin contains six core Met (sequence positions: 44, 86, 207, 253, 257, 288) as well as six core Leu (sequence positions: 72, 76, 79, 125, 128, 131). However, the protein is a membrane-bound receptor, which complicates the synthesis of the isolated protein. Transmembrane (TM) proteins are prone to aggregation when expressed recombinantly in *E. coli* (Kronqvist et al. 2017). Therefore, the proteins could be expressed alternatively in oocytes of *Xenopus laevis* (African clawed frog). Then, the photocurrents measured after light impulse with the TEVC (two-electrode voltage-clamp) electrophysiological method, could function as a readout for changed protein function. The hypothesis suggests that if new leucine residues are introduced in the protein, it should become more rigid, showing less or no conformational change at all, which would be detectable via the reduced change in voltage. On the other hand, new methionine residues should make the protein more dynamic and possibly increase the function. It would be interesting to understand the fundamental functional principle of proteins such as membrane-associated proteins considering Met since this class of proteins accounts for 20-35 % of the entire proteome (Stevens and Arkin 2000).

6.3. Towards more stable and functionalised proteins

Approximately 15 % of the proteins in a single *E. coli* cell are not very stable, with a free enthalpy of only ~ 4 kcal/mol at 37 °C (Ghosh and Dill 2010). An increase of only 4 K, destabilises up to 20 % of the proteins, which has dramatic consequences for the entire cell. Increased stability is, therefore, beneficial. Hsp90, which helps to fold other proteins correctly, can be impaired as well if the surrounding temperature rises too high. Notably, if Hsp90 fails to perform its key function, many other proteins are also prone to be folded incorrectly. Rapid evolution or the genesis of cancer has been connected to this very phenomenon (Rutherford and Lindquist 1998; Queitsch et al. 2002).

The impact of a designed and targeted synthesis of more functional proteins is self-evident. It will always be of major interest to optimise the performance of proteins originating from drug design or enhance or decrease enzyme activity, amongst others. Even the alteration of applications in material science seems conceivable. Certainly, there will also be cases in which the effect of a particular protein should be weakened, which could also be achievable with the tools mentioned above. Finally, predicting thermal stability based solely on the protein structure has significant theoretical consequences for protein design (Miotto et al. 2019). Ideally, aspects of the results presented in this thesis could be the basis for a significant leap in the understanding between the molecular composition and the function of proteins, therefore facilitating a novel approach to protein design and engineering.

7.Bibliography

Alsallaq, R.; Zhou, H.-X. (2008): Electrostatic rate enhancement and transient complex of protein-protein association. In: *Proteins* 71 (1), p. 320–335. DOI: 10.1002/prot.21679.

Andersson, M.; Chen, G.; Otikovs, M.; Landreh, M.; Nordling, K.; Kronqvist, N. et al. (2014): Carbonic Anhydrase Generates CO2 and H+ That Drive Spider Silk Formation Via Opposite Effects on the Terminal Domains. In: *PLoS Biology* 12 (8), e1001921. DOI: 10.1371/journal.pbio.1001921.

Andersson, M.; Jia, Q.; Abella, A.; Lee, X.-Y.; Landreh, M.; Purhonen, P. et al. (2017): Biomimetic spinning of artificial spider silk from a chimeric minispidroin. In: *Nature chemical biology* 13 (3), p. 262–264. DOI: 10.1038/nchembio.2269.

Andersson, M.; Johansson, J.; Rising, A. (2016): Silk Spinning in Silkworms and Spiders. In: *International Journal of Molecular Sciences* 17 (8). DOI: 10.3390/ijms17081290.

Askarieh, G.; Hedhammar, M.; Nordling, K.; Saenz, A.; Casals, C.; Rising, A. et al. (2010): Self-assembly of spider silk proteins is controlled by a pH-sensitive relay. In: *Nature* 465 (7295), p. 236–238. DOI: 10.1038/nature08962.

Becktel, W. J.; Schellman, J. A. (1987): Protein stability curves. In: *Biopolymers* (26), p. 1859–1877.

Bernstein, H. D.; Poritz, M. A.; Strub, K.; Hoben, P. J.; Brenner, S.; Walter, P. (1989): Model for signal sequence recognition from amino-acid sequence of 54K subunit of signal recognition particle. In: *Nature* 340 (6233), p. 482–486. DOI: 10.1038/340482a0.

Birnboim, H. C.; Doly, J. (1979): A rapid alkaline extraction procedure for screening recombinant plasmid DNA. In: *Nucleic acids research* 7 (6), p. 1513–1523.

Blackledge, T. A.; Hayashi, C. Y. (2006): Silken toolkits: biomechanics of silk fibers spun by the orb web spider Argiope argentata (Fabricius 1775). In: *The Journal of experimental biology* 209, p. 2452–2461. DOI: 10.1242/jeb.02275.

Bowen, C. H.; Dai, B.; Sargent, C. J.; Bai, W.; Ladiwala, P.; Feng, H. et al. (2018): Recombinant Spidroins Fully Replicate Primary Mechanical Properties of Natural Spider Silk. In: *Biomacromolecules* 19 (9), p. 3853–3860. DOI: 10.1021/acs.biomac.8b00980.

Brahms, S.; Brahms, J. (1980): Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism. In: *Journal of Molecular Biology* 138 (2), p. 149–178. DOI: 10.1016/0022-2836(80)90282-X.

Buchmann, S. (2018): Untersuchungen zur Funktion und Stabilität des hydrophoben Kerns eines Spinnenseidenproteins. In: *Bachelor-Thesis*. University of Würzburg.

Chen, B.; Zhong, D.; Monteiro, A. (2006): Comparative genomics and evolution of the HSP90 family of genes across all kingdoms of organisms. In: *BMC genomics* 7, p. 156. DOI: 10.1186/1471-2164-7-156.

Chinali, A.; Vater, W.; Rudakoff, B.; Sponner, A.; Unger, E.; Grosse, F. et al. (2010): Containment of Extended Length Polymorphisms in Silk Proteins. In: *Journal of Molecular Evolution* 70 (4), p. 325–338. DOI: 10.1007/s00239-010-9326-2.

Chung, H.; Kim, T. Y.; Lee, S. Y. (2012): Recent advances in production of recombinant spider silk proteins. In: *Phosphorus biotechnology - Pharmaceutical biotechnology* 23 (6), p. 957–964. DOI: 10.1016/j.copbio.2012.03.013.

Cirino, P. C.; Tang, Y.; Takahashi, K.; Tirrell, D. A.; Arnold, F. H. (2003): Global incorporation of norleucine in place of methionine in cytochrome P450 BM-3 heme domain increases peroxygenase activity. In: *Biotechnology and bioengineering* 83 (6), p. 729–734. DOI: 10.1002/bit.10718.

Clark, David P.; Pazdernik, Nanette J. (eds.) (2013): Molecular Biology. (Second Edition). Boston: Academic Press.

Collin, M. A.; Clarke, T. H.; Ayoub, N. A.; Hayashi, C. Y. (2018): Genomic perspectives of spider silk genes through target capture sequencing: Conservation of stabilization mechanisms and homology-based structural models of spidroin terminal regions. In: *International journal of biological macromolecules* 113, p. 829–840. DOI: 10.1016/j.ijbiomac.2018.02.032.

Cooper, S.; Khatib, F.; Treuille, A.; Barbero, J.; Lee, J.; Beenen, M. et al. (2010): Predicting protein structures with a multiplayer online game. In: *Nature* 466 (7307), p. 756–760. DOI: 10.1038/nature09304.

Csermely, P.; Schnaider, T.; Soti, C.; Prohászka, Z.; Nardai, G. (1998): The 90-kDa Molecular Chaperone Family: Structure, Function, and Clinical Applications. A Comprehensive Review. In: *Pharmacology & Therapeutics* 79 (2), p. 129–168. DOI: 10.1016/S0163-7258(98)00013-8.

de los Rios, M. A.; Plaxco, K. W. (2005): Apparent Debye−Huckel Electrostatic Effects in the Folding of a Simple, Single Domain Protein. In: *Biochemistry* 44 (4), p. 1243–1250. DOI: 10.1021/bi048444l.

Debye, P.; Hückel, E. (eds.) (1923): Zur Theorie der Elektrolyte. Gefrierpunktserniedrigung und verwandte Erscheinungen. In: Schriftleitung der "Naturwissenschaften" Ergebnisse der Exakten Naturwissenschaften. Berlin, Heidelberg: Springer.

Denny, M. (1976): Physical properties of spiders silk and their role in the design of orb-webs. In: *Journal of Experimental Biology* 65 (2), p. 483.

Doose, S.; Neuweiler, H.; Sauer, M. (2009): Fluorescence quenching by photoinduced electron transfer: a reporter for conformational dynamics of macromolecules. In: *Chemphyschem : a European journal of chemical physics and physical chemistry* 10 (9-10), p. 1389–1398. DOI: 10.1002/cphc.200900238.

Doose, S.; Sauer, M. (2007): PET-FCS zum Studium von Konformationsdynamik. In: *Biospektrum* 13, p. 384–385.

Eftink, M. R. (1991): Fluorescence Techniques for Studying Protein Structure. In: *Methods of Biochemical Analysis* (eds. Suelter, C. H. and Suelter, C. H.). DOI: 10.1002/9780470110560.ch3.

