Structural and Biochemical Characterization of the GABA$_A$ Receptor Interacting Protein Muskelin

Strukturelle und Biochemische Charakterisierung des GABA$_A$ Rezeptor interagierenden Proteins Muskelin

Doctoral Thesis

for a doctoral degree at the Graduate School of Life Sciences,

Julius-Maximilians-Universität Würzburg,

Section Biomedicine

submitted by

Carolyn Francesca Delto

born in Stuttgart

Würzburg 2015

Submitted on: ....

Members of the *Promotionskomitee*:

Chairperson:	Prof. Dr. Thomas Müller
Primary Supervisor:	Prof. Dr. Hermann Schindelin
Supervisor (Second):	Prof. Dr. Matthias Kneussel
Supervisor (Third):	Prof. Dr. Thomas Raabe

Date of Public Defence: ....

Date of Receipt of Certificates: ....

Für Ralf

und den unwahrscheinlichsten Ausgang einer Fehlerbetrachtung in der Kalorimetrie

ZUSAMMENFASSUNG

Das Protein Muskelin wurde in einer Studie aus dem Jahr 2011 als zentraler Koordinator des retrograden Transports von GABA$_A$-Rezeptoren in Neuronen beschrieben. Da Muskelin den Transport des Rezeptors sowohl entlang von Aktinfilamenten als auch Mikrotubuli steuert, könnte es der erste bekannte Vertreter einer neuen Klasse von Regulatoren sein, die den Transport einer Fracht über die Grenzen dieser beiden unabhängigen Systeme von Transportwegen und der damit assoziierten Motorproteine hinweg koordinieren. Um Grundlagen für das Verständnis der Wirkungsweise von Muskelin zu schaffen, war das Ziel dieser Arbeit die biochemische und strukturelle Charakterisierung von Muskelin und seiner Interaktion mit dem GABA$_A$-Rezeptor.

Ein Schwerpunkt der Arbeit lag dabei auf der Analyse der Oligomerisierung von Muskelin. Wie gezeigt werden konnte, beruht die Oligomerisierung auf zwei unabhängigen, von verschiedenen Domänen des Proteins vermittelten Bindungen: zum Einen auf einer bereits bekannten Wechselwirkung der N-terminalen Discoidin-Domäne und des C-terminalen Teils (anschaulich als *Head-to-tail*, englisch für Kopf-an-Schwanz, bezeichnet), zum Anderen auf einer in der Literatur bisher außer Acht gelassenen Dimerisierung des LisH-Motivs in Muskelin.

Für die detaillierten Studien beider Bindungen lieferte die Aufklärung der Kristallstruktur eines Teilstücks von Muskelin, das die Discoidin-Domäne und das LisH-Motiv umfasst, eine wichtige Grundlage. Das Teilstück kristallisierte als Dimer, wobei die Dimerisierung ausschließlich über das LisH-Motiv vermittelt wurde. Die biochemischen Analysen bestätigten, dass das LisH-Motiv in Muskelin als Dimerisierungselement wirkt, und zeigten darüber hinaus, dass die C-terminale Domäne des Proteins diese Dimerisierung wesentlich stabilisiert. Zudem offenbarte die Kristallstruktur den molekularen Aufbau der Oberfläche des Kopfes in der *Head-to-tail*-Bindung, der Discoidin-Domäne. Die Kartierung der zur Bindung beitragenden Aminosäuren belegte, dass die Bindungsstelle der *Head-to-tail*-Interaktion mit der generischen Ligandenbindungsstelle der Discoidin-Domäne zusammenfällt.

Als Teil der Analysen wurden zur LisH-Dimerisierung oder zur *Head-to-tail*-Bindung kritisch beitragende Aminosäurenseitenketten identifiziert, durch deren Mutationen

I

spezifisch jeweils eine der beiden Interaktionen unterbunden werden konnte. Diese Mutationen ermöglichten es, das Zusammenspiel der Bindungen in der Oligomerisierung zu untersuchen. Es konnte gezeigt werden, dass rekombinantes Muskelin ein Tetramer bildet, wozu beide Interaktionen, die LisH-Dimerisierung und die *Head-to-tail*-Bindung, unabhängig beitragen. Wurde jeweils eine der beiden Interaktionen durch Mutation gestört, konnte nur noch ein über die jeweils andere Interaktion vermitteltes Dimer gebildet werden, bei gleichzeitiger Störung beider Interaktionen lag das Protein als Monomer vor. Darüber hinaus konnte Frank Heisler in der Arbeitsgruppe von Matthias Kneussel mit Hilfe dieser Mutationen zeigen, dass eine Störung der LisH-Dimerisierung drastische Auswirkungen auf Muskelin in Zellen hat. Die Störung der LisH-Dimerisierung führte zu einer vollständigen Umverteilung des sonst im Zytoplasma lokalisierten Muskelins in den Kern, begleitet von einer starken Beeinträchtigung seiner Funktion im Transport des $GABA_A$-Rezeptors. Auf diesen Ergebnissen aufbauend wurde durch Analysen der oligomeren Zustände von Muskelin-Varianten, für die in der Literatur eine veränderte Lokalisation beschrieben worden war, die entscheidende Bedeutung der LisH-Dimerisierung für die subzelluläre Verteilung und damit die Rolle von Muskelin in der Zelle bekräftigt.

Die biochemischen Studien der Interaktion von Muskelin und der α1-Untereinheit des $GABA_A$-Rezeptors demonstrierten eine direkte Bindung mit mikromolarer Affinität, die in Muskelin vorwiegend durch die Kelch-repeat-Domäne vermittelt wird. Für die Bindungsstelle auf Seite des $GABA_A$-Rezeptors wurde bestätigt, dass die dreizehn C-terminalen Reste der intrazellulären Domäne entscheidend sind. In Übereinstimmung mit der starken Konservierung dieser Reste in verschiedenen α-Untereinheiten des $GABA_A$-Rezeptors, konnte gezeigt werden, dass *in vitro* eine Bindung von Muskelin auch an die intrazelluläre Domäne der α2- und α5-Untereinheit möglich ist. Anhand des Vergleichs der Bindungsstelle zwischen den homologen Untereinheiten lassen sich erste Rückschlüsse auf die Details der Interaktion ziehen, die als Anknüpfungspunkt für kommende Studien dienen können.

Diese Arbeit liefert damit wesentliche Beiträge zum Verständnis von Muskelin, insbesondere der Bedeutung seiner Oligomerisierung. Sie bietet zudem ein experimentelles Rahmenwerk für zukünftige Studien, die sich verwandten Themen, wie der Charakterisierung weiterer Interaktionen von Muskelin, oder den in dieser Arbeit aufgeworfenen Fragen widmen.

Summary

In a study from 2011, the protein muskelin was described as a central coordinator of the retrograde transport of $GABA_A$ receptors in neurons. As muskelin governs the transport along actin filaments as well as microtubules, it might be the first representative of a novel class of regulators, which coordinate cargo transport across the borders of these two independent systems of transport paths and their associated motorproteins. To establish a basis for understanding the mode of operation of muskelin, the aim of this thesis was an in-depth biochemical and structural characterization of muskelin and its interaction with the $GABA_A$ receptor.

One focus of the work was the analysis of the oligomerization of muskelin. As could be demonstrated, the oligomerization is based on two independent interactions mediated by different domains of the protein: a known interaction of the N-terminal discoidin domain with the C-terminal portion, termed head-to-tail interaction, and a dimerization of the LisH motif in muskelin that was so far neglected in the literature. For the detailed studies of both binding events, the solution of a crystal structure of a fragment of muskelin, comprising the Discoidin domain and the LisH motif, was an important basis. The fragment crystallized as a dimer, with dimerization being mediated solely by the LisH motif. Biochemical analysis corroborated that the LisH motif in muskelin serves as a dimerization element, and, moreover, showed that the C-terminal domain of the protein substantially stabilizes this dimerization. In addition, the crystal structure revealed the molecular composition of the surface of the head in the head-to-tail interaction, namely the discoidin domain. This information enabled to map the amino acids contributing to binding, which showed that the binding site of the head-to-tail interaction coincides with the generic ligand binding site of the discoidin domain.

As part of the analyses, residues that are critical for LisH-dimerization and the head-to-tail binding, respectively, were identified, whose mutation specifically interfered with each of the interactions separately. These mutations allowed to investigate the interplay of these interactions during oligomerization. It could be shown that recombinant muskelin assembles into a tetramer to which both interactions, the LisH-

dimerization and the head-to-tail binding, contribute independently. When one of the two interactions was disturbed, only a dimer mediated via the respective other interaction could be formed; when both interactions were disturbed, the protein was present as monomer. Furthermore, Frank Heisler in the group of Matthias Kneussel was able to show the drastic impact of an impaired LisH-dimerization on muskelin in cells using these mutations. Disturbing the LisH-dimerization led to a complete redistribution of the originally cytoplasmic muskelin to the nucleus which was accompanied by a severe impairment of its function during GABA$_A$ receptor transport. Following up on these results in an analysis of muskelin variants, for which alterations of the subcellular localization had been published earlier, the crucial influence of LisH-dimerization to the subcellular localization and thereby the role of muskelin in the cell was confirmed.

The biochemical studies of the interaction of muskelin and the α1 subunit of the GABA$_A$ receptor demonstrated a direct binding with an affinity in the low micromolar range, which is mediated primarily by the kelch repeat domain in muskelin. For the binding site on the GABA$_A$ receptor, it was confirmed that the thirteen most C-terminal residues of the intracellular domain are critical for the binding of muskelin. In accordance with the strong conservation of these residues among the α subunits of the GABA$_A$ receptor, it could be shown that an interaction with muskelin *in vitro* is also possible for the α2 and α5 subunits. Based on the comparison of the binding sites between the homologous subunits, tentative conclusions can be drawn about the details of the binding, which may serve as a starting point for follow-up studies.

This thesis thereby makes valuable contributions to the understanding of muskelin, in particular the significance of its oligomerization. It furthermore provides an experimental framework for future studies that address related topics, such as the characterization of other muskelin interaction partners, or the questions raised in this work.

Contents

VI

INTRODUCTION

Reading this text is an instructive example of the capabilities and operating principles of a highly developed nervous system. Receptors in the retina detect the light reflected from or emitted by the medium carrying the text, and translate the stimuli to electrical activity, the coding of information in the nervous system. There, the information is processed in several steps while it is conveyed through different centers in the brain - from the retina to the midbrain, the thalamus, the primary visual cortex, the inferior temporal cortex and further on. Muscle activity in the eye is coordinated to focus the letters and follow the text lines while at the same time the optical stimuli are assembled to patterns, these patterns are interpreted as text by drawing on learned abilities and memory, and, finally, the information content of the text is extracted, is converted to memory itself and can elicit emotions such as excitement, accompanied by physiological reactions like an acceleration of the heartbeat.

To put it more generally, the purpose of a nervous system is to coordinate the activities of functional units in a complex organism and to adapt them to inner and outer conditions. It is accordingly designed to receive information about internal and external stimuli, to integrate all information it receives and to derive and send appropriate instructions to the body. Two specialized cell types have evolved to fulfill these functions: the neurons as information receiving, integrating and propagating units and the glia cells, which provide a specific surrounding that neurons depend upon. The neurons are interconnected to a high degree, forming a huge, complex and highly organized network[1]. In this network, information is represented as electrical activity: electrical signals are generated, propagated and processed based on the transmissive properties of the node points and the connections between them.

---

[1] Most recent estimates for a human brain are $(86 \pm 8) \times 10^9$ neurons (Azevedo et al., 2009) that are connected by 150,000 to 180,000 km of myelinated nerve fibers and form 150×10^{12} synapses in the cerebral cortex alone (Drachman, 2005).

1.1 Neuronal communication

The nature of the electrical signals used in the neuronal network is that of changes in the electrical potential across the plasma membrane of neurons (figure 1.1). Thus a vital property of neurons is a well-defined membrane potential[2] that can be quickly and significantly altered during signaling. Two major classes of electrical signals can be distinguished: the typical form in which input is received is that of variable *localized graded potentials*, the typical form in which output is encoded is that of stereotypic *action potentials*. Localized graded potentials vary in their amplitude and decrease rapidly over time and space, while action potentials are binary events with a fixed amplitude that can be propagated over long distances without decaying.

Four regions can be distinguished in the architecture of neurons based on the function they carry: to receive input, to integrate input, to conduct signals, or to generate the output. Typical morphological features relate to these functional regions. Branched, tapering projections termed *dendrites* provide an enlarged area to receive and integrate input. They are connected to the cell body, which contains the nucleus and major organelles, and also receives and integrates input. If the summed input potentials increase the membrane potential to above a certain threshold, an action potential is triggered at the beginning of a specialized projection of the neuron termed *axon*. The axon conducts the action potentials and at its ends forms the connections to the downstream neurons termed *synapses*, where the output signal is transmitted.

1.1.1 Synaptic transmission

The site of signal transmission between two neurons, the synapse, can either be a conductive link of the cells (electrical synapse) or a specialized junction where the electrical signal is converted to the release of neurotransmitters (chemical synapse). Electrical synapses enable bidirectional flows of charge from cell to cell, thereby mediating symmetrical, fast and reliable transmission. Chemical synapses offer more variability in the evoked signal - depending on the transmitter and receptor involved, it can be excitatory or inhibitory, fast or sustained, amplified or reduced - and they are the prevailing type of synapse in the central nervous system (CNS). Chemical synapses

---

[2] The membrane potential arises from actively sustained asymmetric distributions of ions (mainly Na^+, K^+ and Cl^-) and selective permeabilities. In the resting state, inverse concentration gradients of Na^+ and K^+ (Na^+ high outside, K^+ high inside the cell) are maintained by the sodium-potassium pump. Selective pores make the membrane much more permeable for K^+ than for Na^+. K^+ moving out of the cell along its concentration gradient creates a negative potential at the plasma membrane. The equilibrium of fluxes (active transport and passive diffusion) is reached at the resting potential, typically -70 to -80 mV, close to the equilibrium potential of K^+. For Cl^-, the contribution of active transport and selective permeability are diverse between neurons, but commonly, the equilibrium potential of Cl^- is close to that of K^+, and Cl^- currents do not drastically alter, but rather stabilize the resting potential.
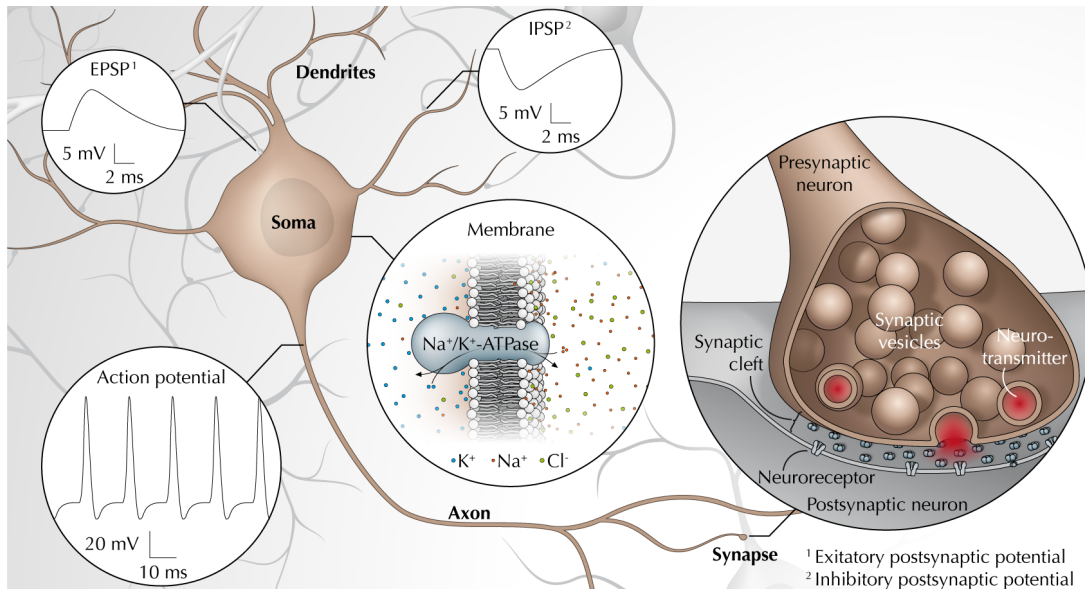
Figure 1.1 – Cellular basis of neuronal communication. Simplified schematic illustration of a neuron with morphological features indicated and different types of electrical signals assigned.

are asymmetrical and pass information directionally. The output sending presynaptic cell forms the presynaptic terminal with an active zone, where synaptic vesicles filled with neurotransmitters are docked to the plasma membrane. In response to an incoming action potential the vesicles fuse with the presynaptic membrane and release the neurotransmitters into the synaptic cleft. The opposing membrane of the input receiving postsynaptic cell contains a high density of complementary receptors that bind the neurotransmitter and evoke the response. These receptors are embedded in a complex of interlinked molecules, the postsynaptic density.