Eisoldt, L.; Thamm, C.; Scheibel, T. (2011): The role of terminal domains during storage and assembly of spider silk proteins. In: *Biopolymers* (97 (6)), p. 355–361. DOI: 10.1002/bip.22006.

Elices, M.; Plaza, G. R.; Arnedo, M. A.; Pérez-Rigueiro, J.; Torres, F. G.; Guinea, G. V. (2009): Mechanical behavior of silk during the evolution of orb-web spinning spiders. In: *Biomacromolecules* 10 (7), p. 1904–1910. DOI: 10.1021/bm900312c.

Fischer, E. (ed.) (1906): Synthese von Polypeptiden. In: *Chloride der Aminosäuren und ihrer Acylderivate*. Berlin, Heidelberg: Springer.

Gaines, J. C.; Virrueta, A.; Buch, D. A.; Fleishman, S. J.; O'Hern, C. S.; Regan, L. (2017): Collective repacking reveals that the structures of protein cores are uniquely specified by steric repulsive interactions. In: *Protein engineering, design & selection : PEDS* 30 (5), p. 387–394. DOI: 10.1093/protein/gzx011.

Gaines, W. A.; Sehorn, M. G.; Marcotte, W. R. (2010): Spidroin N-terminal Domain Promotes a pH-dependent Association of Silk Proteins during Self-assembly. In: *Journal of Biological Chemistry* (285 (52)), p. 40745–40753. DOI: 10.1074/jbc.M110.163121.

Garb, J. E.; Ayoub, N. A.; Hayashi, C. Y. (2010): Untangling spider silk evolution with spidroin terminal domains. In: *BMC Evolutionary Biology* 10 (1), p. 243. DOI: 10.1186/1471-2148-10-243.

Garb, J. E.; Haney, R. A.; Schwager, E. E.; Gregorič, M.; Kuntner, M.; Agnarsson, I.; Blackledge, T. A. (2019): The transcriptome of Darwin's bark spider silk glands predicts proteins contributing to dragline silk toughness. In: *Communications biology* 2, p. 275. DOI: 10.1038/s42003-019-0496-1.

Gellman, S. H. (1991): On the role of methionine residues in the sequence-independent recognition of nonpolar protein surfaces. In: *Biochemistry* 30 (27), p. 6633–6636. DOI: 10.1021/bi00241a001.

Gelman, H.; Gruebele, M. (2014): Fast protein folding kinetics. In: *Quarterly reviews of biophysics* 47 (2), p. 95–142. DOI: 10.1017/S003358351400002X.

Ghosh, K.; Dill, K. (2010): Cellular proteomes have broad distributions of protein stability. In: *Biophysical journal* 99 (12), p. 3996–4002. DOI: 10.1016/j.bpj.2010.10.036.

Gilles, A.; Marlière, P.; Rose, T.; Sarfati, R.; Longin, R.; Meier, A. et al. (1988): Conservative Replacement of Methionine by Norleucine in Escherichia coli Adenylate Kinase. In: *The Journal of biological chemistry* (263(17)), p. 8204–8209.

Gnesa, E.; Hsia, Y.; Yarger, J. L.; Weber, W.; Lin-Cereghino, J.; Lin-Cereghino, G. et al. (2012): Conserved C-terminal domain of spider tubuliform spidroin 1 contributes to extensibility in synthetic fibers. In: *Biomacromolecules* 13 (2), p. 304–312. DOI: 10.1021/bm201262n.

Gómez-Tamayo, J. C.; Cordomí, A.; Olivella, M.; Mayol, E.; Fourmy, D.; Pardo, L. (2016): Analysis of the interactions of sulfur-containing amino acids in membrane proteins. In: *Protein science : a publication of the Protein Society* 25 (8), p. 1517–1524. DOI: 10.1002/pro.2955.

Gosline, J. M.; Denny, M. W.; DeMont, M. E. (1984): Spider silk as rubber. In: *Nature* 309 (5968), p. 551–552. DOI: 10.1038/309551a0.

Gosline, J. M.; Guerette, P. A.; Ortlepp, C. S.; Savage, K. N. (1999): The mechanical design of spider silks: From fibroin sequence to mechanical function. In: *Journal of Experimental Biology* (202), p. 3295–3303.

Greenfield, N. J. (2006): Using circular dichroism spectra to estimate protein secondary structure. In: *Nature Protocols* 1 (6), p. 2876–2890. DOI: 10.1038/nprot.2006.202.

Hagn, F.; Eisoldt, L.; Hardy, J. G.; Vendrely, C.; Coles, M.; Scheibel, T.; Kessler, H. (2010): A conserved spider silk domain acts as a molecular switch that controls fibre assembly. In: *Nature* 465 (7295), p. 239–242. DOI: 10.1038/nature08936.

Hagn, F.; Thamm, C.; Scheibel, T.; Kessler, H. (2011): pH-dependent dimerization and salt-dependent stabilization of the N-terminal domain of spider dragline silk--implications for fiber formation. In: *Angewandte Chemie (International ed. in English)* 50 (1), p. 310–313. DOI: 10.1002/anie.201003795.

Hartley, H. (1951): Origin of the Word 'Protein'. In: *Nature* 168 (4267), p. 244. DOI: 10.1038/168244a0.

Hayashi, C. Y.; Shipley, N. H.; Lewis, R. V. (1999): Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins. In: *International journal of biological macromolecules* 24 (2), p. 271–275. DOI: 10.1016/S0141-8130(98)00089-0.

Hedhammar, M.; Rising, A.; Grip, S.; Martinez, A. S.; Nordling, K.; Casals, C. et al. (2008): Structural properties of recombinant nonrepetitive and repetitive parts of major ampullate spidroin 1 from Euprosthenops australis: implications for fiber formation. In: *Biochemistry* 47 (11), p. 3407–3417. DOI: 10.1021/bi702432y.

Heiby, J. C. (2015): Faltung, Dynamik und Funktion der N-terminalen Domäne einer Spinnenseidenprotein-Familie. In: *Master thesis*. University of Würzburg.

Heiby, J. C.; Goretzki, B.; Johnson, C. M.; Hellmich, U. A.; Neuweiler, H. (2019): Methionine in a protein hydrophobic core drives tight interactions required for assembly of spider silk. In: *Nature Communications* 10 (1), p. 4378. DOI: 10.1038/s41467-019-12365-5.

Heiby, J. C.; Rajab, S.; Rat, C.; Johnson, C. M.; Neuweiler, H. (2017): Conservation of folding and association within a family of spidroin N-terminal domains. In: *Scientific reports* 7 (1), p. 16789. DOI: 10.1038/s41598-017-16881-6.

Heim, M.; Keerl, D.; Scheibel, T. (2009): Spider silk: from soluble protein to extraordinary fiber. In: *Angewandte Chemie (International ed. in English)* 48 (20), p. 3584–3596. DOI: 10.1002/anie.200803341.

Henzler-Wildman, K.; Kern, D. (2007): Dynamic personalities of proteins. In: *Nature* 450 (7172), p. 964–972. DOI: 10.1038/nature06522.

Hofmeister, F. (1902): Über Bau und Gruppierung der Eiweisskörper. In: *Ergebnisse der Physiologie* 1 (1), p. 759–802. DOI: 10.1007/BF02323641.

Hoskins, J. R.; Wickner, S.; Doyle, S. M. (2018): Bacterial Hsp90 ATPase Assays. In: *Methods in molecular biology (Clifton, N.J.)* 1709, p. 199–207. DOI: 10.1007/978-1-4939-7477-1_15.

Hu, X.; Vasanthavada, K.; Kohler, K.; McNary, S.; Moore, A. M. F.; Vierra, C. A. (2006): Molecular mechanisms of spider silk. In: *Cellular and Molecular Life Sciences* 63 (17), p. 1986–1999. DOI: 10.1007/s00018-006-6090-y.

Huang, W.; Zhang, Y.; Chen, Y.; Wang, Y.; Yuan, W.; Zhang, N. et al. (2017): From EST to novel spider silk gene identification for production of spidroin-based biomaterials. In: *Scientific reports* 7 (1), p. 13354. DOI: 10.1038/s41598-017-13876-1.

Hulme, E. C.; Trevethick, M. A. (2010): Ligand binding assays at equilibrium: validation and interpretation. In: *British Journal of Pharmacology* 161 (6), p. 1219–1237. DOI: 10.1111/j.1476-5381.2009.00604.x.

Hutchison, C. A.; Phillips, S.; Edgell, M. H.; Gillam, S.; Jahnke, P.; Smith, M. (1978): Mutagenesis at a specific position in a DNA sequence. In: *The Journal of biological chemistry* 253 (18), p. 6551–6560.

Inoue, H.; Nojima, H.; Okayama, H. (1990): High efficiency transformation of Escherichia coli with plasmids. In: *Gene* 96 (1), p. 23–28. DOI: 10.1016/0378-1119(90)90336-P.

Jackson, S. E. (1998): How do small single-domain proteins fold? In: *Folding and Design* 3 (4), R81-R91. DOI: 10.1016/S1359-0278(98)00033-9.

Jackson, S. E.; Fersht, A. R. (1991): Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. In: *Biochemistry* 30 (43), p. 10428–10435. DOI: 10.1021/bi00107a010.