A chemical synapse is classified as excitatory when its activation results in a depolarization of the postsynaptic membrane, thereby bringing the membrane potential closer to the threshold at which the postsynaptic neuron itself generates an action potential. It is classified as inhibitory when its activity reduces the probability of the postsynaptic neuron to pass this threshold, either by hyperpolarizing the postsynaptic membrane or by short-circuiting depolarizing currents. This distinction is typically related to certain properties of synapses, as the neurotransmitter that is predominately released, the localization (axodentritic, axosomatic, axo-axonic) or the morphology. The great majority of excitatory synapses throughout the CNS use glutamate as neurotransmitter; for inhibitory synapses the most common neurotransmitter differs between the brain, where γ-aminobutyric acid (GABA) is by far most common, and the spinal cord, where glycine is the major inhibitory transmitter.

3
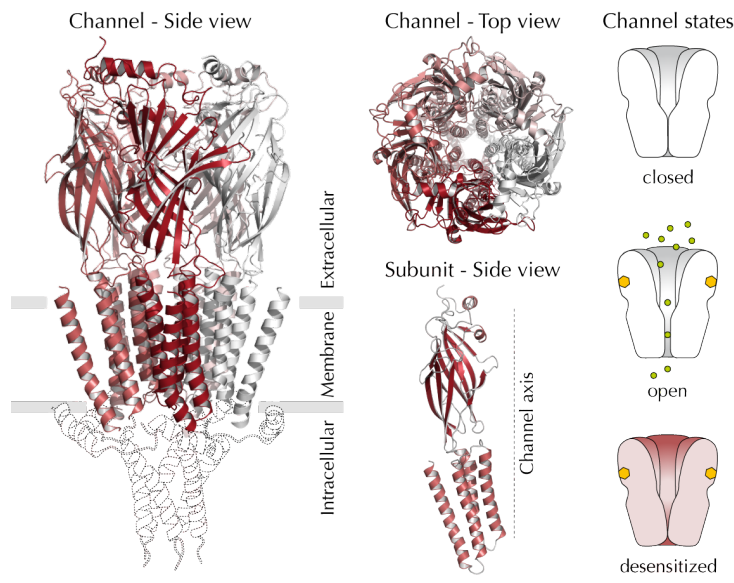
Figure 1.2 – **Structure of a GABA_A receptor,** the GABA_AR β3 homopentamer in side (*left*) and top view (*center, top*) with the subunits colored in red to gray shading. To account for the intracelluar part missing in the crystallized construct, the largest partial structure of a homologous domain available so far, that of the mouse serotonin 5-HT$_3$ receptor, is indicated in dashed lines in the side view. *Center, bottom:* Side view of a single subunit with the channel axis indicated. *Right:* The three functional states of the channel, with the proposed state of the crystallized receptor colored in red. (PDB codes: 4cof, 4pir)

1.1.2 Neurotransmitter receptors

The receptors for neurotransmitters fall in two major classes: ionotropic receptors that are ligand-gated ion channels and metabotropic receptors that trigger an intracellular signaling cascade. The former mediate fast synaptic transmission, whereas the latter produce slower, but longer-lasting reactions that modulate neuronal properties. Most neuroreceptors belong to one of four super-families: the family of pentameric ligand gated ion channels and the family of ionotropic glutamate receptors form the majority of ionotropic receptors, the family of G protein-coupled receptors and the family of receptor tyrosine kinases form the majority of metabotropic receptors.

The GABA_A receptor

GABA, the chief inhibitory neurotransmitter in the mammalian brain, acts on both ionotropic GABA_A and metabotropic GABA_B receptors, which are crucial for signaling in many central processes. Their importance is underscored, in particular for GABA_A receptors, by the variety of disorders caused by disturbed receptor function, such as epilepsy, schizophrenia, anxiety and sleep disorders (Moehler, 2006), and the large number of clinically important drugs that target these receptors, including benzodiazepines, barbiturates, several general anesthetics and ethanol (Sieghart, 1995). The GABA_A receptors belong to the super-family of pentameric ligand gated ion channels and share the typical architecture (figure 1.2), illustrated by several structures deter-

mined for other family members[3] and recently the human GABA$_A$R β3 channel (Miller and Aricescu, 2014). A single receptor has a cylindrical shape[4] and consists of five subunits which are symmetrically arranged around a central pore. The N-terminal half of each subunit folds into a curled β-sandwich and constitutes the extracellular domain. It is followed by four transmembrane α-helices of which the second helix tapers the inside of the pore. A large insertion between the third and the fourth helix forms the intracellular domain of the receptor. It contains few structured elements, hence the major portion of it is assumed to be flexible, and structural information about this part of the GABA$_A$ receptor is missing[5].

The receptors are able to adopt several conformational states: the closed resting state, the agonist-bound open state and the agonist-bound closed state (termed desensitized) as well as variations of and transitional states between these major states. The extracellular domain harbors the binding sites for the neurotransmitter and for many modulators of the channel. Upon binding of the neurotransmitter, conformational transitions in the extracellular domain induce rearrangements of the transmembrane helices and the intracellular domain, resulting in the opening of the channel. Structural details in the architecture of the transmembrane domain determine the selectivity of the ion channel[6], which is permeable mainly to chloride in GABA$_A$ receptors. The transmembrane domain is also the binding site for several modulators of channel function, such as hydrophobic anesthetics, lipids, or channel blockers. The intracellular domain contains several phosphorylation sites and is the interface for interaction with all intracellular proteins that mediate modulation, trafficking and anchoring of the receptors (Chen and Olsen, 2007).

GABA$_A$ receptors with few exceptions are heteropentamers assembled from a large set of homologous subunits. In humans, 19 different genes encode for GABA$_A$ receptor subunits[7]. The high number of possible combinations based on known assembly

---

[3] That is: the electron microscopy structure of the nicotinic acetylcholine receptor of the electric ray *Torpedo marmorata* (Unwin, 2005) and crystal structures of the prokaryotic homologs in *Erwinia chrysanthemi* (Hilf and Dutzler, 2008) and *Gloeobacter violaceus* (Bocquet et al., 2009), a glutamate receptor of *Caenorhabditis elegans* (Hibbs and Gouaux, 2011) and a mouse serotonin 5-HT$_3$ receptor (Hassaine et al., 2014).

[4] Without the intracellular part, the cylinder is 110 Å in height with an outer diameter of 60 to 80 Å.

[5] Based on the structures of the nicotinic acetylcholine receptor from an electric ray (Unwin, 2005) and the mouse serotonin 5-HT$_3$ receptor (Hassaine et al., 2014), the intracellular domain is predicted to include at least one α-helix as a continuous extension to the fourth transmembrane helix. Together, these intracellular helices form a conical vestibule.

[6] Only three mutations in the second transmembrane helix can invert the ion selectivity of the acetylcholine receptor (Galzi et al., 1992) and the glycine receptor (Keramidas et al., 2000).

[7] That is: α(1-6), β(1-3), γ(1-3), δ, ε, θ, πand ρ(1-3), with additional variation due to alternative splicing of α2, α4, α5, β2, β3, γ2, γ3, ε and ρ(Simon et al., 2004) and RNA editing of α3 (Daniel and Öhman, 2009).

rules is opposed by a very restricted set of combinations that occur *in situ*[8] (Olsen and Sieghart, 2008). The GABA$_A$ receptor subtypes differ in their biophysical and pharmacological properties as well as their temporal and spatial expression patterns, both at the cellular and subcellular level (Sieghart and Sperk, 2002). The composition of the major subtype in the adult brain is (in clockwise order viewed from the extracellular side) γ2α1β2α1β2 (Sigel and Steinmann, 2012).

The subcellular localization of a receptor subtype is critical for its role in signaling, with synaptic receptors mediating phasic (fast, transient) inhibition and extrasynaptic receptors mediating tonic (long-lasting) inhibition. Synaptically localized subtypes (typically composed of two α1, α2 or α3 together with two β and a single γ subunit) experience transiently high GABA concentrations during synaptic signaling[9], resulting in their short activation and thereby an increase in anion conductance for milliseconds. Extrasynaptically localized subtypes (most prominently α4βδ, α5βγ2 and α6βδ) are exposed to steady ambient GABA levels and stochastic activation of these receptors leads to persistent, noisy currents. Notably, while an extrasynaptic localization can be exclusive, this does not seem to apply for a synaptic localization. Although typical synaptic receptors are highly enriched at postsynaptic sites, in absolute numbers the majority of these receptors is in fact localized extrasynaptically. The preferential subcellular localization is often reflected in the sensitivity to GABA, which is commonly significantly lower for synaptic than for extrasynaptic receptors[10] and thus seems to be tailored to the GABA concentration the receptor experiences.

Modulation of GABA$_A$ receptor signaling

The diversity of GABA$_A$ receptor subtypes, as reflected in their subcellular distribution and subtleties of response to GABA, gives a first impression of the complexity of signaling via GABA$_A$ receptors. More generally, signaling is shaped by several factors at different levels. The properties of single receptors are directly modulated by small endogenous ligands such as neurosteroids (Herd et al., 2007), by phosphorylation (Brandon et al., 2002) and by interacting proteins, the most prominent example being the GABA$_A$ receptor associated protein GABARAP (Chen and Olsen, 2007). At synapses, the amount of neurotransmitter released, its residence time together with

---

[8] 11 conclusively identified, 6 strong candidates not conclusively identified, 9 tentative candidates

[9] Upon neurotransmitter release, the GABA concentration in the synaptic cleft peaks in the millimolar range, followed by a rapid decay with time constants in the range of $100\,\mu s$ to $500\,\mu s$ due to diffusion away from the release site and clearance by GABA transporters in nerve terminals and glia

[10] Sensitivity to GABA expressed as EC$_{50}$, that is the GABA concentration producing half of the maximal response, of abundant synaptic receptors: α1β2γ2 11-20 μM (Verdoorn, 1994; Saxena and Macdonald, 1996), α2β1/3γ2 5-17 μM (Jones et al., 1995; Böhme et al., 2004) - and common extrasynaptic receptors: α4β3δ 0.5 μM (Brown et al., 2002) or α6β2/3δ 0.2-0.3 μM (Saxena and Macdonald, 1996; Feng and Macdonald, 2004)

the number of receptors and their binding kinetics determine receptor occupancy and thereby the strength of the response. The receptor occupancy at synapses varies substantially between different types of neurons, and can even differ between synapses on the same cell (Hájos et al., 2000).

All influencing factors are tightly controlled, which applies especially to the number of receptors. The amount of a certain receptor subtype delivered to the surface depends on the level of gene expression (Grabenstatter et al., 2012), the controlled assembly of the specific subunit combination, the rate of endoplasmic reticulum (ER) associated degradation of the receptors, the rate of translocation from the ER to the Golgi apparatus and from the Golgi apparatus to the plasma membrane (reviewed in Jacob et al., 2008; Luscher et al., 2011). Once inserted in the neuronal plasma membrane, the number of receptors is locally controlled by anchoring and clustering of the receptors. This is well established for postsynaptic sites where the protein gephyrin is the key factor for clustering (Tretter et al., 2012), but also happens outside synapses as in the case of the $\alpha5\beta\gamma2$ subtype that is clustered extrasynaptically by radixin (Loebrich et al., 2006). GABA$_A$ receptors undergo constitutive and regulated endocytosis, after which they recycle back to the cell surface or are transferred to degradation. Delivery, endocytosis, recycling and transfer to degradation all rely on directed transport along cytoskeletal filaments, and transport processes are specifically regulated to adjust the overall number of a receptor subtype on the cell surface (reviewed in Jacob et al., 2008; Luscher et al., 2011).

The role of muskelin in GABA$_A$ receptor transport

Muskelin was identified as direct interaction partner of the $\alpha1$ subunit, present in the most abundant GABA$_A$ receptor subtype, and was found to influence the internalization and degradation of GABA$_A$R $\alpha1$ (Heisler et al., 2011). More specifically, muskelin and GABA$_A$R $\alpha1$ are cotransported in the retrograde direction, and associate with both the myosinVI and the dynein complex, the motor protein complexes that move cargo along actin filaments and microtubules, respectively. In muskelin knockout mice, both transport steps were disturbed, resulting in overall enhanced GABA$_A$R $\alpha1$ surface levels, the functional consequences of which were illustrated by alterations in network oscillations in the hippocampus. The most intriguing aspect of the putative role of muskelin in GABA$_A$R $\alpha1$ trafficking is the regulation of transport along both actin filaments and microtubules. It suggests muskelin as the first transport factor known to be able to accompany its cargo in subsequent steps in different cytoskeletal transport systems. It is this special role during GABA$_A$ receptor transport that motivated a detailed characterization of muskelin and its interaction with GABA$_A$R $\alpha1$, the central topic of this thesis.

7

1.2 Muskelin

The role as novel coordinator of intracellular transport raised interest in the protein muskelin, which prior to that was the subject of only a few studies. Based on information available on muskelin, it can be described as an evolutionarily well conserved (in eukaryotes), almost ubiquitously expressed multi-domain protein whose function is only poorly understood. These aspects will be described in more detail in the following sections.

1.2.1 Functional context of muskelin

Initially, muskelin was discovered via its role in the cellular response to the extracellular matrix component thrombospondin-1 (Adams et al., 1998). Subsequently, several further binding partners were identified, casting spotlights on the assumed functions of muskelin. The phenotype of the muskelin knockout in mice provided further insights: the knockout mice undergo a change of coat color, which was interpreted as a sign of muskelin's involvement in the intracellular trafficking of pigment granules (Heisler et al., 2011). Overall, three major fields of activity of muskelin emerged: the regulation of cytoskeletal organization, the regulation of intracellular transport and being a member of a complex with putative ubiquitin-ligase activity (figure 1.3). It is an open question whether these aspects are different facets of the same underlying activity. In the following, the reported functions and binding partners of muskelin will be described briefly to illustrate their involvement in these superordinated activities. It was not possible to integrate one binding partner, the T-box transcription factor 20, in this scheme of muskelin functions, as only the interaction, but not its functional relevance has been described (Debenedittis et al., 2011).
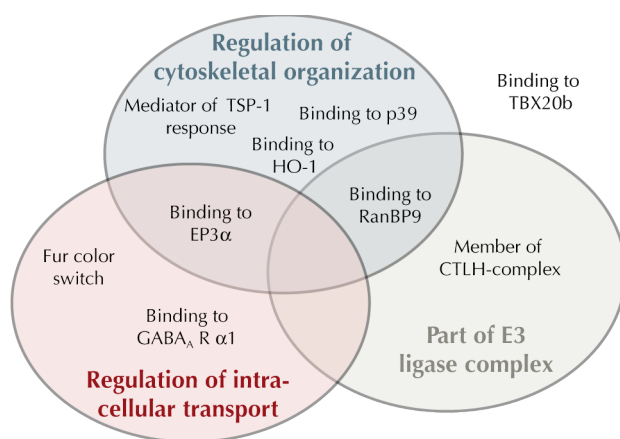
Figure 1.3 – Functional context of muskelin with superordinated activities depicted as set diagrams and reported binding partners and functions integrated as described in the text.

Mediator of Thrombospondin-1 response

Thrombospondin-1 is a matricellular protein[11] that modulates cell behavior in several physiological and pathophysiological processes such as blood hemostasis (Bonnefoy et al., 2008), wound healing (Bornstein et al., 2004), synaptogenesis (Risher and Eroglu, 2012), angiogenesis (Armstrong and Bornstein, 2003) and neoplasia (Lawler and Detmar, 2004). Thrombospondin-1 is a large glycoprotein comprising several domains (Lawler and Hynes, 1986) via which it is able to interact with a variety of binding partners, including cytokines, growth factors and multiple cell surface receptors. Depending on the composition of the environment and the present receptors, thrombospondin-1 assembles distinct multi-protein signaling complexes and evokes a specific cellular response (Bornstein, 1995; Chen et al., 2000).

In a screen for factors promoting cell attachment to the C-terminal domain of thrombospondin-1, muskelin was discovered and shown to act as an important regulator in cell responses to thrombospondin-1, that is cell attachment, cell spreading and cytoskeletal reorganization (Adams et al., 1998). Evolutionary, muskelin is markedly older (see section 1.2.3) than the family of thrombospondins, especially the subgroup of trimeric thrombospondins that occur only in higher chordates (Adams et al., 2003; Prag et al., 2004). The role as mediator of the cellular response to thrombospondin-1 was the first hint at a regulatory role of muskelin in cytoskeletal reorganization processes.

Interaction partner of prostaglandin receptor EP3α

In a yeast two-hybrid screen, muskelin was identified as a binding partner of the C-terminal tail of the EP3 subtype of prostaglandin E_2 receptors (Hasegawa et al., 2000). These receptors belong to the class of G-protein coupled receptors and are expressed in three different splice variants, EP3α, EP3β and EP3γ, of which muskelin was found to specifically bind to the EP3α variant (Hasegawa et al., 2000). The presence of muskelin counteracted the agonist-induced internalization of the receptor and increased down-stream signaling, that is the inhibition of adenylate cyclase by the G-protein G_i, in human embryonic kidney (HEK) 293 cells. It has been hypothesized that this role of muskelin is related to its function in the cellular response to thrombospondin-1 via integrin-associated protein, a thrombospondin-1 receptor that also interacts with G_i (Hasegawa et al., 2000). On the other side, the effect on the internalization of the receptor resembles the role of muskelin in $GABA_A$ receptor transport, although the effect is inverse for the two receptors.