Jaudzems, K.; Askarieh, G.; Landreh, M.; Nordling, K.; Hedhammar, M.; Jörnvall, H. et al. (2012): pH-Dependent Dimerization of Spider Silk N-Terminal Domain Requires Relocation of a Wedged Tryptophan Side Chain. In: *Journal of Molecular Biology* 422 (422), p. 477–487. DOI: 10.1016/j.jmb.2012.06.004.

Jensen, M. H.; Sukumaran, M.; Johnson, C. M.; Greger, I. H.; Neuweiler, H. (2011): Intrinsic Motions in the N-Terminal Domain of an Ionotropic Glutamate Receptor Detected by Fluorescence Correlation Spectroscopy. In: *Journal of Molecular Biology* 414 (1), p. 96–105. DOI: 10.1016/j.jmb.2011.09.037.

Jordan, I. K.; Kondrashov, F. A.; Adzhubei, I. A.; Wolf, Y. I.; Koonin, E. V.; Kondrashov, A. S.; Sunyaev, S. (2005): A universal trend of amino acid gain and loss in protein evolution. In: *Nature* 433 (7026), p. 633–638. DOI: 10.1038/nature03306.

Kelly, S. M.; Price, N. C. (2000): The Use of Circular Dichroism in the Investigation of Protein Structure and Function. In: *Current Protein & Peptide Science* 1 (4), p. 349–384. DOI: 10.2174/1389203003381315.

Kim, J.; Doose, S.; Neuweiler, H.; Sauer, M. (2006): The initial step of DNA hairpin folding: a kinetic analysis using fluorescence correlation spectroscopy. In: *Nucleic acids research* 34 (9), p. 2516–2527. DOI: 10.1093/nar/gkl221.

Kluge, J. A.; Rabotyagova, O.; Leisk, G. G.; Kaplan, D. L. (2008): Spider silks and their applications. In: *Trends in Biotechnology* 26 (5), p. 244–251. DOI: 10.1016/j.tibtech.2008.02.006.

Koepnick, B.; Flatten, J.; Husain, T.; Ford, A.; Silva, D.-A.; Bick, M. J. et al. (2019): De novo protein design by citizen scientists. In: *Nature* 570 (7761), p. 390–394. DOI: 10.1038/s41586-019-1274-4.

Koeppel, A.; Holland, C. (2017): Progress and Trends in Artificial Silk Spinning. A Systematic Review. In: *ACS biomaterials science & engineering* 3 (3), p. 226–237. DOI: 10.1021/acsbiomaterials.6b00669.

Kozak, M. (1999): Initiation of translation in prokaryotes and eukaryotes. In: *Gene* 234 (2), p. 187–208. DOI: 10.1016/S0378-1119(99)00210-3.

Krichevsky, O.; Bonnet, G. (2002): Fluorescence correlation spectroscopy: the technique and its applications. In: *Reports on Progress in Physics* 65 (2), p. 251–297. DOI: 10.1088/0034-4885/65/2/203.

Kronqvist, N.; Otikovs, M.; Chmyrov, V.; Chen, G.; Andersson, M.; Nordling, K. et al. (2014): Sequential pH-driven dimerization and stabilization of the N-terminal domain enables rapid spider silk formation. In: *Nature Communications. DOI:* 10.1038/ncomms4254.

Kronqvist, N.; Sarr, M.; Lindqvist, A.; Nordling, K.; Otikovs, M.; Venturi, L. et al. (2017): Efficient protein production inspired by how spiders make silk. In: *Nature Communications* 8, p. 15504. DOI: 10.1038/ncomms15504.

Kubelka, J.; Hofrichter, J.; Eaton, W. A. (2004): The protein folding 'speed limit'. In: *Current opinion in structural biology* 14 (1), p. 76–88. DOI: 10.1016/j.sbi.2004.01.013.

Kuntner, M.; Coddington, J. A.; Hormiga, G. (2008): Phylogeny of extant nephilid orb-weaving spiders (Araneae, Nephilidae). Testing morphological and ethological homologies. In: *Cladistics* 24 (2), p. 147–217. DOI: 10.1111/j.1096-0031.2007.00176.x.

Kyte, J.; Doolittle, R. F. (1982): A simple method for displaying the hydropathic character of a protein. In: *Journal of Molecular Biology* 157 (1), p. 105–132. DOI: 10.1016/0022-2836(82)90515-0.

Levine, R. L.; Moskovitz, J.; Stadtman, E. R. (2000): Oxidation of Methionine in Proteins: Roles in Antioxidant Defense and Cellular Regulation. In: *IUBMB Life* 50 (4-5), p. 301–307. DOI: 10.1080/713803735.

Levinthal, C. (1969): How to fold graciously. transcript, 1969.

Lim, N. C. H.; Jackson, S. E. (2015): Mechanistic insights into the folding of knotted proteins in vitro and in vivo. In: *Journal of Molecular Biology* 427 (2), p. 248–258. DOI: 10.1016/j.jmb.2014.09.007.

Liu, H.; Naismith, J. H. (2008): An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. In: *BMC biotechnology* 8, p. 91. DOI: 10.1186/1472-6750-8-91.

Lum, J. K.; Neuweiler, H.; Fersht, A. R. (2012): Long-Range Modulation of Chain Motions within the Intrinsically Disordered Transactivation Domain of Tumor Suppressor p53. In: *J. Am. Chem. Soc.* 134 (3), p. 1617–1622. DOI: 10.1021/ja2078619.

Madurga, R.; Blackledge, T. A.; Perea, B.; Plaza, G. R.; Riekel, C.; Burghammer, M. et al. (2015): Persistence and variation in microstructural design during the evolution of spider silk. In: *Scientific reports* 5 (14820). DOI: 10.1038/srep14820.

Mallam, A. L.; Jackson, S. E. (2007): The dimerization of an alpha/beta-knotted protein is essential for structure and function. In: *Structure (London, England : 1993)* 15 (1), p. 111–122. DOI: 10.1016/j.str.2006.11.007.

Marmé, N.; Knemeyer, J.-P.; Sauer, M.; Wolfrum, J. (2003): Inter- and Intramolecular Fluorescence Quenching of Organic Dyes by Tryptophan. In: *Bioconjugate Chem* 14 (6), p. 1133–1139. DOI: 10.1021/bc0341324.

Miotto, M.; Olimpieri, P. P.; Di Rienzo, L.; Ambrosetti, F.; Corsi, P.; Lepore, R. et al. (2019): Insights on protein thermal stability: a graph representation of molecular interactions. In: *Bioinformatics (Oxford, England)* 35 (15), p. 2569–2577. DOI: 10.1093/bioinformatics/bty1011.

Mita, K.; Ichimura, S.; James, T. C. (1994): Highly repetitive structure and its organization of the silk fibroin gene. In: *Journal of Molecular Evolution* 38 (6), p. 583–592. DOI: 10.1007/BF00175878.

Mulder, G. J.; Miquel, F. A.; Wenckebach, W. (1838): Sur la composition de quelques substances animales. In: *Bulletin des sciences physiques et naturelles en néerlande.* Natural History Museum Library.

Myers, J. K.; Pace, C. N.; Scholtz, J. M. (1995): Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. In: *Protein Science* (4), p. 2138–2148.

Neuweiler, H.; Banachewicz, W.; Fersht, A. R. (2010): Kinetics of chain motions within a protein-folding intermediate. In: *Proceedings of the National Academy of Sciences of the USA* 107 (51), p. 22106. DOI: 10.1073/pnas.1011666107.

Neuweiler, H.; Doose, S.; Sauer, M. (2005): A microscopic view of miniprotein folding: Enhanced folding efficiency through formation of an intermediate. In: *Proceedings of the National Academy of Sciences of the USA* 102 (46), p. 16650. DOI: 10.1073/pnas.0507351102.

Neuweiler, H.; Löllmann, M.; Doose, S.; Sauer, M. (2007): Dynamics of Unfolded Polypeptide Chains in Crowded Environment Studied by Fluorescence Correlation Spectroscopy. In: *Journal of Molecular Biology* 365 (3), p. 856–869. DOI: 10.1016/j.jmb.2006.10.021.

Neuweiler, H.; Schulz, A.; Böhmer, M.; Enderlein, J.; Sauer, M. (2003): Measurement of Submicrosecond Intramolecular Contact Formation in Peptides at the Single-Molecule Level. In: *Journal of the American Chemical Society* 125 (18), p. 5324–5330. DOI: 10.1021/ja034040p.

Nicholson, E. M.; Scholtz, J. M. (1996): Conformational Stability of the Escherichia coli HPr Protein: Test of the Linear Extrapolation Method and a Thermodynamic Characterization of Cold Denaturation. In: *Biochemistry* 35 (35), p. 11369–11378. DOI: 10.1021/bi960863y.

Nie, S.; Zare, R. N. (1997): Optical detection of single molecules. In: *Annual Review of Biophysics and Biomolecular Structure* 26 (1), p. 567–596. DOI: 10.1146/annurev.biophys.26.1.567.

Okada, T.; Sugihara, M.; Bondar, A.-N.; Elstner, M.; Entel, P.; Buss, V. (2004): The retinal conformation and its environment in rhodopsin in light of a new 2.2 A crystal structure. In: *Journal of Molecular Biology* 342 (2), p. 571–583. DOI: 10.1016/j.jmb.2004.07.044.

O'Neil, K. T.; DeGrado, W. F. (1990): How calmodulin binds its targets: sequence independent recognition of amphiphilic α-helices. In: *Trends in Biochemical Sciences* 15 (2), p. 59–64. DOI: 10.1016/0968-0004(90)90177-D.