---

[11] The term 'matricellular' has been coined for a small group of functionally related, non-structural components of the extracellular matrix that modulate cell behavior by binding to structural matrix proteins, bioactive molecules and cell surface receptors (Bornstein, 1995).

Interaction partner of p39, a specific activator of cyclin-dependent-like kinase 5

Cyclin-dependent-like kinase 5 (Cdk5) is an atypic cyclin-dependent kinase, atypic as it depends in its activity not on a cyclin family member, but its specific activators p35 and p39, and is not involved in the regulation of cell cycle (Lew et al., 1992). The major place of Cdk5 activity is the CNS, and its best known role is the regulation of the cytoarchitecture (Dhavan and Tsai, 2001). In a yeast two-hybrid screen, muskelin was found to bind to the p39 activator of Cdk5 and p39 was shown to redistribute muskelin to the peripheral actin cytoskeleton (Ledee et al., 2005). The integration of muskelin in Cdk5 signaling was further consolidated in a recent follow-up study which demonstrated for one thing that p39 recruits muskelin to a complex with myosin II on actin stress fibers, for another thing that interfering with muskelin expression had the same effect on stress fiber formation and cell migration as inhibiting Cdk5, and, finally, that muskelin itself is a substrate for Cdk5 (Tripathi et al., 2014). Together, these studies enforce a role for muskelin in the signaling network regulating the cytoarchitecture.

Complex partner of heme-oxygenase-1

In a recent study on the influence of heme-oxygenase-1 (HO-1) on the cell morphology in prostate cancer cells, muskelin was identified being associated with HO-1 (Gueron et al., 2014). Induction of HO-1 activity led to an enhanced muskelin expression and a partial relocation of HO-1, muskelin and its binding partner Ran-binding protein 9 (RanBP9) (described in the next section) to the nucleus. Although the direct link of this up-regulation of muskelin to the observed effects on cell morphology was not established in this study, muskelin was considered a strong candidate. This would add a further node in the protein network around muskelin in the regulation of the cytoarchitecture.

Interaction partner of RanBP9 and member of the CTLH complex

A further aspect of muskelin function is its role in the CTLH complex, a putative E3-Ligase complex. This complex was first described to be assembled around RanBP9 (Nishitani et al., 2001) and subsequently additional subunits were identified: first muskelin and the two-hybrid associated protein to RanBP9 (Twa1) (Umeda et al., 2003), then the macrophage erythroblast attacher (MAEA, also known as p48 EMLP), the protein RMD5 homolog A (Rmnd5a, also known as p44 CTLH) and the Armadillo repeat-containing protein 8 (ARMC8) (Kobayashi et al., 2007). Upon identification of the latter three, on one hand the striking similarities in their domain architecture (figure 1.4 on the facing page), which resulted in the name CTLH complex, and, on the other hand, the homology to the glucose induced degradation (Gid) complex in
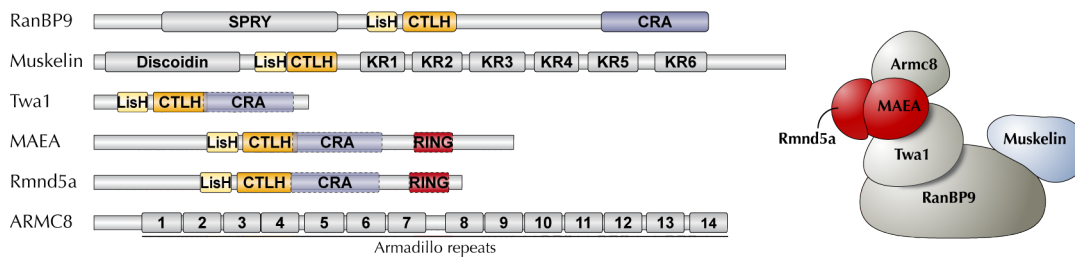
Figure 1.4 – Domain architecture of the CTLH complex subunits and topology of the CTLH complex. *Left:* Schematic representation of the domain architecture of the individual subunits, with recurring domains in color and unique domains in grey (SPRY: SPla kinase and RYanodine receptor domain, LisH: LisH motif, CTLH: CTLH motif, CRA: CT11-RanBPM domain, KR: Kelch repeat, RING: Really Interesting New Gene domain). Domains that were identified with lower confidence are rendered semi-transparent with dashed borders. *Right:* Putative topology of the CTLH complex based on findings for the homologous Gid complex (Menssen et al., 2012), with subunits that supposedly confer ubiquitin ligase activity in red and the accessory subunit muskelin in blue.

yeast was realized (Kobayashi et al., 2007). Whereas the function and topology of the CTLH complex are so far unknown, inferences can be drawn from studies on the Gid complex in *Saccharomyces cerevisiae* (*S. cerevisiae*). The Gid complex is needed for the ubiquitin-dependent degradation of one of the key enzymes in gluconeogenesis, fructose-1,6-bisphosphatase, which represents a central step in the glucose-dependent regulation of carbon metabolism (Hämmerle et al., 1998; Schüle et al., 2000; Regelmann et al., 2003). This activity depends on the ubiquitin ligase activity of the complex provided by two subunits, Gid2 (homologue of Rmnd5a) and Gid9 (homologue of MAEA), that both carry a degenerate RING finger domain (Santt et al., 2008; Braun et al., 2011). As the ubiquitin ligase activity of Rmnd5a is retained in the human homolog (van Wijk et al., 2009), the overall function of an E3 ligase likely applies to the CTLH complex, as well, yet this remains to be demonstrated. Results from studies on the topology of the Gid complex (Menssen et al., 2012) are in good agreement with conclusions drawn from phylogenetic analyses of the complex partners (Francis et al., 2013). Together, they suggest that the core of the complex is formed by RanBP9, Twa1, MAEA and Rmnd5a, while muskelin and ARMC8 are peripheral subunits, which supposedly integrate the complex into the specific cellular signaling network.

So far, no study addressed the connection between the CTLH complex E3 ligase activity and the specific functions of muskelin. A first indication of an interplay between muskelin's role in the CTLH complex and in regulation of the cytoskeleton can, however, be found in a study by Valiyaveettil et al. (2008) which suggested that the binding to RanBP9 is needed for the morphological activity of muskelin. This conclusion was drawn based on the observation that mutations or deletions in muskelin concurrently

11

changed its abilities to bind to RanBP9 and to alter cell morphology, and that effects of depletion of either RanBP9 or muskelin on cell morphology were similar (Valiyaveettil et al., 2008). A further indication of an interplay in muskelin's activities is that ubiquitylation is known to be crucial for the control of internalization, intracellular sorting and degradation of neurotransmitter receptors (Schwarz and Patrick, 2012) including the $GABA_A$ receptors (Saliba et al., 2007; Arancibia-Cárcamo et al., 2009). The $GABA_A R$ $\alpha 1$ subunit has also been found to be ubiquitylated (Na et al., 2012), and its ubiquitin-dependent degradation has recently been linked to the pathophysiology of autism spectrum disorders (Crider et al., 2014). Hence it is also conceivable that muskelin's function as part of a ubiquitin ligase complex is linked to its role in the regulation of the intracellular transport of the $GABA_A$ receptors.

1.2.2 Expression profiles of muskelin

Matching the variety of interaction partners, muskelin was reported to be expressed in many tissues of embryonic and adult mice, the model organism for which expression profiles have been analyzed in most detail. The highest RNA levels in embryonic tissues were found in all parts of the CNS, in muscle tissues, in epithelia of nose and tongue, in tooth buds, the liver and the developing kidneys (Tagnaouti et al., 2007; Prag et al., 2007). In adult mice, muskelin mRNA was detected at high level in liver, skeletal muscle, kidney and testis and at lower level in brain, heart, spleen and lung (Adams et al., 1998). A more detailed study on the expression of muskelin in the CNS (Tagnaouti et al., 2007) corroborated a widespread distribution of muskelin transcripts in the CNS and moreover showed that protein levels were in good agreement with the transcript levels. Similar to the widespread expression in mice, the *Human Protein Atlas* reports a medium to high expression of muskelin in the majority of analyzed tissues (48 out of 80), with a strong expression in most parts of the gastrointestinal tract, in the tubuli of the kidneys, in the parathyroid glands and in a subset of neuronal cells, especially the in Purkinje cells (figure 1.5 on the next page).

On the subcellular level, muskelin was repeatedly found to be localized to the cytoplasm, and to a lower degree the nucleus, but not the nucleoli (Adams et al. (1998); Hasegawa et al. (2000); Ledee et al. (2005); Tagnaouti et al. (2007); Valiyaveettil et al. (2008); Heisler et al. (2011), and the *Human Protein Atlas*). In cultured hippocampal neurons, the endogenously expressed muskelin displayed a punctate distribution throughout all parts of the neurons, often close to cell margins in the neurite projections, but was also present in a significant fraction in the nucleus (Tagnaouti et al., 2007). When overexpressed in different mammalian cell lines, muskelin showed a tendency to form small cytoplasmic particles (Prag et al., 2004), which was proposed to reflect the ability of muskelin to self-associate.

Expression level

☐ not detectable
▨ low
▨ medium
▪ high

Tissue

Cerebral cortex
Hippocampus
Lateral ventricle
Cerebellum
Nasopharynx
Oral mucosa
Salivary gland
Esophagus
Tonsil
Thyroid gland
Parathyroid gland
Bone marrow
Lymph node
Lung
Heart muscle
Breast
Liver
Gallbladder
Pancreas
Spleen
Adrenal gland
Kidney
Stomach
Duodenum
Small intestine
Appendix
Colon
Rectum
Urinary bladder
Testis
Epididymis
Prostate
Seminal vesicle
Vagina
Cervix
Endometrium
Fallopian tube
Ovary

Figure 1.5 – Expression pattern of muskelin in the human body based on the entry for muskelin in the *Human Protein Atlas*. Experimental evidence is based on a single antibody used for staining and therefore rated as uncertain. Soft tissues (no expression except for medium expression in skin) are not shown.

Unexpectedly based on the lack of obvious paralogs, the substantial degree of sequence conservation (see next section) and its ubiquitous expression, a loss of muskelin in knockout mice results only in a very mild phenotype (Heisler et al., 2011). The muskelin knockout mice are viable and fertile, with no major changes in development, anatomy and behavior. The most prominent phenotypic feature is a lightening of the fur color that homozygous knockout mice undergo shortly after birth. The initially black animals develop a markedly dilute color. Less obvious changes are enhanced surface levels of GABA$_A$R α1 and an altered form of network oscillations in the hippocampus, both providing evidence for a role of muskelin in the transport of GABA$_A$ receptors. So far, it cannot be excluded that the loss of muskelin is clinically relevant, as it might be compensated during development and may be revealed only under specific challenges. In this light, it is interesting to note that muskelin has been found to be upregulated under certain physiological stress conditions, such as hyperosmolarity (Boyd et al., 2005) and transient ischemia in the brain (Dhodda et al., 2004).

### 1.2.3 Sequence conservation in muskelin

Muskelin was described to constitute a unique family of proteins which is found only within the opisthokont lineage of eukaryotes (Adams, 2002; Prag et al., 2004), and this assignment is confirmed by latest phylogenetic analyses (figure 1.6 on the facing page). Within the opisthokont lineage, muskelin is not present in all species, as for example no homolog is found in *S. cerevisiae* and in nematodes. Yet, when present, muskelin shows a substantial degree of conservation, even between different kingdoms of eukaryotes, for example, the sequence identity between the human and the *S. pombe* ortholog is 28%. The conservation is almost perfect among mammals and still very high among vertebrates, for instance the sequence identity between human and zebrafish muskelin amounts to 89%.

Figure 1.6 *(facing page)* – **Sequence conservation in muskelin.** *Top:* Phylogenetic tree calculated based on a set of sequences identified by a *BLAST* search of the *UniProt* database with the human muskelin, manually pruned of isoforms and fragments. The final set contained 102 sequences, but for better readability, not all species are shown in the tree. Fungal species are shaded in blue, species belonging to *Animalia* are shaded in brown or red according to the phylum (A: *Annelida*, E: *Echinodermata*, Cn: *Cnidaria*, Pl: *Placozoa*, Po: *Porifera*), chordates are shaded in tints of red according to their class. For typical model organisms, the sequence identity to the human ortholog is indicated. *Bottom:* The domain architecture of human muskelin (*top*, Discoidin, LisH, CTLH and kelch repeat (KR) domains) and of the *Schizosaccharomyces pombe* (*S. pombe*) ortholog (*bottom*, the positions of the LisH and CTLH motifs are inferred from the sequence alignment as indicated) are shown next to the predicted secondary structure (ss pred, with β-strands in light and α-helices in dark grey) and disordered regions (do pred, in brown) obtained by *PROFphd (PredictProtein)* and *Phyre*[2], respectively. In the middle, a heat-plot of conservation (blue - low, red - high) is included, which was calculated based on the alignment used for creating the tree and is plotted on the human sequence.
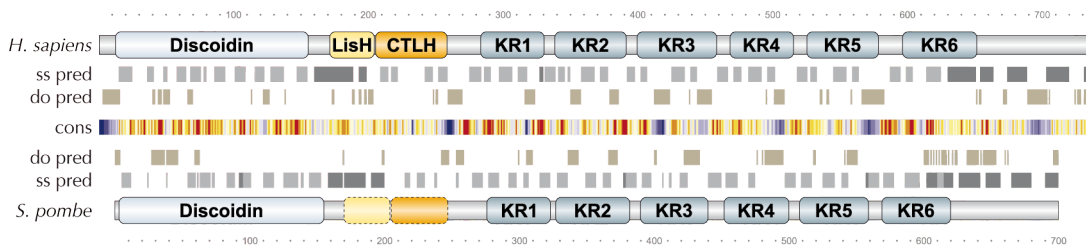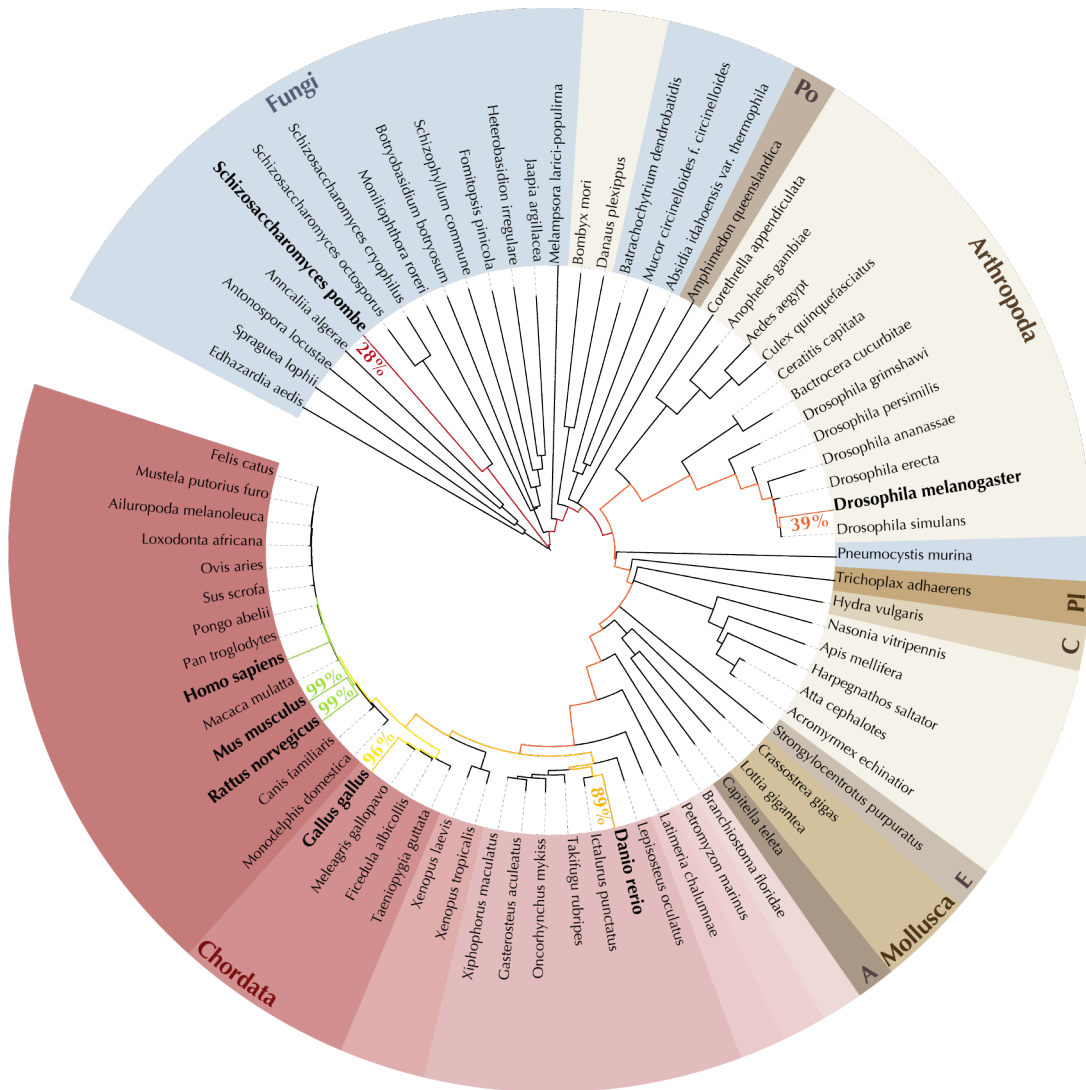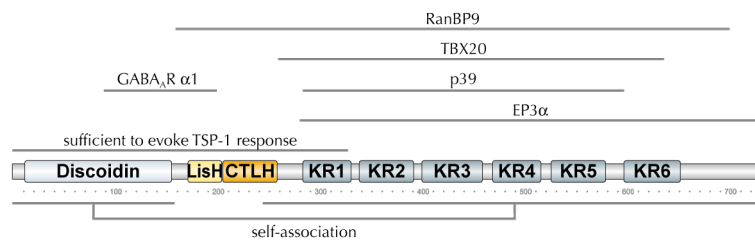
Figure 1.7 – Interaction sites in muskelin of binding partners that have been mapped indicated by thin black lines above and parts of muskelin involved in head-to-tail binding indicated below the domain architecture of human muskelin as annotated in the *UniProt* database.

Along the amino acid sequence of muskelin, the degree of conservation is rather uniform, with the exception of the N-terminus and a few short stretches that correspond to disordered loops. Notably, the distinctive region of high conservation (residues 131 to 152 in human sequence; termed MIND motif) postulated by Adams (2002), which was based on the comparison of the human, mouse, rat and *Drosophila melanogaster* sequences, is not recognizable when the much larger set of homologs available to date is analyzed (compare figure 1.6 on the previous page). Broken down to the domains in muskelin (section 1.2.4), the highest degree of conservation is found in the Discoidin domain and in the kelch repeat domain, while the LisH motif, the CTLH motif and the C-terminal module are less conserved. Accordingly, the overall domain architecture is well conserved throughout muskelin homologs, too, although the LisH motif and the CTLH motif are not found in all homologs, likely due to the lower conservation obliterating the family resemblance. Even one of the most distant relatives to mammalian muskelin, the *S. pombe* homolog, harbors a discoidin domain directly followed by a helical motif, by a short β-stranded segment, then the kelch repeat domain and finally a C-terminal helical portion, just as its mammalian counterparts.

1.2.4 The domain architecture

Muskelin features a characteristic domain composition, the first domain described upon the discovery of muskelin, the kelch repeat domain, even became part of muskelin's name[12]. Subsequently, the LisH and CTLH motif (Emes and Ponting, 2001) and the discoidin domain (Prag et al., 2004) were assigned to the muskelin sequence, leaving only the C-terminal part of muskelin unannotated. This domain composition is unique to muskelin and has been well conserved in muskelin throughout evolution (see section 1.2.3). Overall, muskelin contains few and rather short linker segments and its domains are in close succession, indicating a rather compact fold of the protein.

---

[12] Taken from Adams et al. (1998): "The novel protein identified herein has been named muskelin, to indicate its identification in a mouse muscle cell line and the presence of kelch-like motifs within the amino acid sequence."
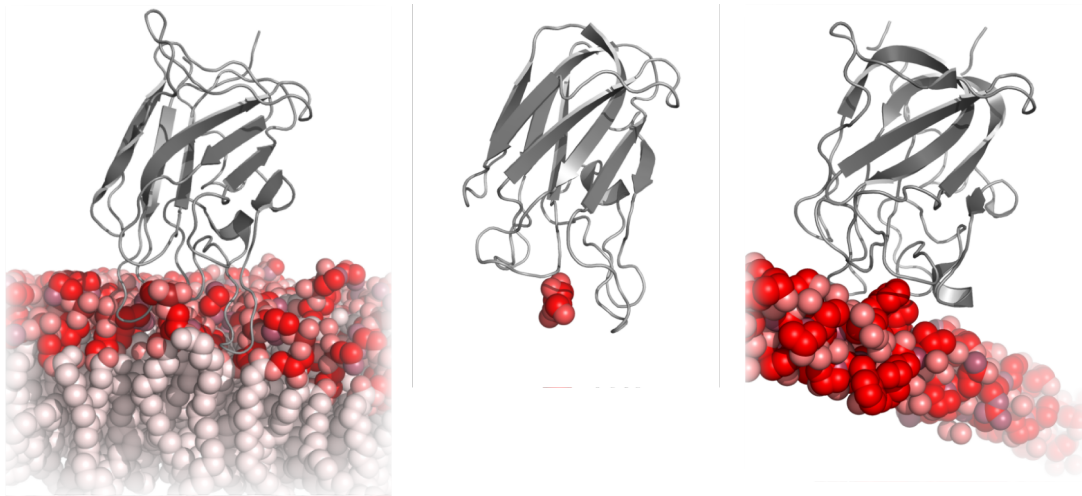
Figure 1.8 – Discoidin domains bind to a diverse set of ligands. Discoidin domains of coagulation factor V (*left*), sialidase (*middle*) and discoidin domain receptor 2 (*right*) in grey are shown with their respective ligands, phospholipids (modeled), β-D-galactose and collagen, in red. (PDB codes: 1czt, 2bzd, 2wuh)

Notably, the N-terminal discoidin domain was shown to bind to the C-terminal part in muskelin, and this interaction was suggested to underly a regulated oligomerization of muskelin (Prag et al., 2004, 2007). Indeed, recombinant muskelin was later shown to form a large oligomer *in vitro* (Kiedzierska et al., 2008).

The multidomain architecture has fueled speculations that muskelin acts as a scaffolding or adaptor protein, binding various partners via its different domains. The mapping of the identified interactions shows a prevalence of most partners binding to the Kelch repeat domain (figure 1.7). Yet, a simultaneous interaction with two different partners has not been demonstrated so far.

The discoidin domain

Preceded by only a short, unstructured and variable sequence, the most N-terminal domain of muskelin is a discoidin domain, also known as F5/F8 type C domain (Prag et al., 2004). Generally, discoidin domains are 150 amino acids long and fold into a distorted β-sandwich. The members of the discoidin domain family are evolutionarily widespread and are found in bacteria as well as eukaryotes[13]. Discoidin domains have evolved as versatile binding modules, with a diverse set of ligands including carbohydrates as for sialidase (Newstead et al., 2005), lipids as for the coagulation

---

[13] Sequences assigned to PF00754 in the *PFAM* database: 60.06% bacterial, 39.57% eukaryotic; 0.03% archeal, 0.3% viral, 0.04% unclassified. Sequences assigned to SM00231 in the *SMART* database: 19.9% bacterial, 79.92% eukaryotic, 0.02% archeal, 0.07% viral, 0.10% unclassified.

factors V and VIII (Macedo-Ribeiro et al., 1999; Pratt et al., 1999) and proteins as for the discoidin receptor 2 (Carafoli et al., 2009) or neuropilin (Vander Kooi et al., 2007) (figure 1.8). Their versatility is based on a conserved β-sandwich core fold that provides the framework for a very variable binding site formed by the inter-strand loops on the bottom of the sandwich (β1-β2, β3-β4 and β7-β8). For the majority of discoidin domains, the primary interaction site is formed by the β1-β2 and the β3-β4 loop, which can be so salient that they are referred to as 'spikes', for example in the coagulation factors V and VIII (Macedo-Ribeiro et al., 1999; Pratt et al., 1999). The β7-β8 loop is often rather buried and forms the base of the binding pocket, but it contributes significantly to the binding of the ligand in many cases (Abbott et al., 2007; Carafoli et al., 2009; Mathieu et al., 2010).

The LIS1 homology motif

The LisH motif succeeding the discoidin domain is a roughly 30 amino acids long motif consisting of two α-helices, that is almost exclusively found in eukaryotic proteins[14]. The crystal structures of the LisH-containing proteins, LIS1 (Kim et al., 2004), fibroblast growth factor receptor 1 oncogene partner (FOP) (Mikolajka et al., 2006) and transducin β-like protein 1 (TBL1) (Oberoi et al., 2011), illustrate that the LisH motif is a dimerization motif, forming a four-helix bundle upon dimerization (figure 1.9 on the facing page). The dimer interfaces typically have a mixed composition, but hydrophobic residues prevail. The structures moreover reveal an important role for a highly conserved glutamate at the end of the second helix. In line with its high degree of conservation, this residue contributes significantly to the stability of the LisH dimer by reaching across the interface and bridging the second helices with hydrogen bonds to the backbone amides of the opposing helix and a charged interaction with the helix dipole. For the LisH motif in muskelin, a possible dimerization and contribution to oligomerization had neither been discussed nor investigated so far.

The C-terminal to LisH motif

Directly in succession to the LisH motif, a CTLH motif is ensuing. CTLH motifs have so far been identified only in eukaryotic proteins. As indicated by its name, it is often found directly C-terminal to a LisH motif, but both the CTLH and the LisH motif also occur in absence of each other. The striking cumulative occurrence of this motif in proteins organized within one complex has resulted in naming of the mammalian form of this putative E3 ligase complex as CTLH-complex (see section 1.2.1). Neither

---

[14] Sequences assigned to PF08513 in the *PFAM* database: 0.15% bacterial, 99.85% eukaryotic. Sequences assigned to SM00667 in the *SMART* database: 1.24% bacterial, 98.72% eukaryotic, 0.03% archeal, 0.01% viral.
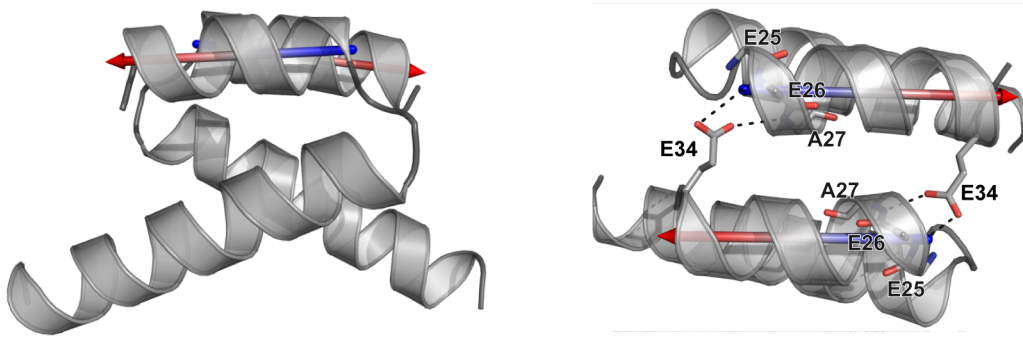
Figure 1.9 – Architecture of LisH motifs as illustrated by Lis1 as example, which is shown in ribbon representations in side (*left*) and top view (*right*). In top view, the side chain of Glu34 and the backbone of the first residues in the second helix are shown in stick representation with hydrogen bonds indicated by dashed lines. The vectors along the axes of the second helices are shown colored in a blue to red gradient. (PDB code: 1uuj)
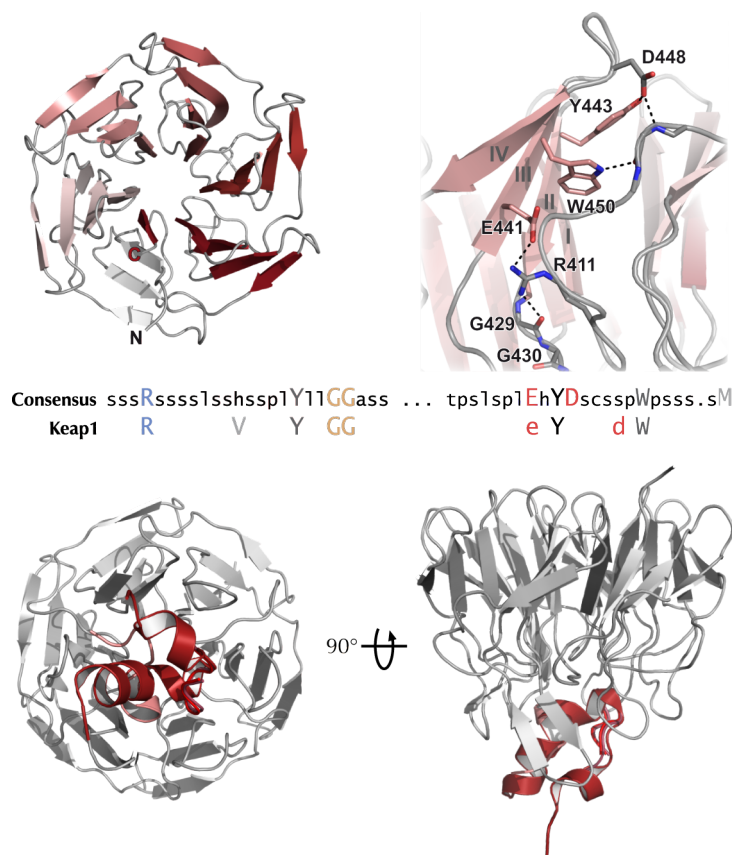
a function, nor a structure is known for the CTLH motif so far but the consensus motif is predicted to adopt an α-helical fold. The CTLH motif in muskelin seems to be atypical as it is predicted to fold into four β-strands.

The kelch repeat domain

The largest domain in muskelin is a kelch repeat domain, formed by six repeats of the kelch motif. The kelch motif is an evolutionarily old, 44 to 56 amino acid long motif found in archaea, bacteria and eukaryotes[15]. The signature sequence of kelch motifs is a short stretch of four hydrophobic amino acids preceding a glycine doublet that is followed in a variable distance by a pair of aromatic residues, typically a tyrosine, and a highly-conserved tryptophane, separated by six residues from each other (Xue and Cooley, 1993; Bork and Doolittle, 1994).

Kelch motifs occur as groups of five to seven repeats, together forming a kelch repeat domain. They adopt the tertiary structure of a β-propeller (figure 1.10 on the next page) in which each repeat forms one twisted four-stranded, antiparallel β-sheet corresponding to one blade of the propeller (Bork and Doolittle (1994), based on Ito et al. (1991)). Additional atypical blades, not formed by a kelch repeat, can be part of the β-propeller assembly, as illustrated by the crystal structure of the protein kelch-related protein 1 (Krp1) (Gray et al., 2009). Despite the ubiquitous distribution of kelch domain containing proteins, the structures of kelch repeat domains from

---

[15] The *PFAM* database lists 6 kelch motifs that potentially have redundant sets of sequences and therefore are not considered here. Sequences assigned to SM00612 in the *SMART* database: 9.51% bacterial, 89.36% eukaryotic, 0.17% archeal, 0.96% viral, 0.01% unclassified kelch motif sequences.

Consensus sssRsssslsshssplYllGGass ... tpslsplEhYDscsspWpsss.sM
Keap1   R   V   Y   GG   e Y   d W

Figure 1.10 – Architecture and ligand binding properties of kelch repeat domains illustrated by Keap1 as example. *Top*: Kelch repeat domain in ribbon representation with β-strands colored in grey to red from N- to C-terminus viewed from top (left) and close-up view of the third blade with highly conserved residues, their hydrogen-bonded interaction partners indicated (*left*). Below, the 50% consensus sequence of SM00612 aligned with the consensus of the Keap1 kelch motifs derived by Li (2004). Conserved amino acids are colored according to their grouping (see Appendix, section 6.2). *Bottom*: Kelch repeat domain (grey) with different ligands bound (red) in side and bottom view, obtained by structural alignment of the kelch repeat domain in Keap1-Nrf2, Keap1-Prothymosin α, Keap1-sequestosome-1 and KLHL3-WNK4 peptide complexes. (PDB codes: 3wn7, 2z32, 3ade, 4ch9)

only ten proteins have been solved to date[16]. Most kelch repeat domains use a closing mechanism to stabilize the circular arrangement of the blades: the first or last β-sheet consists of only three strands and is completed by a β-strand from the C- or N-terminus, respectively. Therefore, these arrangements are referred to as C-terminal or N-terminal closure mechanisms, respectively[17].

The most conserved residues in the kelch motif hold key structural roles: The conserved glycine doublet allows for unusual (and unaccessible to all other amino acids)

---

[16] Galactose oxidase from *Gibberella zeae*, YjhT from *Escherichia coli* (*E. coli*), Krp1 from rat, human KLHL2, human KLHL7, human KLHL12, human KLHL40, rat KLHL41 and Keap1 from human and mouse, the last of which (starting with human KLHL2) all belong to the family of BTB-Kelch proteins (Canning et al., 2013).

[17] The protein based on whose sequence the kelch repeat was initially defined (Xue and Cooley, 1993) and subsequently identified as common sequence pattern (Bork and Doolittle, 1994) uses an C-terminal closing mechanism. This led to an initial definition of the kelch repeat motif starting with the second of the four β-strands within one blade, omitting the first one. This is still reflected in the motif defined in the SMART database to date as well as the kelch 3 motif in the PFAM database.

conformations of the backbone, and moreover engages in hydrogen bonds, especially with a likewise highly conserved arginine in the loop connecting two blades that, in turn, interacts with a further residue (typically a glutamate or aspartate) in the same blade as the glycine doublet. This intrablade hydrogen-bond network likely sustains the folding of the individual blade. The conserved tyrosine/tryptophane pair significantly contributes to a well-packed hydrophobic interblade core and additionally these residues are intimately involved in hydrogen bonding to the backbone in the neighboring blade, thereby linking one blade to the next.