OriginPro 2016G. Version OriginPro 2016G: OriginLab Corporation, Northampton, MA, USA.

Otikovs, M.; Chen, G.; Nordling, K.; Landreh, M.; Meng, Q.; Jörnvall, H. et al. (2015): Diversified Structural Basis of a Conserved Molecular Mechanism for pH-Dependent Dimerization in Spider Silk N-Terminal Domains. In: *ChemBioChem: a European journal of chemical biology* 16 (12), p. 1720–1724. DOI: 10.1002/cbic.201500263.

Pal, D.; Chakrabarti, P. (2001): Non-hydrogen bond interactions involving the methionine sulfur atom. In: *Journal of biomolecular structure & dynamics* 19 (1), p. 115–128. DOI: 10.1080/07391102.2001.10506725.

Pan, Y.; Sackmann, E. K.; Wypisniak, K.; Hornsby, M.; Datwani, S. S.; Herr, A. E. (2016): Determination of equilibrium dissociation constants for recombinant antibodies by high-throughput affinity electrophoresis. In: *Scientific reports* 6, p. 39774. DOI: 10.1038/srep39774.

Pérez-Rigueiro, J.; Madurga, R.; Gañán-Calvo, A. M.; Plaza, G. R.; Elices, M.; López, P. A. et al. (2018): Straining Flow Spinning of Artificial Silk Fibers: A Review. In: *Biomimetics* 3 (4). DOI: 10.3390/biomimetics3040029.

Queitsch, C.; Sangster, T. A.; Lindquist, S. (2002): Hsp90 as a capacitor of phenotypic variation. In: *Nature* 417 (6889), p. 618–624. DOI: 10.1038/nature749.

Radchenko, D. S.; Kattge, S.; Kara, S.; Ulrich, A. S.; Afonin, S. (2016): Does a methionine-to-norleucine substitution in PGLa influence peptide-membrane interactions? In: *Biochimica et biophysica acta* 1858 (9), p. 2019–2027. DOI: 10.1016/j.bbamem.2016.06.002.

Radtke, C. (2016): Natural Occurring Silks and Their Analogues as Materials for Nerve Conduits. In: *International Journal of Molecular Sciences* 17 (10). DOI: 10.3390/ijms17101754.

Rajab, S. (2015): Untersuchung zur Faltung und Funktion der N-terminalen Domäne eines Spinnenseidenproteins der Braunen Witwe. In: *Bachelor thesis*. University of Würzburg.

Rat, C. (2015): Untersuchung zu Faltung und Funktion der N-terminalen Domäne eines Spinnenseidenproteins der Goldenen Seidenspinne. In: *Bachelor-Thesis*. University of Würzburg.

Rat, C. (2018): Untersuchungen zur Faltung und Assoziation der C-terminalen Domäne eines Spinnenseidenproteins. In: *Master thesis*. University of Würzburg.

Rat, C.; Heiby, J. C.; Bunz, J. P.; Neuweiler, H. (2018): Two-step self-assembly of a spider silk molecular clamp. In: *Nature Communications* 9 (1), p. 4779. DOI: 10.1038/s41467-018-07227-5.

Rehm, D.; Weller, A. (1969): Kinetik und Mechanismus der Elektronübertragung bei der Fluoreszenzlöschung in Acetonitril. In: *Berichte der Bunsengesellschaft für physikalische Chemie* 73 (8-9), p. 834–839. DOI: 10.1002/bbpc.19690730818.

Richter, K.; Muschler, P.; Hainzl, O.; Buchner, J. (2001): Coordinated ATP hydrolysis by the Hsp90 dimer. In: *The Journal of biological chemistry* 276 (36), p. 33689–33696. DOI: 10.1074/jbc.M103832200.

Ries, J.; Schwarze, S.; Johnson, C. M.; Neuweiler, H. (2014): Microsecond folding and domain motions of a spider silk protein structural switch. In: *Journal of the American Chemical Society* 136 (49), p. 17136–17144. DOI: 10.1021/ja508760a.

Rising, A.; Johansson, J. (2015): Toward spinning artificial spider silk. In: *Nature chemical biology* 11 (5), p. 309–315. DOI: 10.1038/nchembio.1789.

Rutherford, S. L.; Lindquist, S. (1998): Hsp90 as a capacitor for morphological evolution. In: *Nature* 396 (6709), p. 336–342. DOI: 10.1038/24550.

Sadava, D.; Hillis, D. M.; Heller, H. C.; Hacker, S. D.; Markl, J. (2019): Purves Biologie. Berlin, Heidelberg: Springer Berlin Heidelberg.

Sanggaard, K. W.; Bechsgaard, J. S.; Fang, X.; Duan, J.; Dyrlund, T. F.; Gupta, V. et al. (2014): Spider genomes provide insight into composition and evolution of venom and silk. In: *Nature Communications* 5, p. 3765. DOI: 10.1038/ncomms4765.

Santoro, M. M.; Bolen, D. W. (1988): Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. In: *Biochemistry* 27 (21), p. 8063–8068. DOI: 10.1021/bi00421a014.

Sauer, M.; Neuweiler, H. (2014): PET-FCS: probing rapid structural fluctuations of proteins and nucleic acids by single-molecule fluorescence quenching. In: *Methods in molecular biology (Clifton, N.J.)* 1076, p. 597–615. DOI: 10.1007/978-1-62703-649-8_27.

Schacht, K.; Jüngst, T.; Schweinlin, M.; Ewald, A.; Groll, J.; Scheibel, T. (2015): Dreidimensional gedruckte, zellbeladene Konstrukte aus Spinnenseide. In: *Angewandte Chemie* 127 (9), p. 2858–2862. DOI: 10.1002/ange.201409846.

Schreiber, G.; Haran, G.; Zhou, H.-X. (2009): Fundamental aspects of protein-protein association kinetics. In: *Chemical reviews* 109 (3), p. 839–860. DOI: 10.1021/cr800373w.

Schulze, A.; Beliu, G.; Helmerich, D. A.; Schubert, J.; Pearl, L. H.; Prodromou, C.; Neuweiler, H. (2016): Cooperation of local motions in the Hsp90 molecular chaperone ATPase mechanism. In: *Nature chemical biology* 12 (8), p. 628–635. DOI: 10.1038/nchembio.2111.

Schwarze, S.; Zwettler, F. U.; Johnson, C. M.; Neuweiler, H. (2013): The N-terminal domains of spider silk proteins assemble ultrafast and protected from charge screening. In: *Nature Communications. DOI:* 10.1038/ncomms3815.

Stark, M.; Grip, S.; Rising, A.; Hedhammar, M.; Engström, W.; Hjälm, G.; Johansson, J. (2007): Macroscopic Fibers Self-Assembled from Recombinant Miniature Spider Silk Proteins. In: *Biomacromolecules* 8 (5), p. 1695–1701. DOI: 10.1021/bm070049y.

Stevens, T. J.; Arkin, I. T. (2000): Do more complex organisms have a greater proportion of membrane proteins in their genomes? In: *Proteins* 39 (4), p. 417–420. DOI: 10.1002/(SICI)1097-0134(20000601)39:4<417::AID-PROT140>3.0.CO;2-Y.

Tanford, C. (1968): Protein denaturation. In: *Advances in Protein Chemistry* (23), p. 121–282.

Tanford, C. (1970): Protein denaturation. In: *Advances in Protein Chemistry* (24), p. 1–95. DOI: 10.1016/S0065-3233(08)60241-7.

Tastan, O.; Yu, E.; Ganapathiraju, M.; Aref, A.; Rader, A. J.; Klein-Seetharaman, J. (2007): Comparison of Stability Predictions and Simulated Unfolding of Rhodopsin Structures†. In: *Photochemistry and Photobiology* 83 (2), p. 351–363. DOI: 10.1562/2006-06-20-RA-942.

Teufel, D. P.; Johnson, C. M.; Lum, J. K.; Neuweiler, H. (2011): Backbone-Driven Collapse in Unfolded Protein Chains. In: *Journal of Molecular Biology* 409 (2), p. 250–262. DOI: 10.1016/j.jmb.2011.03.066.

The PyMOL Molecular Graphics System. Version Version 2.0 Schrödinger: LLC.

Tokareva, O.; Michalczechen-Lacerda, V. A.; Rech, E. L.; Kaplan, D. L. (2013): Recombinant DNA production of spider silk proteins. In: *Microbial Biotechnology* 6 (6), p. 651–663. DOI: 10.1111/1751-7915.12081.

Valley, C. C.; Cembran, A.; Perlmutter, J. D.; Lewis, A. K.; Labello, N. P.; Gao, J.; Sachs, J. N. (2012): The methionine-aromatic motif plays a unique role in stabilizing protein structure. In: *The Journal of biological chemistry* 287 (42), p. 34979–34991. DOI: 10.1074/jbc.M112.374504.

Venyaminov, S. Y.; Vassilenko, K. S. (1994): Determination of Protein Tertiary Structure Class from Circular Dichroism Spectra. In: *Analytical Biochemistry* 222 (1), p. 176–184. DOI: 10.1006/abio.1994.1470.

Vepari, C.; Kaplan, D. L. (2007): Silk as a biomaterial. In: *Polymers in Biomedical Applications* 32 (8), p. 991–1007. DOI: 10.1016/j.progpolymsci.2007.05.013.