Notably, there is a markable difference in the two faces of the propeller: the $\beta1$-$\beta2$-loops and $\beta3$-$\beta4$-loops on one face are short, while the $\beta2$-$\beta3$ loops and the loops connecting two repeats, the $\beta4$-$\beta'1$-loops, on the other face are rather extended. These long loops seem to be of special importance for the function of kelch repeat proteins, as they form the binding site in most kelch repeat domain-protein complexes that were structurally characterized so far[18] or harbor the catalytic site[19]. This suggests a preferred interaction site for kelch-ligand/substrate interactions, but its significance is limited by the small number of kelch repeat domain complexes that have been characterized in detail. The Krp1 protein is an exception to the rule, as the Lasp1 binding site was mapped to the side of the β-propeller.

Given the low primary sequence identity of kelch repeat motifs, and the substantial structural variation in the loops that form the putative primary active/binding site, it is not surprising that kelch repeat domains are found in a variety of functional contexts, like enzymatic activity in galactose oxidase, or as versatile protein binding module, for example in the family of BTB-Kelch proteins (Canning et al., 2013). In muskelin, the kelch repeat domain seems to be used as protein binding module, as it contributes to the binding for the majority of known interaction partners (see figure 1.7 on page 16).

The C-terminal module

The C-terminal part of muskelin is unique, it constitutes the only major region to which no motif or domain can be assigned. It is predicted to fold into several α-helices with rather short intermediary linkers, suggesting that it adopts a well-defined tertiary structure. The C-terminal module of several vertebrate muskelin homologs harbors a recognition motif of serine/threonine kinases, whose phosphorylation status is implied to determine the binding to RanBP9 and the subcellular targeting of muskelin (Valiyaveettil et al., 2008), but the structural basis of these effects is so far unknown.

---

[18] That is the interaction of KLHL3 and WNK4 (Schumacher et al., 2014), Krp1 and the LIM and SH3 domain protein 1 (Lasp1) (Gray et al., 2009), Keap1 and Nrf2 (Padmanabhan et al., 2006; Lo et al., 2006; Tong et al., 2007; Fukutomi et al., 2014), Keap1 and Prothymosin α (Padmanabhan et al., 2008) as well as Keap1 and sequestosome-1 (Komatsu et al., 2010; Ichimura et al., 2013).

[19] As in galactose oxidase (Ito et al., 1991) and supposedly in YjhT (Severi et al., 2008).

1.3 Aim of the work

At the beginning of this work, most of the studies on muskelin focussed on the iden-
tification of a novel interaction, in some cases extended for its functional relevance.
Few studies dealt with the biochemical and cell biological basis of muskelin function,
and none had provided structural information. The ability of muskelin to undergo
a head-to-tail interaction had been suggested to underly muskelin oligomerization
(Prag et al., 2004) and recombinant muskelin was shown to form oligomers *in vitro*
(Kiedzierska et al., 2008) but these observations had not been brought into relation.
More precisely, neither a contribution of the head-to-tail interaction to the assembly
of the *in vitro* oligomer nor a physiological relevance of the oligomerization observed
in vitro were demonstrated. Moreover, a putative role of the LisH motif in muskelin
oligomerization had not been considered. As the oligomerization was suggested as
target in the regulation of muskelin function, but the potential consequences of an
altered oligomerization remained to be investigated, it was of special interest to gain
a comprehensive understanding of oligomerization, the underlying interactions and
the possibilities to influence it.

The major aim of this thesis was a biochemical and to the extent deemed possible
structural characterization of muskelin, particularly its oligomerization, to provide
the basis for a mechanistic understanding of its function. In addition, the interaction
between muskelin and the GABA$_A$ receptor was intended to be analyzed in detail using
isolated recombinant proteins with a view to assess the basis of subunit specificity and
potential ways of regulation.

Materials and Methods

2.1 Materials

2.1.1 Consumables

Chemicals

Chemicals were purchased at analytical grade or better. For the preparation of buffers and solutions, reagent-grade water provided by a TKA GenPure system was used.

Name	CASRN*	Supplier
2'-Deoxyadenosine 5'-triphosphate (dATP), sodium salt solution	1927-31-7	New England Biolabs
2'-Deoxycytidine 5'-triphosphate (dCTP), sodium salt solution	102783-51-7	New England Biolabs
2'-Deoxyguanosine 5'-triphosphate (dGTP), sodium salt solution	93919-41-6	New England Biolabs
2'-Deoxythymidine 5'-triphosphate (dTTP), sodium salt solution	18423-43-3	New England Biolabs
2-Propanol	67-63-0	Carl Roth
Acetic acid	64-19-7	Carl Roth
Agarose NEEO ultra quality	9012-36-6	Carl Roth
Agarose HEEO ultra quality	9012-36-6	Carl Roth
Ammonium citrate tribasic	3458-72-8	Sigma-Aldrich
Ampicillin sodium salt	69-52-3	Carl Roth
Ammonium persulfate	7727-54-0	Carl Roth
Bis[sulfosuccinimidyl] suberate (BS^3)	82436-77-9	Thermo Scientific
Bis[2-hydroxyethyl]-amino-tris[hydroxymethyl]-methane	6976-37-0	Carl Roth
Bromophenol blue	115-39-9	Carl Roth
Chloramphenicol	56-75-7	Carl Roth
Coomassie Brilliant Blue G-250	6104-58-1	Carl Roth
Coomassie Brilliant Blue R-250	6104-59-2	Carl Roth
Disodium hydrogen phosphate (Na_2HPO_4)	7558-79-4	Carl Roth
Dithiothreitol (DTT)	3483-12-3	Carl Roth
Ethylenediaminetetraacetic acid (EDTA)	60-00-4	Carl Roth

Name	CASRN*	Supplier
Ethanol	64-17-5	Carl Roth
Ethidium bromide	1239-45-8	Carl Roth
Ethylene glycol	107-21-1	Sigma-Aldrich
Glycerol	56-81-5	Carl Roth
Glycine	56-40-6	Carl Roth
Guanidinium chloride	50-01-1	Carl Roth
Hydrochloric acid (HCl)	7647-01-0	Carl Roth
Hydrogen peroxide (H_2O_2)	7722-84-1	Carl Roth
Imidazole	288-32-4	Carl Roth
Isopropyl-β-D-thiogalactopyranoside (IPTG)	367-93-1	Carl Roth
Kanamycin sulfate	25389-94-0	Carl Roth
Orange G	1936-15-8	Sigma-Aldrich
Phenylmethylsulfonyl fluoride (PMSF)	329-98-6	Carl Roth
Piperazine-N,N′-bis[2-ethanesulfonic acid]	5625-37-6	Carl Roth
Polyethylene glycol (PEG) 2000 monomethyl ether	9004-74-4	Sigma-Aldrich
PEG 3350	25322-68-3	Sigma-Aldrich
Potassium chloride (KCl)	7447-40-7	Carl Roth
Potassium dihydrogen phosphate (KH_2PO_4)	7778-77-0	Carl Roth
Sodium acetate trihydrate	6131-90-4	Sigma-Aldrich
Sodium citrate tribasic dihydrate	6132-04-3	Sigma-Aldrich
Sodium chloride (NaCl)	7647-14-5	Carl Roth
Sodium dodecyl sulfate (SDS)	151-21-3	Carl Roth
Sodium malonate dibasic monohydrate	26522-85-0	Sigma-Aldrich
Sodium dihydrogen phosphate (NaH_2PO_4)	7558-80-7	Carl Roth
Sodium sulfate (Na_2SO_4)	7757-82-6	Carl Roth
Sypro® Orange	*not assigned*	Sigma-Aldrich
Tetramethylethylenediamine	110-18-9	Carl Roth
Trifluoroacetic acid	76-05-1	Carl Roth
Tris[2-carboxyethyl]phosphine (TCEP)	51805-45-9	Carl Roth
Tris[hydroxymethyl]-aminomethane (Tris)	77-86-1	Carl Roth
Triton® X-100	9002-93-1	Carl Roth
Xylene cyanol	2650-17-1	Sigma-Aldrich

*Chemical Abstracts Service Registry Number (CASRN) is a registered trademark of the american chemical society

Manufactured composites

Name	Type	Supplier
Bayer silicon grease medium viscosity	Silicon grease	Jena Biosciences
cOmplete, EDTA free	Protease inhibitor cocktail	Roche
GeneRuler™ 1 kb DNA Ladder	DNA molecular-weight size marker	Thermo Fisher Scientific
GeneRuler™ 100 bp DNA Ladder	DNA molecular-weight size marker	Thermo Fisher Scientific
Lysogeny broth (LB) medium (Lennox)	Standard medium for *E. coli* culture	Roth

Name	Type	Supplier
Midori green Advance DNA stain	DNA stain	Biozym Scientific
PageRuler™ Prestained Protein Ladder	Protein molecular-weight size marker	Thermo Fisher Scientific
PageRuler™ Plus Prestained Protein Ladder	Protein molecular-weight size marker	Thermo Fisher Scientific
peqGOLD PCR Mastermix S	DNA polymerase reaction mastermix	peqlab
Rotiphorese® Gel 30 (37.5:1)	30% acrylamide/bisacrylamide solution	Carl Roth
Spectra™ Multicolor Broad Range Protein Ladder	Protein molecular-weight size marker	Thermo Fisher Scientific

Disposable labware

The following list details the disposable labware used, but basic laboratory equipment (such as reaction tubes, weighing dishes) is interchangeable and therefore not included.

Type	Model	Supplier
24-well plate	Crystalgen SuperClear™ Plate	Jena Biosciences
96-well crystallization plate	Crystalquick™ 1 square well, flat bottom, low profile	Greiner Bio-One
96-well thin wall PCR plate	Microplate 96 well, PP, for PCR	Greiner Bio-One
Centrifugal concentrator	Amicon® Ultra-4 and -15	Millipore
Centrifugal concentrator	Vivaspin 500, 6 and 20	Sartorius
Cuvette	UVette®	Eppendorf
Cover slides	22 mm circular cover slides - siliconized	Jena Biosciences
Dialysis cassette	Slide-A-Lyzer	Thermo Scientific
Dialysis membrane	Spectra/Por®	Spectrum Laboratories
Dialysis tube	D-Tube™ Dialyzer	Novagen
Optical quality sealing foil	VIEWseal™	Greiner Bio-One

Escherichia coli strains

All cells used were chemically competent, that is pretreated with calcium chloride to facilitate attachment of plasmid DNA to the cell membrane.

Name	Genotype	Supplier
DH5α™	F⁻ φ80 *lacZ*ΔM15 Δ(*lacZYA-arg*F)U169 *rec*A1 *end*A1 *hsd*R17(r_K^-, m_K^+) *pho*A *sup*E44 λ⁻ *thi*-1 *gyr*A96 *rel*A1	Invitrogen
BL21-CodonPlus® (DE3)-RIL	*E. coli* B F⁻ *ompT hsdS*(r_B^- m_B^-) *dcm*⁺ Tet^r *gal* λ(DE3) *endA* Hte[*argU ileY leuW* **Cam^r**]*	Stratagene

*Concentration of antibiotic used for selection: chloramphenicol 34 µg/ml

Enzymes

Name	Type	Supplier
Bovine Serum Albumin (BSA)	Carrier protein (used as concentration standard)	New England Biolabs
BspEI	Restriction endonuclease	New England Biolabs
BspHI	Restriction endonuclease	New England Biolabs
CIP	Nonspecific phosphatase	New England Biolabs
DNase I	Deoxyribonuclease	Invitrogen
dtUD1	Protease	*self-purified**
EcoRI-HF®	Restriction endonuclease	New England Biolabs
Ferritin	Carrier protein (used as size standard)	GE Healthcare
NcoI-HF®	Restriction endonuclease	New England Biolabs
NotI-HF®	Restriction endonuclease	New England Biolabs
peqGOLD *Taq*	DNA polymerase	peqlab
Phusion®	High-Fidelity DNA polymerase	New England Biolabs
RecA	Single-stranded DNA binding protein	New England Biolabs
SalI-HF®	Restriction endonuclease	New England Biolabs
SacI	Restriction endonuclease	New England Biolabs
T4 DNA Polymerase	DNA polymerase	New England Biolabs
Trypsin	Protease	Carl Roth
Quick T4 DNA Ligase	DNA ligase	New England Biolabs
XhoI	Restriction endonuclease	New England Biolabs

*see protocol on page 41

Kits

Name	Purpose	Supplier
NucleoSpin® Gel and PCR Clean-up kit	Extraction of DNA fragment from gels and purification of PCR products	MACHEREY-NAGEL
NucleoSpin® Plasmid	Small-scale isolation of plasmid DNA from *E. coli* cells	MACHEREY-NAGEL
NucleoBond® Xtra Midi	Medium-scale isolation of plasmid DNA from *E. coli* cells	MACHEREY-NAGEL

Screens

Name	Type	Supplier
Additive Screen™	Optimization screen, compound screen	Hampton Research
Crystal Screen	Crystallization screen, primary sparse matrix screen	Hampton Research
Crystal Screen 2	Crystallization screen, primary sparse matrix screen	Hampton Research
Detergent Screen™	Optimization screen, detergent screen	Hampton Research
Index	Crystallization screen, primary diverse reagent screen	Hampton Research

Name	Type	Supplier
OptiMix™ 3	Crystallization screen, primary sparse matrix screen	Fluidigm
OptiMix™-PEG	Crystallization screen, primary sparse matrix screen	Fluidigm
PEGs Suite	Crystallization screen, primary diverse reagent screen	Qiagen
Protein Complex Suite	Crystallization screen, primary sparse matrix screen	Qiagen
Salt Screen	Crystallization screen, salt screen for optimization	*Self-designed**
ThermoFlour screen	Optimization screen, systematic buffer screen	Based on Ericsson et al. (2006)*
Wizard Screen 1	Crystallization screen, primary sparse matrix screen	Emerald BioSystems
Wizard Screen 2	Crystallization screen, primary sparse matrix screen	Emerald BioSystems

*Compositions are given in the appendix, section 6.3 on page 152

Peptides

All peptides in the following list were readily soluble in aqueous buffers at pH 7.5 - 8.0, except for Mkln (G475-D484), which was first dissolved in TFA and afterwards diluted into the respective buffer.

Origin	Name	Sequence	Modification	Supplier
Mkln (G280-D289)	KR1 β2-β3	GWDGTQDLAD	N-terminal biotinylation	GenScript
Mkln (R315-H328)	KR1 β4-β1'	RDTEKENGPSARSC	N-terminal biotinylation	GenScript
Mkln (R345-S359)	KR2 β2-β3	RYLDSSVRNSKSLKS	N-terminal biotinylation	GenScript
Mkln (E376-H390)	KR2 β4-β1'	EDTAADGGPKLVFDH	N-terminal biotinylation	GenScript
Mkln (G405-G427)	KR3 β2-β3	GGRILTCNGSVDDSRA SEPQFSG	N-terminal biotinylation	GenScript
Mkln (R442-G458)	KR3 β4-β1'	REDSCNAGPEDIQSRIG	N-terminal biotinylation	GenScript
Mkln (G475-D484)	KR4 β2-β3	GQRSKTYLND	N-terminal biotinylation	GenScript
Mkln (G501-Q516)	KR4 β4-β1'	GTKKDSGMVPMTGFTQ	N-terminal biotinylation	GenScript
Mkln (G532-N546)	KR5 β2-β3	GLSKDKEKREENVRN	N-terminal biotinylation	GenScript
Mkln (K563-F585)	KR5 β4-β1'	KNDQAAKENLSKSLQEEEPCPRF	N-terminal biotinylation	GenScript
Mkln (G602-D617)	KR6 β2-β3	GGNPGKSCSPKMRLDD	N-terminal biotinylation	GenScript

Plasmids

The following list contains preexisting plasmids. All used expression plasmids including those generated in the course of this work can be found in the list of expression plasmids in the appendix (section 6.4 on page 155).