Villar, H. O.; Koehler, R. T. (2000): Amino acid preferences of small, naturally occurring polypeptides. In: *Biopolymers* 53 (3), p. 226–232.

Vollrath, F. (2000): Strength and structure of spiders' silks. In: *Reviews in Molecular Biotechnology* 74 (2), p. 67–83. DOI: 10.1016/S1389-0352(00)00006-4.

Vollrath, F.; Knight, D. P. (1999): Structure and function of the silk production pathway in the Spider Nephila edulis. In: *International journal of biological macromolecules* 24 (2), p. 243–249. DOI: 10.1016/S0141-8130(98)00095-6.

Vollrath, F.; Knight, D. P. (2001): Liquid crystalline spinning of spider silk. In: *Nature* (410), p. 541–548.

Wiegel, J. (1990): Temperature spans for growth: Hypothesis and discussion. In: *FEMS Microbiology Letters* 75 (2-3), p. 155–169. DOI: 10.1111/j.1574-6968.1990.tb04092.x.

Xia, X.-X.; Qian, Z.-G.; Ki, C. S.; Park, Y. H.; Kaplan, D. L.; Lee, S. Y. (2010): Native-sized recombinant spider silk protein produced in metabolically engineered Escherichia coli results in a strong fiber. In: *Proceedings of the National Academy of Sciences of the USA* 107 (32), p. 14059. DOI: 10.1073/pnas.1003366107.

Xia, Y.; Chu, W.; Qi, Q.; Xun, L. (2015): New insights into the QuikChange™ process guide the use of Phusion DNA polymerase for site-directed mutagenesis. In: *Nucleic acids research* 43 (2), e12. DOI: 10.1093/nar/gku1189.

Xu, H.; Xiao, T.; Chen, C.-H.; Li, W.; Meyer, C. A.; Wu, Q. et al. (2015): Sequence determinants of improved CRISPR sgRNA design. In: *Genome research* 25 (8), p. 1147–1157. DOI: 10.1101/gr.191452.115.

Yuan, T.; Weljie, A. M.; Vogel, H. J. (1998): Tryptophan Fluorescence Quenching by Methionine and Selenomethionine Residues of Calmodulin: Orientation of Peptide and Protein Binding. In: *Biochemistry* 37 (9), p. 3187–3195. DOI: 10.1021/bi9716579.

Zeng, F.; Zhang, S.; Hao, Z.; Duan, S.; Meng, Y.; Li, P. et al. (2018): Efficient strategy for introducing large and multiple changes in plasmid DNA. In: *Scientific reports* 8 (1), p. 1714. DOI: 10.1038/s41598-018-20169-8.

Zhang, F. (2019): Development of CRISPR-Cas systems for genome editing and beyond. In: *Quarterly reviews of biophysics* 52, p. 653. DOI: 10.1017/S0033583519000052.

Zhang, X.; Xia, L.; Day, B. A.; Harris, T. I.; Oliveira, P.; Knittel, C. et al. (2019): CRISPR/Cas9 Initiated Transgenic Silkworms as a Natural Spinner of Spider Silk. In: *Biomacromolecules* 20 (6), p. 2252–2264. DOI: 10.1021/acs.biomac.9b00193.

Zheng, X.; Bi, C.; Li, Z.; Podariu, M.; Hage, D. S. (2015): Analytical methods for kinetic studies of biological interactions: A review. In: *Journal of Pharmaceutical and Biomedical Analysis* 113, p. 163–180. DOI: 10.1016/j.jpba.2015.01.042.

Zwanzig, R.; Szabo, A.; Bagchi, B. (1992): Levinthal's paradox. In: *Proceedings of the National Academy of Sciences of the USA* 89 (1), p. 20–22. DOI: 10.1073/pnas.89.1.20.

8. Appendix

Annexe A: Protein sequences

Sequence 8. 1: *E. australis* MaSp1-NTD

GSGNSHTTPW TNPGLAENFM NSFMQGLSSM PGFTASQLDD MSTIAQSMVQ
SIQSLAAQGR TSPNKLQALN MAFASSMAEI AASEEGGGSL STKTSSIASA
MSNAFLQTTG VVNQPFINEI TQLVSMFAQA GMNDVSA

Number of amino acids: 137; MW: 14.2 kDa; Total number of charged residues: 11; Ext.-coefficient: 5 500 $M^{-1}\,cm^{-1}$; (calculated with ProtParam)

Sequence 8. 2: *L. hesperus* MaSp1-NTD

GSLAQANTPW SSKANADAFI NSFISAASNT GSFSQDQMED MSLIGNTLMA
AMDNMGGRIT PSKLQALDMA FASSVAEIAA SEGGDLGVTT NAIADALTSA
FYQTTGVVNS RFISEIRSLI GMFAQASAND VYA

Number of amino acids: 133; MW: 13.7 kDa; Total number of charged residues: 17; Ext.-coefficient: 8 480 $M^{-1}\,cm^{-1}$; (calculated with ProtParam)

Sequence 8. 3: *L. geometricus* MaSp1-NTD

GSLGQANTPW SSKQNADAFI SAFMTAASQS GAFSSDQIDD MSVISNTLMA
AMDNMGGRIT PSKLQALDMA FASSVAEIAA VEGQNIGVTT NAISDALTSA
FYQTTGVVNN KFISEIRSLI NMFAQASAND VYS

Number of amino acids: 133; MW: 13.9 kDa; Total number of charged residues: 16; Ext.-coefficient: 8 480 $M^{-1}\,cm^{-1}$; (calculated with ProtParam)

Sequence 8. 4: *N. clavipes* MaSp1-NTD

GSAQGQNTPW SSTELADAFI NAFMNEAGRT GAFTADQLDD MSTIGDTIKT
AMDKMARSNK SSKGKLQALN MAFASSMAEI AAVEQGGLSV DAKTNAIADS
LNSAFYQTTG AANPQFVNEI RSLINMFAQS SANEVSY

Number of amino acids: 137; MW: 14.4 kDa; Total number of charged residues: 23; Ext.-coefficient: 8 480 $M^{-1}\,cm^{-1}$; (calculated with ProtParam)

Sequence 8. 5: *E. australis* MaSp1-CTD

GNSGIQGYGQ SSASASAAAS AASTVANSVS RLSSPSAVSR VSSAVSSLVS
NGQVNMAALP NIISNISSSV SASAPGASGC EVIVQALLEV ITALVQIVSS
SSVGYINPSA VNQITNVVAN AMAQVMG

Number of amino acids: 127; MW: 12.2 kDa; Total number of charged residues: 4; Ext.-coefficient: 2 980 $M^{-1}\,cm^{-1}$; (calculated with ProtParam)

Sequence 8. 6: Mini-Spidroin *E. australis* MaSp1 NTD-rep4-CTD

GSGNSHTTPW TNPGLAENFM NSFMQGLSSM PGFTASQLDD MSTIAQSMVQ
SIQSLAAQGR TSPNKLQALN MAFASSMAEI AASEEGGGSL STKTSSIASA
MSNAFLQTTG VVNQPFINEI TQLVSMFAQA GMNDVSAGGS GNSGIQGQGG
YGGLGQGGYG QGAGSSAAAA AAAAAAAAGG QGGQGQGGYG QGSGGSAAAA
AAAAAAAAAA AGRGQGGYGQ GSGGNAAAAA AAAAAAAAAA GQGGQGGYGR
QSQGAGSAAA AAAAAAAAAA AGSGQGGYGG QGQGGYGQSG SEFGNSGIQG
YGQSSASASA AASAASTVAN SVSRLSSPSA VSRVSSAVSS LVSNGQVNMA
ALPNIISNIS SSVSASAPGA SGCEVIVQAL LEVITALVQI VSSSSVGYIN
PSAVNQITNV VANAMAQVMG

Number of amino acids: 420; MW: 39.1 kDa; Total number of charged residues: 18; Ext.-coefficient: 18 910 $M^{-1}\,cm^{-1}$; (calculated with ProtParam)

Sequence 8. 7: Mini-Spidroin *L. hesperus* MaSp1 NTD-rep17-CTD

TQSLYAL AQANTPWSSK ANADAFINSF ISAASNTGSF SQDQMEDMSL
IGNTLMAAMD NMGGRITPSK LQALDMAFAS SVAEIAASEG GDLGVTTNAI
ADALTSAFYQ TTGVVNSRFI SEIRSLIGMF AQASANDVYA SAGSSGGGGY
GASSASAASA SAAAPSGVAY QAPAQAQISF TLRGQQPVSY GQGGAGPGGA
GAAAAAAAAA GGAGQGGQGG YGQGGYGQGG AGQGGSGAAA AAAAAAGGTG
QGGAGQGGAG AAAAAAAAAG GAGQGGQGGY GQGGYGQGGT GQGGAGAAAA
AAAAGGAGQG GQGGYGQGGY GQGGYGQGGS GAAAAAAAAA GGAGQGGQGG
YGQGGYGQGG AGQGGAGAAA AAAAAAGGAG QGGYGRGGAG QGGAAAAAAA
AAGAGQGGYG GQGAGQGGSG AAAAAAAAGG AGQGGQGGYG QGGYGQGGSG
AAAAAAAAGG AGQGGQGGYG QGGYGQGGAG QGGAGAAAAA AAAGGAGQGG
QGGYGQGGYG QGGAGQGGAG AAAAAAAAGG AGQGGQGGYG QGGYGQGGAG
QGGAGAAAAA AAAGGAGQGG QGGYGQGGYG QGGAGQGGAA AAAAAAAGGA
GQGGYGRGGA GQGGAAAAAG AGQGGYGGQG AGQGGAGAAA AAAAAGGAGQ
GGQGGYGRGG YGQGGAGQGG AGAAAAAAAA GGAGQGGQGG YGQGGYGQGG
AGQGGAAAAA AAAGGAGQGG YGRGGAGQGG AAAAAAAAAG AGQGGYGGQG
AGQGGAGAAA AAAAASGPGQ IYYGPQSVAA PAAAAASALA APATSARISS
HASALLSNGP TNPASISNVI SNAVSQISSS NPGASACDVL VQALLELVTA
LLTIIGSSNI GSVNYDSSGQ YAQVVTQSVQ NAFA