Name	Description	Resistance*	Reference
pETM-11	*E. coli* expression vector providing N-terminal hexahistidine (6xHis)-tobacco etch virus (TEV) protease recognition site and a C-terminal 6xHis sequence	Kanamycin	Dümmler et al. (2005)

Name	Description	Resistance*	Reference
pETM-30	*E. coli* expression vector providing N-terminal 6xHis-glutathione S transferase (GST)-TEV protease recognition site and a C-terminal 6xHis sequence	Kanamycin	Dümmler et al. (2005)
pETM-41	*E. coli* expression vector providing N-terminal 6xHis-maltose-binding protein-TEV protease recognition site and a C-terminal 6xHis sequence	Kanamycin	Dümmler et al. (2005)
pETHSUL	*E. coli* expression vector providing an N-terminal 6xHis-small ubiquitin like modifier (SUMO)- , and a C-terminal 6xHis sequence	Ampicillin	Weeks et al. (2007)
pSUPER	*E. coli* expression plasmid of the doubly tagged catalytic domain of the *S. cerevisiae* ubiquitin-like-specific protease 1 (dtUD1) with an N-terminal 6xHis-SUMO-sequence	Ampicillin	Weeks et al. (2007)
Mkln-pGEX	*E. coli* expression plasmid of *Rattus norvegicus* (*R. norvegicus*) muskelin with an N-terminal GST-tag	Ampicillin	Heisler et al. (2011)
Mkln-pBK-CMV	Expression plasmid for expression of muskelin in eukaryotic cells	Kanamycin	Heisler et al. (2011)
mRFP-Mkln-pEGFP-C1′	Expression plasmid for expression of monomeric red fluorescent protein (mRFP)-muskelin in eukaryotic cells	Kanamycin	Heisler et al. (2011)
mCherry-Mkln-pEGFP-C1′	Expression plasmid for expression of mCherry-muskelin in eukaryotic cells	Kanamycin	Heisler et al. (2011)
α1L-pETM-11	*E. coli* expression plasmid of *R. norvegicus* GABA$_A$R α1$_{334\text{-}420}$ with an N-terminal 6xHis-TEV recognition site sequence	Kanamycin	Mukherjee et al. (2011)
α2L(V371G)-pETM-11	*E. coli* expression plasmid of *Homo sapiens* GABA$_A$R α2$_{335\text{-}419}$ V371G with an N-terminal 6xHis-TEV recognition site sequence	Kanamycin	*provided by H. Maric***
α3L-pETM-11	*E. coli* expression plasmid of *R. norvegicus* GABA$_A$R α3$_{360\text{-}458}$ with an N-terminal 6xHis-TEV recognition site sequence	Kanamycin	Tretter et al. (2011)
α5L-pETM-11	*E. coli* expression plasmid of *R. norvegicus* GABA$_A$R α5$_{342\text{-}429}$ with an N-terminal 6xHis-TEV recognition site sequence	Kanamycin	*provided by H. Maric***

*Concentrations of antibiotics used for selection: ampicillin 100 μg/ml, kanamycin 50 μg/ml

**AG Schindelin, Rudolf Virchow Center for Biomedical Research (RVZ), University of Würzburg.

Primers

5′-3′ sequence	Dir	Temp (pos)	Ext
Generation of pETM-SUMO			
ATACCATGAAACATCACCATCATCATCAC	fwd	6xHis-Sumo	HO
GTGATGATGATGGTGATGTTTCATGGTAT	rev	pET-backbone	HO

5'-3' sequence	Dir	Temp (pos)	Ext
GGATCCGGTACCACTAGTTACCATGGCGCCACCAATCTGTTCG	rev	Sumo	HO
GGTGGCGCCATGGTAACTAGTGGTACCGGATCCGA	fwd	pET-backbone	HO
Subcloning			
CGATCCATGGCGGCTGGTGGTGCTGTTGCT	fwd	Mkln (M1)	NcoI
CGTAGGATCCATGAAACATCACCATCACCA	fwd	Mkln (M1)	BamHI
CGTAGAATTCATGGCGGCTGGTGG	fwd	Mkln (M1)	EcoRI
CTGACCATGGAGTGCCGGCTTCTGC	fwd	Mkln (E12)	NcoI
GCATGCGGCCGCTTAATCAATGCCACTGAGTTCAAC	rev	Mkln (D156)	Stop NotI
CGTACCATGGATCCTGATATAGTGCAGCCC	fwd	Mkln (D157)	NcoI
GTCAGCGGCCGCCTAGTGTTGTCTGAAGTGTTTTAGAC	rev	Mkln (N187)	Stop NotI
CGTACCATGGAGCATCCAATGTTGAC	fwd	Mkln (E205)	NcoI
GCATGCGGCCGCTTACTCCAGTGCAATCTTGGTT	rev	Mkln (E205)	Stop NotI
GCATGCGGCCGCTCACTGTTGACTGATATACTGATTGAACAAGC	rev	Mkln (Q245)	NotI
CGTACCATGGAGTATAAGCCACGGTGGAGTCA	for	Mkln (E246)	NcoI
GCATCCATGGGTTATAAGCCACGGTGGAGTCA	fwd	Mkln (Y247)	NcoI
CTGAGCGGCCGCCTAATACTCCTGTTGACTGATATACTG	rev	Mkln (Y247)	Stop NotI
GCATGCGGCCGCTTACCCTCTCATTCCTGGACG	rev	Mkln (G272)	NotI
CGATCCATGGGCCATCAGATGGTCATTG	fwd	Mkln (G273)	NcoI
CGATCCATGGGACCTTCGAAGGACTACTTACTGAG	for	Mkln (P626)	NcoI
CTAGCGGCCGCTCATCTGCACAACTTCAGCGAC	rev	Mkln (R625)	Stop NotI
GCATGCGGCCGCTTACACGTCTGAAAAGCCTAAAGCTGT	rev	Mkln (V700)	Stop NotI
CTAGCCATGGTGGATCACACCTATGCTCA	fwd	Mkln (V700)	NcoI
CCTGGTAGACCTCATCACACTGGCGGCCGCTAGC	rev	Mkln (L735)	NotI
CCTGGTAGACCTCATCACACTGTGAGCGGCCGCTAGC	rev	Mkln (L735)	Stop NotI
GCATGTCGACCAGTGTGATGAGGTCTACCAGG	rev	Mkln (L735)	SalI
Mutagenesis			
GGCGCCCGAGTCTCGGCTTCTGCCCTAC	fwd	Mkln (C13S)	
GTAGGGCAGAAGCCGAGACTCGGGCGCC	rev	Mkln (C13S)	
GATGGTCTTCAGCGAGCAACTATCC	fwd	Mkln (E50A)	
GGATAGTTGCTCGCTGAAGACCATC	rev	Mkln (E50A)	
TCAGAGAGCAACGCTCCTCCCCAGTA	fwd	Mkln (Y53A)	
TACTGGGGAGGAGCGTTGCTCTCTGA	rev	Mkln (Y53A)	
TCAGAGAGCAACTATGAACCCCAGTACTTGATTC	fwd	Mkln (P54E)	
GAATCAAGTACTGGGGTTCATAGTTGCTCTCTGA	rev	Mkln (P54E)	
CTATCCTCCCCAGGAATTGATTCTGAAGC	fwd	Mkln (Y57E)	
GCTTCAGAATCAATTCCTGGGGAGGATAG	rev	Mkln (Y57E)	
GAGAAAACTCATGTCAGCAATTTGAAGAAATTC	fwd	Mkln (C82S)	
GAATTTCTTCAAATTGCTGACATGAGTTTTCTC	rev	Mkln (C82S)	
TTGATGAACAGATGTTCCCTTCTCGATTCA	fwd	Mkln (C128S)	
TGAATCGAGAAGGGAACATCTGTTCATCAA	rev	Mkln (C128S)	
GGGGCCCAGCGCTAATTTTAGCATC	fwd	Mkln (F143A)	
GATGCTAAAATTAGCGCTGGGCCCC	rev	Mkln (F143A)	
GGGCCCAGCTTTCGTTTTAGCATCTG	fwd	Mkln (N144R)	
CAGATGCTAAAACGAAAGCTGGGCCC	rev	Mkln (N144R)	

29

5'-3' sequence	Dir	Temp (pos)	Ext
GGGCCCAGCTTTTGGTTTAGCATCTG	fwd	Mkln (N144W)	
CAGATGCTAAACCAAAAGCTGGGCCC	rev	Mkln (N144W)	
GATATAGTGCAGCCCTCTCTGAACTGGTAC	fwd	Mkln (C164S)	
GTACCAGTTCAGAGAGGGCTGCACTATATC	rev	Mkln (C164S)	
GGTACAGCAAGTACGCCGAACAGGAAGCCAT	fwd	Mkln (R172A)	
ATGGCTTCCTGTTCGGCGTACTTGCTGTACC	rev	Mkln (R172A)	
CAGCAAGTACCGCGCACAGGAAGCAATTCG	fwd	Mkln (E173A)	
CGAATTGCTTCCTGTGCGCGGTACTTGCTG	rev	Mkln (E173A)	
GGAAGCCATTCGCCTTTCTCTAAAACACTTCAG	fwd	Mkln (C180S)	
CTGAAGTGTTTTAGAGAAAGGCGAATGGCTTCC	rev	Mkln (C180S)	
CAGGAAGCCATTCGCCTTCAGCTAAAACACTTCAGAC	fwd	Mkln (C180Q)	
GTCTGAAGTGTTTTAGCTGAAGGCGAATGGCTTCCTG	rev	Mkln (C180Q)	
CAGGAAGCTATTCGCCTTTGGCTAAAACACTTCAG	fwd	Mkln (C180W)	
CTGAAGTGTTTTAGCCAAAGGCGAATAGCTTCCTG	rev	Mkln (C180W)	
CCTTTGTCTAAAACACGCCAGACAGCACAACTATAC	fwd	Mkln (F184A)	
GTATAGTTGTGCTGTCTGGCGTGTTTTAGACAAAGG	rev	Mkln (F184A)	
CAAAAGCCTCTGTATAGTTGTGTTGTCTCTCGTGTTTTAGAC	rev	Mkln (F184E)	
CAAAAGCCTCTGTATAGTTGTGTTGTCTCTGGTGTTTTAGAC	rev	Mkln (F184Q)	
CAAAAGCCTCTGTATAGTTGTGTTGTCTCTTGTGTTTTAGAC	rev	Mkln (F184K)	
CATTCGCCTTTGTCTAGCAGCCTTCAGCAACAC	fwd	Mkln (K182A H183A)	
GTGTTGTCTGAAGGCTGCTAGACAAAGGCGAATG	rev	Mkln (K182A H183A)	
GTCTAAAACACTTCGCACAACACAACTATACAGAGGC	fwd	Mkln (R185A)	
GCCTCTGTATAGTTGTGTTGTGCGAAGTGTTTTAGAC	rev	Mkln (R185A)	
CACTTCAGACAACACGCGTATACAGAGGCTTT	fwd	Mkln (N188A)	
AAAGCCTCTGTATACGCGTGTTGTCTGAAGTG	rev	Mkln (N188A)	
CAACTATACAGAGGCTTTTGAATCCGAGCAGAAGAAAACC	fwd	Mkln (L196E)	
CAACTATACAGAGGCTTTTGAATCCCAACAGAAGAAAACC	fwd	Mkln (L196Q)	
CAACTATACAGAGGCTTTTGAATCCAAGCAGAAGAAAACC	fwd	Mkln (L196K)	
GAGAAAGAGAATGGTCCTGAGGCCAGATCATGTC	fwd	Mkln (S324E)	
GACATGATCTGGCCTCAGGACCATTCTCTTTCTC	rev	Mkln (S324E)	
GAGAAAGAGAATGGTCCTGCAGCCAGATCATGTC	fwd	Mkln (S324A)	
GACATGATCTGGCTGCAGGACCATTCTCTTTCTC	rev	Mkln (S324A)	
CCAATGACCGGATTCGAACAGAGAGCAACTATTG	fwd	Mkln (T515E)	
CAATAGTTGCTCTCTGTTCGAATCCGGTCATTGG	rev	Mkln (T515E)	
GACCGGATTCGCACAGAGAGCAACTATTG	fwd	Mkln (T515A)	
CAATAGTTGCTCTCTGTGCGAATCCGGTC	rev	Mkln (T515A)	

Alanine substitution mutagenesis

TGCCCAGAAAGCTGCAGCACCAAACAAATAAACAGTCTCTGTCTG	rev	Mkln (G280)	HO 3xA
TTGTTTGGTGCTGCAGCTTTCTGGGCATACAGTGTGAAGGAG	fwd	Mkln (D289)	HO 3xA
AGCTGCAGCGGCTGCTGCGGCGGCGGCTGCTGAAATACATGTCCACTGGTTCTCC	rev	Mkln (R315)	10xA
GCAGCCGCCGCCGCAGCAGCCGCTGCAGCTAAAATGTGCATTGATATTCAGCGGAGA	fwd	Mkln (H328)	10xA
GTAGAAATCAGCTGCAGCGCCCAATGTGTAGATTTGTCTCC	rev	Mkln (R345)	HO 3xA
ACATTGGGCGCTGCAGCTGATTTCTACCGTTACGACATTGAC	fwd	Mkln (S359)	HO 3xA
AGCTGCAGCGGCTGCTGCGGCGGCGGCTGCACTTAGTAACATCCAGGTGTTGGT	rev	Mkln (E376)	10xA

5'-3' sequence	Dir	Temp (pos)	Ext
GCAGCCGCCGCCGCAGCAGCCGCTGCAGCTCAGATGTGTATGGACTCAGAAAAG	fwd	Mkln (H381)	10xA
TGCAAACAGAGCTGCAGCAAAGGTATAGATCATATGCTTTTCTGAGTC	rev	Mkln (G405)	HO 3xA
CTATACCTTTGCTGCAGCTCTGTTTGCATTCAACTGCCAGTGT	fwd	Mkln (G427)	HO 3xA
AGCTGCAGCGGCTGCTGCGGCGGCGGCTGCAAGAAGTTTCCAGGTTTGACACTG	rev	Mkln (R442)	10xA
GCAGCCGCCGCCGCAGCAGCCGCTGCAGCTCACTGCATGCTGTTCCACTC	fwd	Mkln (G458)	10xA
ACTAAAAAAAGCTGCAGCACCAAACACGTATAAACAACGG	rev	Mkln (G475)	HO 3xA
GTGTTTGGTGCTGCAGCTTTTTTTAGTTATGATGTGGACTCTGATC	fwd	Mkln (D484)	HO 3xA
AGCTGCAGCGGCTGCTGCGGCGGCGGCTGCATCTGAAATTATGTCTACATGATCAGAG	rev	Mkln (G501)	10xA
GCAGCCGCCGCCGCAGCAGCCGCTGCAGCTAGAGCAACTATTGATCCAGAACTG	fwd	Mkln (Q516)	10xA
CCAGAATGAAGCTGCAGCGGATAAAACATGTATTTCATTCAGTTC	rev	Mkln (G532)	HO 3xA
GTTTTATCCGCTGCAGCTTCATTCTGGATTTATGACATTGTG	fwd	Mkln (N546)	HO 3xA
AGCTGCAGCGGCTGCTGCGGCGGCGGCTGCATAGACACAAGACCAGCTATTCCTC	rev	Mkln (K563)	10xA
GCAGCCGCCGCCGCAGCAGCCGCTGCAGCTGCCCATCAGCTCGTTTATG	fwd	Mkln (F585)	10xA
CGACCAGAAAGCTGCAGCAAATAAATAATGAACCTTGTGTAATTCATC	rev	Mkln (G602)	HO 3xA
TATTTATTTGCTGCAGCTTTCTGGTCGCTGAAGTTGTG	fwd	Mkln (D617)	HO 3xA
Subcloning			
GCTAGCGGCCGCAACTATTTCACCAAGAGAGG	fwd	GABA$_A$R α1 (N334)	NotI
GCTACCATGGCAAACTATTTCACCAAGAGAGG	fwd	GABA$_A$R α1 (N334)	NcoI
CGATGCGGCCGCTCATTCTTTCGGTTCTATGGTCGCAC	rev	GABA$_A$R α1 (E398)	Stop NotI
CGATGCGGCCGCTCATTCTGGCGGTTTTGTTTCAGGC	rev	GABA$_A$R α1 (E409)	Stop NotI
GCATCTCGAGTCGGTCGATTTTGCTGAC	rev	GABA$_A$R α1 (R420)	XhoI
GCATGCGGCCGCTCATTCAGCTGGCTTGTTTTCTGG	rev	GABA$_A$R α2 (E406)	Stop NotI
CGATCCATGGCAAACTACTTCACCAAGC	fwd	GABA$_A$R α3 (N360)	NcoI
CGATGCGGCCGCTCAAGCAGGACTGTCTTGCACATAAG	rev	GABA$_A$R α3 (A445)	Stop NotI
GCTAGCGGCCGCAACTACTTTACAAAGAGAGGATGGGC	fwd	GABA$_A$R α5 (N342)	NotI
CGATCCATGGCAAACTACTTTACAAAGAGAGG	fwd	GABA$_A$R α5 (N342)	NcoI
CGATGCGGCCGCTCAAGATGCTCTGATTGAGGCTGTACC	rev	GABA$_A$R α5 (S410)	Stop NotI
CGATGCGGCCGCTCACTCAGAAGTCTTCTCCTCAGATGC	rev	GABA$_A$R α5 (E416)	Stop NotI
GCATCTCGAGTTTGTCGATCTTGCTGATGCTG	rev	GABA$_A$R α5 (K428)	XhoI

Dir: direction of primer (fwd: forward, rev: reverse), **Temp (pos)**: Template and position (for fwd-primer = first codon included, for rev-primer = last codon included, for mutagenesis-primer = position of mutation), **Ext**: Extension (HO: homologous overhang, Stop: Stop codon, 3xA: 3 alanine codons, 10xA: 10 alanine codons)

2.1.2 Equipment

Instruments

The following list details the instruments used, but basic laboratory equipment (like pipettes, glassware, measuring cylinders) is interchangeable and therefore not listed.

Type	Model	Supplier
Agarose gel electrophoresis system (DNA)	Mini-Sub® Cell GT System	Bio-Rad Laboratories
Agarose gel electrophoresis system (Protein)	Vari-Gel ™ horizontal maxi system	Denville Scientific

Type	Model	Supplier
Analytical ultracentrifuge	Optima™ XL-I	Beckman Coulter
Analytical ultracentrifuge; cell assembly	Cell, assembled, sapphire windows, with Epon-charcoal centerpiece	Beckman Coulter
Analytical ultracentrifuge; rotor	An-50 Ti	Beckman Coulter
Autoclave	Systec V-150	Systec
Balance, analytical	XS 105 DR	Mettler-Toledo
Balance	XS 6002S DR	Mettler-Toledo
Circular dichroism (CD) spectropolarimeter	J-810	JASCO
CD cuvette	QS 1 mm	Hellma
Cell disruption system	M-110P	Microfluidics
Cell disruption system	TS Series Benchtop	Constant Systems Limited
Centrifuge	Rotilabo®-mini-centrifuge Uni-fuge	Carl Roth
Centrifuge	Centrifuge 5417 R	Eppendorf
Centrifuge	Centrifuge 5810 R	Eppendorf
Centrifuge	Avanti J26 XP	Eppendorf
Cryocrystallography - cryo loop	CryoLoop	Hampton Research
Cryocrystallography - sample holder	CrystalCap Magnetic™	Hampton Research
Cryocrystallography - sample vial	CryoVial	Hampton Research
Cryocrystallography - handling tool	CrystalWand™ Magnetic	Hampton Research
Cryocrystallography - sample basket	ESRF/EMBL Sample Changer Basket	Molecular Dimensions
Sample degassing and thermostat system	ThermoVac	MicroCal
FPLC system	ÄKTA avant 25	GE Healthcare
FPLC system	ÄKTA pure 25	GE Healthcare
FPLC system	ÄKTApurifier 10	GE Healthcare
FPLC system	ÄKTAxpress	GE Healthcare
Incubator shaker	LT-X	Kühner
Incubator shaker	ISF1-X	Kühner
Isothermal titration calorimeter	MicroCal ITC$_{200}$	MicroCal
Isothermal titration calorimeter	VP-ITC	MicroCal
Lighting panel	2 E	Carl Roth
Liquid handling robot	Honeybee 963	Digilab
Liquid handling robot	LISSY	Zinsser
Magnetic stirrer	MR 3002	Heidolph Instruments
Microscope	STEMI 2000-C	ZEISS
Microscope	SteREO Discovery.V12	ZEISS
Microscope camera	AxioCam MRc	ZEISS
Microscope light source	KL 2500 LCD	ZEISS
Microscope light source	CL 1500 Eco	ZEISS
MALS detector	DAWN® 8+ HELEOS® II	Wyatt Technology
pH-meter	pH-meter	SCHOTT
Polyacrylamide gel electrophoresis system	Mini-PROTEAN system	Bio-Rad Laboratories
Power supply	PowerPac™ Basic	Bio-Rad Laboratories
PCR-cycler	Mastercycler® EPgradient S	Eppendorf
Real-time (RT) PCR cycler	Mx3005P	Agilent Technologies

Type	Model	Supplier
Refractometer	Optilab T-rEX	Wyatt Technology
Robotic sealing unit for microplates	RoboSeal	HJ-BIOANALYTIC
Rocker	Duomax 1030	Heidolph Instruments
Rotor assembly (4x2,250 ml)	JS-5.0	Beckman Coulter
Rotor assembly (6x1,000ml)	JLA-8.100	Beckman Coulter
Rotor assembly (6x250 ml)	JA-16.250	Beckman Coulter
Rotor assembly (10x100 ml)	JA-18	Beckman Coulter
Rotor assembly (8x50 ml)	JA-25.50	Beckman Coulter
Shaker	VRX basic Vibrax®	Carl Roth
Spectrophotometer	BioPhotometer	Eppendorf
Spectrophotometer	NanoDrop ND 1000	peqlab
Thermomixer	Thermomixer comfort	Eppendorf
Ultra pure water system	TKA GenPure	Thermo Fisher Scientific
Ultrasonic bath sonicator	Sonorex RK 255 H	BANDELIN electronic
UV-illumination table	Electronic UV Transilluminator	Ultra Lum
UV imaging system	Gel Doc™ XR System	Bio-Rad Laboratories
Vortex mixer	Vortex-Genie 2	Scientific Industries
X-ray cryosystem	X-Stream™ 2000	Rigaku
X-ray detector	R-AXIS HTC	Rigaku
X-ray generator	MicroMax™-007 HF	Rigaku
X-ray optics	VariMax™	Rigaku

Chromatography columns and media

Name	Type	Supplier
Econo-Column® 2.5 x 20 cm	Column body, low pressure	Bio-Rad Laboratories
Econo-Column® 5 x 10 cm	Column body, low pressure	Bio-Rad Laboratories
HisTrap™ FF crude	Immobilized metal ion affinity chromatography FPLC column	GE Healthcare
HiLoad™ 16/60 Superdex™ 200	Preparative SEC FPLC column	GE Healthcare
HiLoad™ 26/60 Superdex™ 200	Preparative SEC FPLC column	GE Healthcare
Superdex™ Increase 200 10/300 GL	Analytical SEC FPLC column	GE Healthcare
Superdex™ 200 10/300 GL	Analytical SEC FPLC column	GE Healthcare
Superdex™ 75 10/300 GL	Analytical SEC FPLC column	GE Healthcare
Superose™ 6 10/300 GL	Analytical SEC FPLC column	GE Healthcare
MonoQ™ 10/100 GL	High-resolution AIEX FPLC column	GE Healthcare
Protino® Ni-IDA	Immobilized metal ion affinity chromatography resin	MACHEREY-NAGEL
Streptavidin Agarose	Affinity chromatography medium	Thermo Scientific

2.1.3 Software

Computer Programs

Name	Description/Usage	Supplier/*Reference*
3D-blast	Structure analysis; database search for structural homologues	*Yang and Tung (2006)*
Astra VI	Multi-angle light scattering (MALS) detector and refractometer control; collection and analysis of MALS data	Wyatt
ATSAS	Small-angle X-ray scattering (SAXS) data analysis; software suite	*Petoukhov et al. (2012)*
AxioVision	Recording of microscopy images	ZEISS
BLAST	Biosequence analysis; database search for homologous sequences	*Altschul et al. (1990)*
Buccaneer	X-ray crystallography; automated tracing of protein chains in electron-density maps	*Cowtan (2006)*
CCP4	X-ray crystallography; software suite for determination of macromolecular structures	*Winn et al. (2011)*
CCP4i	X-ray crystallography; graphical interface to *CCP4*	*Potterton et al. (2003)*
cNLS Mapper	Biosequence analysis; prediction of nuclear localization signals	*Kosugi et al. (2009)*
Clustal Omega	Biosequence analysis; multiple sequence alignment	*Sievers et al. (2011)*
Coot	X-ray crystallography; model building software	*Emsley et al. (2010)*
CRYSOL	SAXS data analysis; evaluation of solution scattering for macromolecule with known atomic structure	*Svergun et al. (1995)*
CrystalClear	X-ray data collection and basic processing	Rigaku
DAMMIF	SAXS data analysis; rapid *ab initio* shape determination	*Franke and Svergun (2009)*
ElNémo	Structure analysis; normal mode analysis (NMA)	*Suhre and Sanejouand (2004)*
GASBOR	SAXS data analysis; *ab initio* shape determination	*Svergun et al. (2001)*
GENtle	Analysis and *in silico* editing of DNA sequences, plasmid map database management	`gentle.magnusmanske.de`
GNOM	SAXS data analysis; indirect transform program for SAXS data processing	*Svergun (1992)*
HMMER	Biosequence analysis; sequence alignment and database search	*Finn et al. (2011)*
iMOSFLM	X-ray crystallography; processing of diffraction images, indexing and integrating, data reduction	*Leslie and Powell (2007)*
i-TASSER	Biosequence analysis; protein 3D structure prediction	*Zhang (2008)*
ITC$_{200}$	Isothermal titration calorimetry (ITC) instrument control; experimental design	MicroCal
iTOL	Visualization of phylogenetic trees	*Letunic and Bork (2011)*
MxPro	RT PCR cycler control; recording and basic processing of data	Agilent Technologies
meta-PPISP	Structure analysis; prediction of protein protein interaction sites	*Qin and Zhou (2007)*
Microsoft Excel	Spreadsheet software	Microsoft Cooperation
NES Finder 0.2	Biosequence analysis; prediction of nuclear export signals	`research.nki.nl/ fornerodlab/NES-Finder.htm`
NetNES 1.1	Biosequence analysis; prediction of nuclear export signals	*la Cour et al. (2004)*
ND-1000	Instrument control of spectrophotometer; recording and analysis of spectrophotometric data	peqlab
NLStradamus	Biosequence analysis; prediction of nuclear localization signals	*Nguyen Ba et al. (2009)*
OASIS	X-ray crystallography; SAD phasing software	*Wang et al. (2004)*

Name	Description/Usage	Supplier/*Reference*
Origin	Date analysis and graphing software	OriginLab Corporation
PHASER	X-ray crystallography; software for phasing	*McCoy et al. (2007)*
Phenix	X-ray crystallography; software suite for determination of macro-molecular structures	*Adams et al. (2002)*
Phyre2	Biosequence analysis; protein 3D structure prediction	*Kelley and Sternberg (2009)*
PISA	Structure analysis; interface and assembly analysis	*Krissinel and Henrick (2007)*
PredictProtein	Biosequence analysis; prediction of protein structural and functional features	*Yachdav et al. (2014)*
PRIMUS	SAXS data analysis; primary data reduction and processing of SAXS data	*Konarev et al. (2003)*
ProteomeLab	Analytical ultracentrifuge control; recording of analytical ultra-centrifugation (AUC) data	Beckmann Coulter
PyMOL	3-dimensional visualization and graphical illustration software	Schrödinger
Quantity One®	Ultraviolet (UV) imaging system control; recording and analysis of UV images	BioRad
REFMAC	X-ray crystallography; macromolecular structure refinement	*Vagin et al. (2004)*
Scala	X-ray crystallography; scaling and merging of diffraction data	*Evans (2006)*
Scorecons	Biosequence analysis; scoring of residue conservation based on multiple sequence alignment	*Valdar (2002)*
SEDFIT	Analysis of AUC data	*Schuck (2000)*
SERp Server	Biosequence analysis; identification of candidate mutations for surface entropy reduction	*Goldschmidt et al. (2007)*
SHARP	X-ray crystallography; heavy atom refinement and phasing program	*Bricogne et al. (2003)*
Skylign	Biosequence analysis; generation of hidden Markov model profiles	*Wheeler et al. (2014)*
Spectra Manager	CD spectropolarimeter control; recording and analysis of CD data	JASCO
Thermofluor Script	*Excel* script for thermofluor data analysis	SGC, Oxford
UNICORN	FPLC instrument control; recording, management and analysis of chromatograms	GE Healthcare
XDS	X-ray crystallography; processing of diffraction images, indexing and integrating, data reduction	*Kabsch (2010)*

Databases

Name	Contents	Web address	Reference
Human Protein Atlas	Expression profiles of human proteins	`www.proteinatlas.org`	Uhlén et al. (2015)
PDB	Structural data of biological macro-molecules	`www.rcsb.org`	Berman et al. (2000)
PFAM	Protein families, domains, repeats and motifs	`pfam.xfam.org`	Finn et al. (2014)
SMART	Protein domain profile-hidden Markov models	`smart.embl-heidelberg.de`	Schultz et al. (1998)
UniProt	Protein sequences	`www.uniprot.org`	Consortium (2011)

2.2 Method descriptions

2.2.1 Molecular biology

Separation of deoxyribonucleic acid (DNA) by agarose gel electrophoresis

TAE buffer:	40 mM Tris, 20 mM acetic acid, 1 mM EDTA pH 8.0
DNA sample buffer:	10 mM Tris, 0.3‰ bromphenol blue, 0.3‰ xylene cyanol, 60% glycerol, 60 mM EDTA, pH 7.6

DNA molecules were separated by size via electrophoresis in agarose gels. Agarose gels were prepared by dissolving 0.8 - 1.5% (w/v) NEEO ultra quality agarose in TAE buffer, adding either 0.4 µg/ml ethidium bromide or Midori green Advance DNA stain (dosed according to manufacturer's instructions) and were cast using the Mini-Sub® Cell GT system. DNA samples were mixed 6:1 with DNA sample buffer, and loaded next to a molecular-weight size marker to the gel submerged in TAE buffer in the electrophoresis chamber. Electrophoresis was performed at 100 V for up to 60 min. For subsequent detection, the gel was exposed to UV light on a UV transillumination table at low light intensity for excision of specific fragments or in the UV imaging system for documentation. Extraction of DNA fragments from agarose gels was performed according to the manufacturer's instructions using the NucleoSpin® Gel and PCR Clean-up kit.

Transformation of competent *E. coli* cells

Plasmids were introduced into chemically competent *E. coli* cells by heat shock transformation. Half of a ligation reaction or 50 - 100 ng plasmid DNA was added to one aliquot (50 - 100 µl) of thawed competent cells and incubated on ice for 20 - 30 min. Subsequently, cells were first exposed to 42 °C for 60 s, then chilled on ice. 1 ml of LB-medium was added and the cells were incubated at 37 °C for 45 - 60 min under constant shaking in a Thermomixer. Afterwards, the cells were plated on an LB agar plate with the appropriate antibiotics (see lists of plasmids on page 27 and strains on page 25) and incubated overnight at 37 °C.

Colony PCR

To probe for the presence of a specific DNA sequence in *E. coli* clones, single colonies were used in an analytical polymerase chain reaction (PCR). Cells of the colony were transferred to a PCR-tube and mixed with the PCR-reaction mix supplied with suitable primers (peqGOLD PCR Mastermix S, 2.5 mM $MgCl_2$, 1 µM forward primer, 1 µM

reverse primer). The PCR reaction was performed in a PCR-cycler following the PCR-protocol depicted below, with the duration of the elongation step adjusted to the expected fragment length. Subsequently, the PCR products were analyzed by agarose gel electrophoresis as described on the facing page.
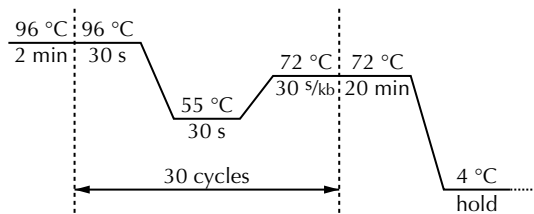
Amplification of plasmid DNA

To amplify plasmid DNA, *E. coli* cells were transformed with the respective plasmid as described on the preceding page. A single colony was transferred afterwards to 5 ml (small scale) or 100 - 200 ml (medium scale) of sterile LB medium containing the appropriate antibiotics (see lists of plasmids on page 27 and strains on page 25) and incubated overnight at $37\,^\circ$C under constant shaking at 200 rpm in an incubator shaker. Subsequently, cells were separated from the medium by centrifugation at 4,000 x g for 15 min at $4\,^\circ$C. The plasmid DNA was isolated following the manufacturer's instructions with the NucleoSpin® Plasmid or NucleoBond® Xtra Midi kit.

Molecular cloning

For molecular cloning, that is the assembly of recombinant DNA molecules, two different methods were applied: classical cloning (using restriction endonucleases) and sequence and ligation independent cloning (SLIC) (Li and Elledge, 2007). To generate fragments with suitable overhangs (containing the restriction endonuclease recognition site for classical cloning or homologous regions for recombination in SLIC) the fragments were amplified by PCR with primers that carried the desired overhang. PCR was performed using the high-fidelity DNA polymerase Phusion® (reaction mix: supplied HF buffer, 0.2 mM dATP, 0.2 mM dCTP, 0.2 mM dGTP, 0.2 mM dTTP, 0.5 μM forward primer, 0.5 μM reverse primer, 5 - 10 ng template DNA, 1 U Phusion® polymerase) in a PCR-cycler following the protocol outlined below with adjusted elongation time and annealing temperature. Afterwards, the PCR products were purified using the NucleoSpin® Gel and PCR Clean-up kit according to the manufacturer's instructions.

37
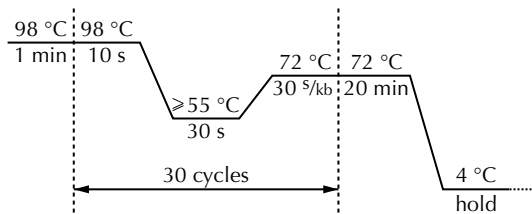
For classical cloning, the template DNA or PCR product and the vector plasmid DNA were digested with suitable restriction enzymes for 1 h at 37 °C (reaction mix was prepared according to the instructions of the supplier). 10 U CIP were added to vector digestion reactions 15 min prior to the end of incubation. The products were separated by agarose gel electrophoresis as described on page 36 and the desired fragment was extracted using the NucleoSpin® Gel and PCR Clean-up kit according to the manufacturer's instructions. For ligation the fragment and the linearized vector were mixed to a 3-fold molar excess of the insert, combined with Quick Ligation Buffer and 2,000 U Quick T4 DNA Ligase, and incubated at room temperature (22 °C) for 5 min. Ligation products were either directly used for transformation of *E. coli* DH5α™ cells or stored at -20 °C.