Number of amino acids: 881; MW: 74.7 kDa; Total number of charged residues: 26; Ext.-coefficient: 63 610 $M^{-1}\,cm^{-1}$; (calculated with ProtParam)

Sequence 8. 8: Mini-Spidroin *L. hesperus* MaSp2 NTD-rep17-CTD

ALGQANT PWSSKENADA FIGAFMNAAS QSGAFSSDQI DDMSVISNTL
MAAMDNMGGR ITQSKLQALD MAFASSVAEI AVADGQNVGA ATNAISDALR
SAFYQTTGVV NNQFITGISS LIGMFAQVSG NEVSYSSAGS SSAAASEAVS
AGQGPAAQPV YAPSGASAAA AAASGAAPAI QQAYERGGSG SAAAAAGSGP
SGYGQGAGGP GGAGAAAGAA AAGGSGPGGY GQGPAAYGPS GPSGQQGYGP
GGSGAAAAAA AAAGSGPSGY GPGAGGPGGA GAAAAAAAAG GSGPGGYGQG
QASYGPSGPS GQQGYGPGGS GAAAAAAAAA GSGPSGYGPG AAAAAAAGSA
GPGTQQGYGP GGSGAAAAAG SGPRGYGPRG PGGAGAAATA ARGSGPGGYG
QGPAGYGTSG PSRQQGYGPG GSGAAAAAAA AAGGAGPGRQ QGYGPGGSGA
AAATAAGGPG YVGQQRYGPG GAGAAAAAAA GSAGPSRQQA YGPGGSGPAA
ATAAAGSGPS GYGPGASGPV GADAAAAAAT GSAGPGRQQA YGPGESGAAA
AAASGAGPGR QLGYGPGGSG AAAAAAAGGP GYGGQQGYGP GGAGAAAAAA
AGGAGPGRQQ TYGPGGSGAA ATAAGGSGPG GYGQGPSGYG PSGPGGQQGY
GPGGSGAAAA AAAGEAGPGR QQGYGPRGSG AAAAAAAGGP GYGGQSGYGP
GGAGAAAAAA AGGAGPGRQQ EYGPGGSGAA AAAAAAAGSG PSGYGPGAAG
PIGPGGAGAA AAGGSGPVGY GQGPSGYGAS GTGGEQDYGP GGSGAAAAAA
AAASGSGGYG PSQYVPSSVA SSAASAASAL SSPTTHARIS SHASTLLSSG
PTNAAALSNV ISNAVSQVSA SNPGSSSCDV LVQALLEIIT ALISILDSSS
VGQVNYGSSG QYAQIVGQSM QQAMG

Number of amino acids: 922; MW: 80.3 kDa; Total number of charged residues: 41; Ext.-coefficient: 69 570 $M^{-1}\,cm^{-1}$; (calculated with ProtParam)

Sequence 8. 9: Full-length Hsp90 (*Saccharomyces cerevisiae*)

GSHHHHHHGM ASETFEFQAE ITQLMSLIIN TVYSNKEIFL RELISNASDA
LDKIRYKSLS DPKQLETEPD LFIRITPKPE QKVLEIRDSG IGMTKAELIN
NLGTIAKSGT KAFMEALSAG ADVSMIGQFG VGFYSLFLVA DRVQVISKSN
DDEQYIWESN AGGSFTVTLD EVNERIGRGT ILRLFLKDDQ LEYLEEKRIK
EVIKRHSEFV AYPIQLVVTK EVEKEVPIPE EEKKDEEKKD EEKKDEDDKK
PKLEEVDEEE EKKPKTKKVK EEVQEIEELN KTKPLWTRNP SDITQEEYNA
FYKSISNDWE DPLYVKHFSV EGQLEFRAIL FIPKRAPFDL FESKKKKNNI
KLYVRRVFIT DEAEDLIPEW LSFVKGVVDS EDLPLNLSRE MLQQNKIMKV
IRKNIVKKLI EAFNEIAEDS EQFEKFYSAF SKNIKLGVHE DTQNRAALAK
LLRYNSTKSV DELTSLTDYV TRMPEHQKNI YYITGESLKA VEKSPFLDAL
KAKNFEVLFL TDPIDEYAFT QLKEFEGKTL VDITKDFELE ETDEEKAERE
KEIKEYEPLT KALKEILGDQ VEKVVVSYKL LDAPAAIRTG QFGWSANMER
IMKAQALRDS SMSSYMSSKK TFEISPKSPI IKELKKRVDE GGAQDKTVKD
LTKLLYETAL LTSGFSLDEP TSFASRINRL ISLGLNIDED EETETAPEAS
TAAPVEEVPA DTEMEEVDA

Number of amino acids: 719; MW: 82.5 kDa; Total number of charged residues: 246; Ext.-coefficient: 57 300 $M^{-1}\,cm^{-1}$; (calculated with ProtParam)

Annexe B: Nucleotide-sequences of the Primers used with PCR

Table 8. 1: Primers for cysteine modification. (f = forward; r = reverse)

Cysteine mutants	Sequence (5' → 3')
Ea G3C-f	CCTGGTTCCGCGTGGATCCTGCAATAGCCATACCACCCCG
Ea G3C-r	CGGGGTGGTATGGCTATTGCAGGATCCACGCGGAACCAGG
Ea Q50C-f	GCACAGAGCCTGGTTUGCAGCATTCAGAGCCTGGC
Ea Q50C-r	GCCAGGCTCTGAATGCTGCAAACCAGGCTCTGTGC
Ea Q50C+L55M-f	GCACAGAGCATGGTTTGCAGCATTCAGAGCATGGC
Ea Q50C+L55M-r	GCCATGCTCTGAATGCTGCAAACCATGCTCTGTGC
Lh A50C-f	GGTAATACCCTGATGTGCGCAATGGATAATATGGG
Lh A50C-r	CCCATATTATCCATTGCGCACATCAGGGTATTACC
Lg A50C-f	GCAATACCCTGATGTGCGCAATGGATAATATGGGTGG
Lg A50C-f	CCACCCATATTATCCATTGCGCACATCAGGGTATTGC
Nc T50C-f	GGTGATACCATTAAATGCGCCATGGATAAAATGGC
Nc T50C-r	GCCATTTTATCCATGGCGCATTTAATGGTATCACC

Table 8. 2: Primers for *E. australis* modifications. (f = forward; r = reverse)

Ea Mutation	Sequence (5' → 3')
Ea M20L-f	GGCAGAAAATTTTCTGAATAGCTTTATGCAGGGCCTGAGC
Ea M20L-r	GCTCAGGCCCTGCATAAAGCTATTCAGAAAATTTTCTGCC
Ea M24L-f	GGCAGAAAATTTTATGAATAGCTTTCTGCAGGGCCTGAGC
Ea M24L-r	GCTCAGGCCCTGCAGAAAGCTATTCATAAAATTTTCTGCC
Ea M41L-f	GCCAGCTGGATGATCTGAGCACCATTGC
Ea M41L-r	GCAATGGTGCTCAGATCATCCAGCTGGC
Ea M48L-f	CCATTGCACAGAGCCTGGTTCAGAGCATTCAGAGCC
Ea M48L-r	GGCTCTGAATGCTCTGAACCAGGCTCTGTGCAATGG
Ea M77L-f	GGCATTTGCAAGCAGCCTGGCAGAAATTGCAGCAAGCG
Ea M77L-r	CGCTTGCTGCAATTTCTGCCAGGCTGCTTGCAAATGCC
Ea M101L-f	GCATTGCAAGCGCACTGAGCAATGCATTTCTGC
Ea M101L-r	GCAGAAATGCATTGCTCAGTGCGCTTGCAATGC
Ea M71L-f	GCAGGCACTGAATCTGGCATTTGCAAGCAGC
Ea M71L-r	GCTGCTTGCAAATGCCAGATTCAGTGCCTGC

Ea M126L-f	CCCAGCTGGTTAGCCTGTTTGCACAGGC
Ea M126L-r	GCCTGTGCAAACAGGCTAACCAGCTGGG
Ea M132L-f	GCACAGGCAGGTCTGAATGATGTTAGCGC
Ea M132L-r	GCGCTAACATCATTCAGACCTGCCTGTGC

Ea M24L+L27M-f	GCTTTCTGCAGGGCATGAGCAGCATGCCTGGTTTTACC
Ea M24L+L27M-r	GGTAAAACCAGGCATGCTGCTCATGCCCTGCAGAAAGC
Ea M40+L38M-f	CCGCAAGCCAGATGGATGATCTGAGCACCATTGC
Ea M40+L38M-r	GCAATGGTGCTCAGATCATCCATCTGGCTTGCGG
Ea L55M-f	CAGAGCATTCAGAGCATGGCAGCACAGGGTCGTACC
Ea L55M-r	GGTACGACCCTGTGCTGCCATGCTCTGAATGCTCTG
Ea L66M-f	GGTCGTACCAGCCCGAATAAAATGCAGGCACTGAATATGGC
Ea L66M-r	GCCATATTCAGTGCCTGCATTTTATTCGGGCTGGTACGACC
Ea L66M+L69M-f	GCCCGAATAAAATGCAGGCAATGAATATGGCATTTGC
Ea L66M+L69M-r	GCAAATGCCATATTCATTGCCTGCATTTTATTCGGGC
Ea L90M-f	GCGAAGAAGGTGGCGGTAGCATGAGCACCAAAACCAGC
Ea L90M-r	GCTGGTTTTGGTGCTCATGCTACCGCCACCTTCTTCGC

Ea M77I-f	GGCATTTGCAAGCAGCATTGCAGAAATTGCAGCAAGCG
Ea M77I-r	CGCTTGCTGCAATTTCTGCAATGCTGCTTGCAAATGCC
Ea M77A-f	GGCATTTGCAAGCAGCGCGGCAGAAATTGCAGCAAGCG
Ea M77A-r	CGCTTGCTGCAATTTCTGCCGCGCTGCTTGCAAATGCC
Ea M77V-f	GGCATTTGCAAGCAGCGTGGCAGAAATTGCAGCAAGCG
Ea M77V-r	CGCTTGCTGCAATTTCTGCCACGCTGCTTGCAAATGCC
Ea M77P-f	GGCATTTGCAAGCAGCCCGGCAGAAATTGCAGCAAGCG
Ea M77P-r	CGCTTGCTGCAATTTCTGCCGGGCTGCTTGCAAATGCC
Ea M77F-f	GGCATTTGCAAGCAGCTTTGCAGAAATTGCAGCAAGCG
Ea M77F-r	CGCTTGCTGCAATTTCTGCAAAGCTGCTTGCAAATGCC

Table 8. 3: Primers for *L. hesperus* and *N. clavipes* modifications made with the QuikChange protocol. (f = forward; r = reverse)

L5 mutants	Sequence (5' → 3')
Lh M38L-f	GCTTTAGCCAGGATCAGCTGGAAGATATGAGCCTGATTGG
Lh M38L-r	CCAATCAGGCTCATATCTTCCAGCTGATCCTGGCTAAAGC
Lh M41L-f	GCCAGGATCAGATGGAAGATCTGAGCCTGATTGGTAATACCC
Lh M41L-r	GGGTATTACCAATCAGGCTCAGATCTTCCATCTGATCCTGGC
Lh M49L-f	GGTAATACCCTGCTGGCAGCAATGGATAATATGGGTGGTCGT
Lh M49L-r	ACGACCACCCATATTATCCATTGCTGCCAGCAGGGTATTACC
Lh M52L-f	GGTAATACCCTGATGGCAGCACTGGATAATATGGGTGGTCGT
Lh M52L-r	ACGACCACCCATATTATCCAGTGCTGCCATCAGGGTATTACC
Lh M55L-f	GGTAATACCCTGATGGCAGCAATGGATAATCTGGGTGGTCGT
Lh M55L-r	ACGACCACCCAGATTATCCATTGCTGCCATCAGGGTATTACC
Nc M24L-f	GCCTTTATTAACGCATTTCTGAATGAAGCAGGTCGTACC
Nc M24L-r	GGTACGACCTGCTTCATTCAGAAATGCGTTAATAAAGGC
Nc M41L-f	GCAGATCAGCTGGATGATCTGAGCACCATTGGTGATACC
Nc M41L-r	GGTATCACCAATGGTGCTCAGATCATCCAGCTGATCTGC
Nc M52L-f	CCATTAAAACCGCCCTGGATAAAATGGCACGTAGC
Nc M52L-r	GCTACGTGCCATTTTATCCAGGGCGGTTTTAATGG
Nc M77L-f	GCATTTGCAAGCAGCCTGGCAGAAATTGCAGC
Nc M77L-r	GCTGCAATTTCTGCCAGGCTGCTTGCAAATGC
Nc M55L-f	CCGCCCTGGATAAACTGGCACGTAGCAATAAAAGC
Nc M55L-r	GCTTTTATTGCTACGTGCCAGTTTATCCAGGGCGG

Table 8. 4: Primers for Hsp90 modifications made with the QuikChange protocol. (f = forward; r = reverse)

Hsp90 mutants	Sequence (5' → 3')
L641M-f	CCATTGACCAAGGCCATGAAAGAAATTTTGGG
L641M-r	CCCAAAATTTCTTTCATGGCCTTGGTCAATGG
L637+641+646M-f	CCAATGACCAAGGCCATGAAAGAAATTATGGGTGACCAAGTGG
L637+641+646M-r	CCACTTGGTCACCCATAATTTCTTTCATGGCCTTGGTCATTGG
L716M-f	CCAATTATCAAGGAAATGAAAAAGAGAGTTGACGAAGG
L716M-r	CCTTCGTCAACTCTCTTTTTCATTTCCTTGATAATTGG
L737M-f	GGACTTGACTAAGTTAATGTATGAAACTGCTTTGTTGACTTCC
L737M-r	GGAAGTCAACAAAGCAGTTTCATACATTAACTTAGTCAAGTCC
I713M+L716M-f	CCAAAATCTCCAATTATGAAGGAAATGAAAAAGAGAGTTGACG
I713M+L716M-r	CGTCAACTCTCTTTTTCATTTCCTTCATAATTGGAGATTTTGG

Annexe C: Vector maps

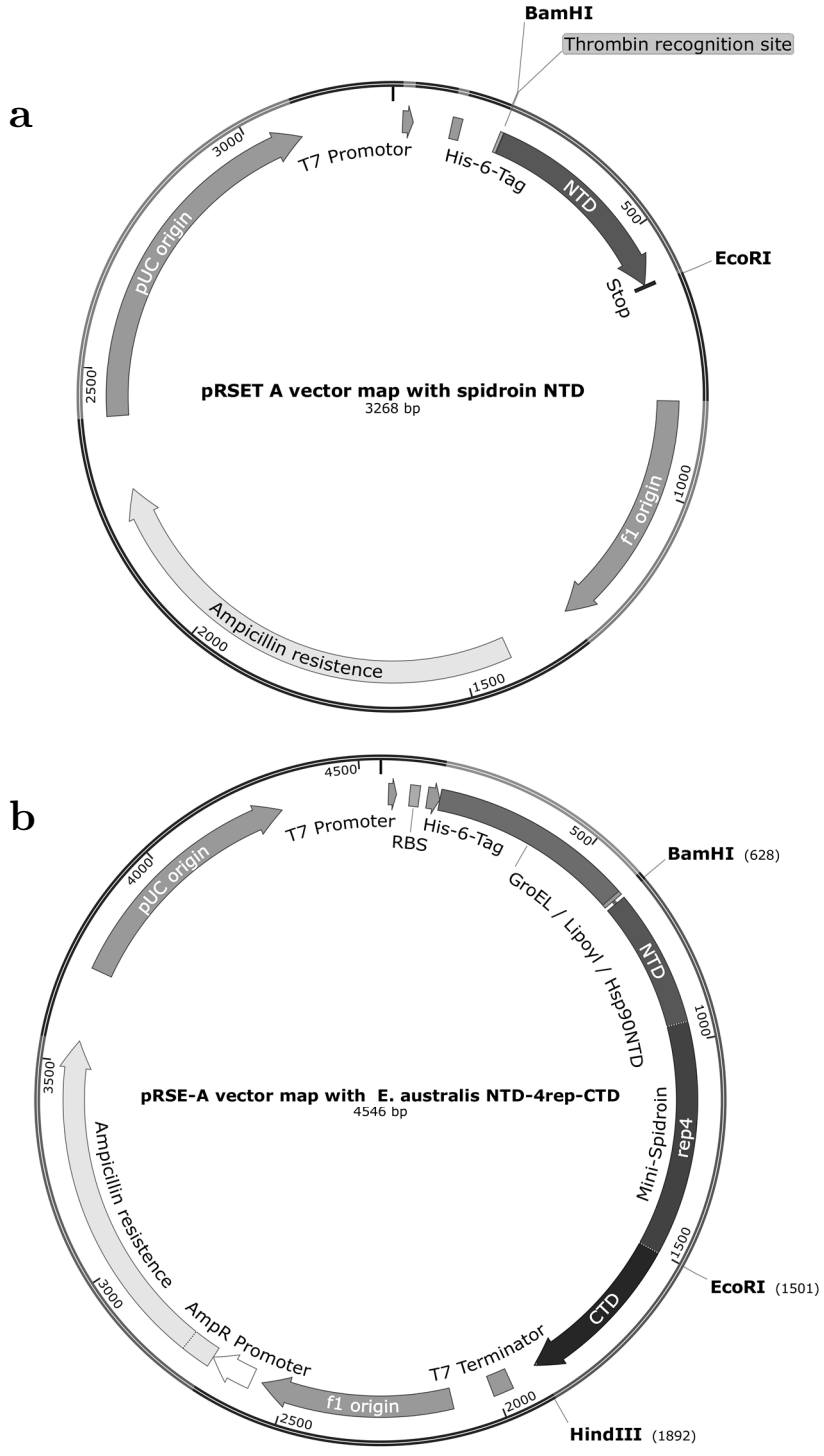


Figure 8. 1: Vector-maps of implemented plasmids. (**a**) pRSET-A vectors with introduced spidroin NTD or (**b**) miniaturised spidroin (exemplary NTD-4rep-CTD) (maps created with SnapGene).

Annexe D: FoldIt protein sequences

Table 8. 5: De novo protein sequences of FoldIt users citizen scientists, with PDB-code and FoldIt-name. Methionine residues highlighted in red, leucine in yellow, isoleucine in blue.

PDB-code	FoldIt-name	Sequence
6MRR	Foldit1	GWSTELEKHR EELKEFLKKE GITNVEIRID NGRLEVRVEG GTERLKRFLE ELRQKLEKKG YTVDIKIE
6MRS	Peak6	GSGRQEKVLK SIEETVRKMG VTMETHRSGN EVKVVIKGLH ESQQEQLKK DVEETSKKQ GVETRIEFH GDTVTIVVR E
6NUK	Ferredog-Diesel	MGHHHHHHGW SENLYFQGST VRIEIRFTNM RREEVQKELE KFKERLKELE KRTGSEIRIE IEERDGEVRV EVEIRNSHEE EVRQIIEEIE RWVRKMGGE LRVEK
6MSP	Foldit3	MGHHHHHHEN LYFQSHMTDE LLERLRQLFE ELHERGTEIV VEVHINGERD EIRVRNISKE ELKKLLERIR EKIEREGSSE VEVNVHSGGQ TWTFNEK

Annexe E: PCR programs

Table 8. 6: Thermocycling parameters for plasmid amplification using the QuikChange protocol for insertions new base pairs for spidroin NTD

Reaction step	Cycle-times	Temperature	Reaction time
1	1	98 °C	30 sec
2 Denaturation		98 °C	30 sec
3 Primer-hybridisation	20	58 °C	30 sec
4 Elongation		72 °C	240 sec
5	1	4 °C	∞

Table 8. 7: Thermocycling parameters for plasmid amplification using the QuikChange protocol for insertions new base pairs for Hsp90

Reaction step	Cycle-times	Temperature	Reaction time
1	1	98 °C	30 sec
2 Denaturation		98 °C	30 sec
3 Primer-hybridisation	20	56 °C	30 sec
4 Elongation		72 °C	240 sec
5 Final Extension	1	72 °C	120 sec
6	1	4 °C	∞

Acknowledgements

I would like to start off by thanking Hannes Neuweiler, my supervisor. You introduced me to the world of protein folding and gave me the opportunity to do my master and PhD thesis with you (and all started with this snake cube puzzle you brought to your lecture, or "Zauberwürfel" as you called it). I am very grateful for the wonderful time in your lab, that you transmitted your passion and enthusiasm for science, your precise thinking and determination. Thank you for letting me work on these fantastic projects.

Markus Sauer, for your support in any form, from funding to the multiple opportunities you gave me to present my work. For always being positive and having positive energy; that you support each and everyone in your group, and never let anyone down. Thank you.

Thomas Müller, who accepted to be my second supervisor and for giving helpful advice in our scientific discussions. For being fair and a good advisor.

To all of my collaboration partners:

Chris Johnson. It was a great pleasure to work with you and thank you for having us in Cambridge back in 2016. Back then, I was blown away from the LMB and still am. What a great place to do science! Thank you for being so canny and a fun person to work with.

Many thanks go to Thomas Bumm, Gernot Stuhler and Dina Kouhestani, for sharing a fantastic project with me and for letting me think outside the box.

Ute Hellmich and Benedikt Goretzki, it was a great pleasure to work with you both. I guess we had what could be called an exemplary collaboration. I hope to be lucky to work with you again in the future.

Many thanks to the department of biotechnology and biophysics, to its fantastic people and the incredible atmosphere they create. For always having cold drinks and snacks, when you are craving for them. Special thanks to Sina, Felix, Alex, Partick, Simon, Andrea, Sebastian, Gerti and Andy for that. Thanks to Suhaila

and Joni for taking over the coffee assignment, and to Fabi to have shown me how to pull strings. Without coffee, the department goes on strike, as you might have already found out ;)

Andrea, Suhaila, Joni and Charlotte, for being an awesome and supportive "Neuweiler-Gang". To Suhaila, Joni and Charlotte, I would like to say: Keep going! You got this!

Of course, particular thanks go to my students: Suhaila, Charlotte, Alex, Sarah! You all taught me something different! Suhaila and Charly, I am glad I worked with you both. Now it's your turn to do/finish your PhD ;)

Many thanks go to the permanent staff, Petra and Lisa, who always answered my typical "student" questions and got us provisioned with lab stuff we needed. Albert and Marcus, always at my side when lab-equipment wasn't working, Oli and Willi for their support in all electrical or computer problems...

- Willi, you will always be remembered.

I'd like to express special thanks to the secretary's office, which might get overlooked sometimes when it comes to acknowledgement speeches. I would like to thank Frau Thal and Sigrid, without whom none of my administrative paperwork and purchase orders would have gone as smoothly. Frau Thal, thank you for always having a fabulous sense of humour! I wish you all the best for your retirement.

To my many friends, I have made in my time at the department. Special thanks go to the "Mensa-team" for never letting me eat alone. To Gerti, for being the best office colleague. Thanks for bringing me snacks when I was doing extra hours in the ensemble lab. Thanks for making me laugh on so many occasions.

Nora. For being a good friend for now many years. For loving coffee and cake as much as I do ;) To Jan, for showing me how to take it easy. To Teresa and Mara, for always being there.

To Caro, Jenny, Julia, Vany, Flo and Andy, for their friendship over the years. You are truly amazing friends. I hope we always will be.

Christian, I don't think I can put in words, what it meant to me that you supported and loved me these last years. Thank you for always being attentive. I'd like to thank you, especially for your incredible support over the past months!! And for your careful proofreading of this work!

To my family. For their support and love, for always having believed in me! I would not be where I am without you! Pierre, Sylvie et Doris, merci de tout cœur!

Publication List

Publications originating from this work:

Heiby, J. C., Rajab, S., Rat, C., Johnson, C. M. & Neuweiler, H. (2017): Conservation of folding and association within a family of spidroin N-terminal domains. In: *Scientific Reports* 7, p. 16789. DOI: 10.1038/s41598-017-16881-6.

Rat, C., **Heiby, J. C.**, Bunz, J. P., Neuweiler, H. (2018): Two-step self-assembly of a spider silk molecular clamp. In: *Nature Communications* 9 (1), p. 4779. DOI: 10.1038/s41467-018-07227-5.

Heiby, J. C., Goretzki, B., Johnson, C. M., Hellmich, U. A., Neuweiler, H. (2019): Methionine in a protein hydrophobic core drives tight interactions required for assembly of spider silk. In: *Nature Communications* 10 (1), p. 4378. DOI: 10.1038/s41467-019-12365-5.

Other publications:

Goretzki, B., **Heiby, J. C.**, Hacker, C., Neuweiler, H. & Hellmich, U. A. (2019): NMR Assignments of a dynamically perturbed and dimerization inhibited N-terminal domain of *E. australis* spider silk protein. In: *Biomolecular NMR Assignments*. In review.

Banaszek, A., Bumm, T., Nowotny, B., Geis, M., Jacob, K., Wölfl, M., Trebing, J., Kouhestani, D., Grimmer, A., Gogishvili, T., Rasche, L., Hönemann, D., Neuweiler, H., **Heiby, J. C.**, Bargou, R., Wajant, H., Einsele, H., Riethmüller G., Stuhler, G. (2019): On-Target Restoration of a Split T Cell-Engaging Antibody for Precision Immunotherapy. In: *Nature Communications*. Accepted.

Kouhestani D., **Heiby, J. C.**, Geis, Alsouri, S., Neuweiler, H., Bumm, T., Einsele, H., Sauer, M., Stuhler, G. (2019): Atypical synapse formation by complementary T-cell engaging hemibodies. Manuscript in preparation.

Scientific Presentations

2017 Joint Meeting of Czech and German Biophysicists 2017, Hünfeld, Germany (poster presentation)

2017 IZKF-Retreat, Kloster Banz, Germany (poster presentation)

2017 28th Faltertage für Protein Faltung, Dynamik und Stabilität, Halle, Germany (poster presentation)

2018 53th Winter seminar, Biophysical Chemistry, Molecular Biology and Cybernetics of cell functions, Klosters, Switzerland (talk and poster presentation)

2018 4th Biophysics by the Sea, Mallorca, Spain (talk)

2018 29th Faltertage für Protein Faltung, Dynamik und Stabilität, Halle, Germany (talk)

2019 54th Winter seminar, Biophysical Chemistry, Molecular Biology and Cybernetics of cell functions, Klosters, Switzerland (talk)

2019 25th International Anniversary Workshop on "Single Molecule Spectroscopy and Super-resolution Microscopy in the Life Sciences", Picoquant, Berlin, Germany (flash talk and poster presentation)