For SLIC, the DNA fragments were incubated with T4 DNA polymerase (reaction mix: NEBuffer 2, 0.2 $^{\mu g}/_{\mu l}$ BSA, 0.3 $^{units}/_{\mu l}$ T4 DNA polymerase) for 30 min at room temperature (22 °C) to generate single-stranded overhangs. The reaction was stopped by addition of dCTP to a final concentration of 1 mM. Annealing reactions were set up by combining the fragments in a suitable molar ratio and mixing with T4 DNA ligase reaction buffer and 1 $^{ng}/_{\mu l}$, incubated for 30 min at room temperature (22 °C) and subsequently either directly used for transformation of *E. coli* DH5α™ cells or stored at -20 °C.

Clones were tested either by colony PCR and then amplified in small scale as described on the previous page, or controlled afterwards by analyzing the fragment pattern in agarose gel electrophoresis upon small-scale digestion of 200 ng DNA with a suitable restriction enzyme performed following the instructions of the supplier. Finally, all recombinant plasmids were verified by dideoxy sequencing executed by the companies Seqlab (Göttingen, Germany) or Eurofins Genomics (Ebersberg, Germany).

Mutagenesis

To introduce specific mutations, substitutions or deletions, DNA fragments with the respective variation were generated by PCR with suitable primers. For some simple mutations, a single PCR step with homologous primers carrying the variation was used to amplify the entire plasmid. The resulting linearized plasmid was controlled and purified by agarose gel electrophoresis with subsequent gel extraction, then further treated

following the standard SLIC procedure described on page 37. Alternatively, for the introduction of more complex variations, overlap extension PCR (Higuchi et al., 1988; Ho et al., 1989) was applied. In initial steps, the primers used for subcloning were employed as flanking primers and the primers designed for the respective mutation to generate the overlap (following the PCR protocol described on page 37 in the section Molecular cloning). The PCR products were controlled and purified by agarose gel electrophoresis with subsequent gel extraction. In the next PCR steps, the fragments were combined, pairing via their appended overlaps, and amplified using the flanking primers (reaction mix: supplied HF buffer, 0.2 mM dATP, 0.2 mM dCTP, 0.2 mM dGTP, 0.2 mM dTTP, 0.5 μM forward primer, 0.5 μM reverse primer, 25 ng fragment 1, 25 ng fragment 2, 1 U Phusion® polymerase; protocol as described above). The products were controlled and purified by agarose gel electrophoresis with subsequent gel extraction. The final product was then subcloned via either of the methods described on page 37.

Generation of pETM-SUMO

The expression vector termed pETM-SUMO was generated by amplifying the pET vector backbone using the pETM-11 vector as template and the coding sequence of the 6xHis-tagged SUMO protein of *S. cerevisiae* using the pETHSUL vector as template in a PCR with primers appending homologous regions to the fragments and assembling the plasmid via SLIC. Clones were amplified and selected based on their BspHI digestion pattern.

Cloning strategies

The expression constructs generated in the course of this work are listed in the appendix (section 6.4 on page 155) and are specified in the following by their identifier (ID) in this list.

All plasmids for expression of wild type muskelin and its shortened variants in *E. coli* (M1-M29, M52-M59, M83, M84, M100, M103-M110) were generated by subcloning the muskelin variant coding sequence as NcoI/NotI fragment, generated in a PCR using the Muskelin-pGEX or M8 plasmid as template, into pETM-11, pETM-30, pETM-41 or pETM-SUMO. Similarly, to generate the plasmids for the expression of the shortened GABA$_A$ receptor intracellular domains in *E. coli*, fragments amplified by PCR using the respective GABA$_A$ receptor intracellular domain as template (G1-G4) were subcloned as NcoI/NotI fragment into pETM-11.

The *E. coli* expression plasmids of muskelin variants carrying single mutations in the head-to-tail interface (M60-M66, M52-M59, M92, M94, M96, M114) or the LisH interface (M85-91, M98, M115, M116), the double phosphomimetic mutation in the Kelch re-

39

peat domain (M46-M51), the variant with the fourfold cysteine to serine mutation in the discoidin domain (M113), the muskelin full-length K182A H183A variant (M99) and all alanine-substitution variants (M67-M82) were built by generating a fragment containing the variation by overlap extension PCR using M8 as initial template and subsequently subcloning it as NcoI/NotI fragment into pETM-SUMO.

The *E. coli* expression plasmids of the shortened variants carrying the K182A/H183A mutation (M101, M102) were built by subcloning a NcoI/NotI fragment generated by PCR using M99 as template into pETM-SUMO.

The *E. coli* expression plasmids of full-length muskelin harboring the triple mutation (M93, M95, M97) were built by using the F184E/L196Q variant M42 as initial template in overlap extension PCR to introduce the third mutation, and subcloning the PCR product as NcoI/NotI fragment into pETM-SUMO.

The expression plasmids of muskelin variants with the double mutation of F184 and L196 (M30-M45, M123-M124) were built by SLIC using M2, M4, M8, M117 or M118 as template in a PCR step with primers carrying the respective mutation.

All further plasmids for expression of muskelin variants in eukaryotic cells were built by transferring the mutation from a previously generated plasmid. The plasmid of the ΔC35 variant (M130) was built by transposing the BspEI/XhoI fragment of M100 to the BspEI/SalI linearized backbone of M119. To build the plasmids of the N144R variant (120, M121), a SacI/BspEI fragment was transposed from M96 to M118 and subsequently an EcoRI/BspEI fragment was transposed from M121 to M117. All plasmids of mCherry-tagged muskelin mutated variants (M122, M125, M128, M129) were built by subcloning a EcoRI/SalI fragment generated by PCR using M42, M96, M97 or M99 as template into M119.

Protein expression

For the recombinant expression of proteins, *E. coli* cells were transformed with the respective expression plasmid as described on page 36 and used to inoculate 5 ml (for small scale expression) or 100 - 200 ml (for large scale expression) of sterile LB medium containing the appropriate antibiotics (see lists of plasmids on page 27 and strains on page 25). Cultures were incubated under constant shaking 37 °C in a shaker incubator overnight. Expression cultures were started by transferring pre-cultured cells to fresh LB-medium containing the appropriate antibiotics, either 5 ml in a sterile culture tube for small scale test expression or 2 l in a sterile 5 l Erlenmeyer flask for large scale expression. Cultures were grown under constant shaking at 37 °C while the optical density at 600 nm (OD_{600}) was monitored. When the desired OD_{600} was reached (0.5 - 0.8 for standard expressions of muskelin), expression was induced by addition of IPTG to a final concentration of 0.5 mM, and the cultures were incubated

at the expression temperature (20 - 22 °C for standard expressions of muskelin) under constant shaking in a shaker incubator. At the end of the expression phase (19 - 21 h for standard expressions of muskelin), the cells were harvested by centrifugation for 15 min at 4,000 x g at 4 °C and the cell pellets were stored until further use at -80 °C.

To test the influence of different expression conditions and durations, the OD_{600} of the correspondingly treated culture was determined and 500 µl samples were taken before induction of expression and after expression. The cells were pelleted by centrifugation for 2 min at 11,000 x g at 4 °C and resuspended in phosphate buffered saline (8.1 mM Na_2HPO_4, 1.6 mM KH_2PO_4, 150 mM NaCl, 2.7 mM KCl) supplemented with 0.5 mM PMSF and 1 $^U/_{ml}$ DNaseI , whereby the volume was scaled by the measured OD_{600} of the sample. To separate the soluble and insoluble protein fraction, cells were first lyzed mechanically by adding glass beads and intensely shaking the suspensions for 15 min at 4 °C. Glass beads were sedimented by centrifugation for 1 min at 200 x g and 4 °C, the supernatant was transferred and soluble and insoluble fractions were separated by centrifugation for 15 min at 20,000 x g and 4 °C. All samples were analyzed by separating the proteins by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and staining with Coomassie as described on page 46.

2.2.2 Protein purification

Purification of SUMO-protease

LEW buffer dtUD1:	50 mM Na_2HPO_4, 250 mM NaCl, 10% (w/v) glycerol, 2 mM DTT, pH 8.0
Elution buffer dtUD1:	50 mM Na_2HPO_4, 250 mM NaCl, 300 mM imidazole, 10% (w/v) glycerol, 2 mM DTT, pH 8.0
Dialysis buffer dtUD1:	20 mM Tris, 150 mM NaCl, 1 mM EDTA, 20% (w/v) glycerol, 5 mM DTT, pH 8.0
Storage buffer dtUD1:	20 mM Tris, 150 mM NaCl, 1 mM EDTA, 50% (w/v) glycerol, 5 mM DTT, pH 8.0

The isolation of dtUD1 from *E. coli* cells basically followed the protocol described by Weeks et al. (2007). Cells were resuspended in 10 $^{ml}/_{(g\ cell\ pellet)}$ LEW buffer supplemented with 0.5 mM PMSF and 1 $^U/_{ml}$ DNase and were lysed in two passages through a cell disruption system at a pressure setting of 1.5 kbar. The resulting cell lysate was centrifuged for 45 min to 60 min at 40,000 x g and 4 °C and the supernatant was loaded by gravity onto a Protino® Ni-IDA resin column, which was pre-equilibrated in LEW buffer. The column was washed with at least 10 column volumes (CVs) of LEW buffer, before proteins were eluted in elution buffer. The protein concentration in the eluate fraction was determined spectrophotometrically (as described on page 45), samples of all purification steps were analyzed by SDS-PAGE (as described

on page 46). The eluate fractions containing the target protein were pooled and dialyzed against dialysis buffer overnight at 4 °C. After dialysis, the protein solution was cleared by centrifugation for 30 min at 70,000 x g and 10 °C, the supernatant was diluted to a protein concentration of 1 $^{mg}/_l$ and divided into 500 µl aliquots, which were flash-frozen in liquid nitrogen and stored at -80 °C until further use.

Affinity purification of the GABA$_A$R α1, GABA$_A$R α3 and GABA$_A$R α5 intracellular loops

Lysis buffer N: 20 mM Tris, 500 mM NaCl, pH 8.0
Washing buffer N: 20 mM Tris, 500 mM NaCl, 20 mM imidazole, pH 8.0
Elution buffer N: 20 mM Tris, 500 mM NaCl, 500 mM imidazole, pH 8.0

The initial purification steps for the intracellular loops of the GABA$_A$R α1, GABA$_A$R α3 and GABA$_A$R α5 subunits were performed under native conditions. Cells were resuspended in 10 $^{ml}/_{(g\ cell\ pellet)}$ lysis buffer N supplemented with protease inhibitor cocktail and 1 $^U/_{ml}$ DNase and were lysed in two passages through a cell disruption system at a pressure setting of 1.5 kbar. The resulting cell lysate was centrifuged for 45 min to 60 min at 40,000 x g and 4 °C. The supernatant was subjected to affinity chromatography, which was performed on an FPLC system with an attached HisTrap™ FF crude column. The sample was loaded completely onto the column, which was pre-equilibrated in lysis buffer N, then the column was washed with 20 CV washing buffer N. Proteins were eluted with a linear gradient from 0% to 100% elution buffer N. Samples of all purification steps were analyzed by SDS-PAGE (as described on page 46), and the eluate fractions that contained the target protein were pooled.

Affinity purification of GABA$_A$R α2 intracellular loop

Lysis buffer N: 20 mM Tris, 500 mM NaCl, pH 8.0
Lysis buffer DN: 20 mM Tris, 500 mM NaCl, 1.5 M guanidinium chloride, pH 8.0
Washing buffer DN: 20 mM Tris, 500 mM NaCl, 20 mM imidazole, 1.5 M guanidinium chloride, pH 8.0
Elution buffer DN: 20 mM Tris, 500 mM NaCl, 500 mM imidazole, 1.5 M guanidinium chloride, pH 8.0

The initial purification steps for the intracellular loop of the GABA$_A$R α2 subunit was performed under denaturing conditions. Cells were resuspended in 10 $^{ml}/_{(g\ cell\ pellet)}$ lysis buffer N supplemented with protease inhibitor cocktail and 1 $^U/_{ml}$ DNase and were lysed in two passages through a cell disruption system at a pressure setting of

1.8 kbar. The resulting cell lysate was supplemented with guanidinium chloride (final concentration: 1.5 M), stirred for 30 min at 4 °C and then centrifuged for 45 min to 60 min at 40,000 x g and 4 °C. The supernatant was subjected to affinity chromatography, which was performed on an FPLC system with an attached HisTrap™ FF crude column. The sample was loaded completely onto the column, which was pre-equilibrated in lysis buffer DN, then the column was washed with 20 CV washing buffer DN. Proteins were eluted with a linear gradient from 0% to 100% elution buffer DN. Samples of all purification steps were analyzed by SDS-PAGE (as described on page 46), and the eluate fractions containing the target protein were pooled.

Affinity purification of muskelin constructs

For the affinity purification of muskelin constructs, different strategies were followed depending on the respective construct and intended use:

Variant I: for shorter constructs - M1, M2, M3, M7 and all derived variants thereof
Variant II: for simple purification of longer constructs - M4, M5, M8, M54, M55, and all derived variants thereof
Variant II: for purification to higher purity of all constructs

Variant I:

LEW buffer Mkln I: 20 mM PIPES, 0.5 M NaCl, 5% (w/v) glycerol, 2 mM DTT, pH 7.5
Elution buffer Mkln I: 20 mM PIPES, 0.2 M NaCl, 5% (w/v) glycerol, 2 mM DTT, pH 7.5

Cells were resuspended in $10\,^{ml}/_{(g\ cell\ pellet)}$ LEW buffer I supplemented with protease inhibitor cocktail and $1\,^{U}/_{ml}$ DNase and were lysed in two passages through a cell disruption system at a pressure setting of 1.5 kbar. The resulting cell lysate was centrifuged for 45 min to 60 min at 40,000 x g and 4 °C and the supernatant was loaded by gravity onto a Protino® Ni-IDA resin column, which was pre-equilibrated in LEW buffer I. The column was washed with at least 10 CV of LEW buffer I, then dtUD1 protease in 2 CV of elution buffer I was added and the on-column cleavage reaction was allowed to proceed overnight at 4 °C for 10 h. Cleaved protein was collected and the remaining protein were washed off with 2 CV of elution buffer I. Samples of all purification steps were analyzed by SDS-PAGE (as described on page 46). The eluate fractions containing the target protein were combined and cleared by centrifugation for 30 min at 70,000 x g and 10 °C.

Variant II:

LEW buffer Mkln II:	20 mM PIPES, 0.5 M NaCl, 5% (w/v) glycerol, 5 mM DTT, pH 7.5
Elution buffer Mkln II:	20 mM PIPES, 0.2 M NaCl, 0.5 M imidazole, 5% (w/v) glycerol, 5 mM DTT, pH 7.5

Cells were resuspended in $10 \, \mathrm{ml/(g \, cell \, pellet)}$ LEW buffer II supplemented with protease inhibitor cocktail and $1 \, \mathrm{U/ml}$ DNase and were lysed in two passages through a cell disruption system at a pressure setting of 1.5 kbar. The resulting cell lysate was centrifuged for 45 min to 60 min at 40,000 x g and 4 °C and the supernatant was loaded by gravity onto a Protino® Ni-IDA resin column, which was pre-equilibrated in LEW buffer II. The column was washed with at least 10 CV of LEW buffer II, then the proteins were eluted in elution buffer II. The protein concentration in the eluate fraction was determined spectrophotometrically (as described on the next page) and samples of all purification steps were analyzed by SDS-PAGE (as described on page 46). The eluate fractions containing the target protein were pooled and the dtUD1 protease was added for overnight cleavage at 4 °C for at least 10 h.

Variant III:

LEW buffer Mkln III:	20 mM Tris, 0.5 M NaCl, 5% (w/v) glycerol, 5 mM DTT, pH 8.0
Elution buffer Mkln III:	20 mM Tris, 0.2 M NaCl, 0.5 M imidazole, 5% (w/v) glycerol, 5 mM DTT, pH 8.0
AIEX buffer A:	20 mM Tris, 50 mM NaCl, 5 mM DTT, pH 8.0
AIEX buffer B:	20 mM Tris, 1 M NaCl, 5 mM DTT, pH 8.0

Cells were resuspended in $10 \, \mathrm{ml/(g \, cell \, pellet)}$ LEW buffer III supplemented with protease inhibitor cocktail and $1 \, \mathrm{U/ml}$ DNase and were lysed in two passages through a cell disruption system at a pressure setting of 1.5 kbar. The resulting cell lysate was centrifuged for 45 min to 60 min at 40,000 x g and 4 °C and the supernatant was loaded by gravity onto a Protino® Ni-IDA resin column, which was pre-equilibrated in LEW buffer III. The column was washed with at least 10 CV of LEW buffer III, then the proteins were eluted in elution buffer III. The protein concentration in the eluate fraction was determined spectrophotometrically (as described on the next page). Protein-containing fractions were pooled and dtUD1 protease was added. The cleavage proceeded overnight during dialysis against AIEX buffer A. The dialysate was cleared by centrifugation for 30 min at 70,000 x g and 10 °C, then subjected to AIEX chromatography using an FPLC system with an attached MonoQ™ 10/100 GL column. The sample was loaded onto the column pre-equilibrated in AIEX buffer A and washed with at

least 10 CV AIEX buffer A. Bound proteins were eluted with a linear gradient from 0% to 40% AIEX buffer B over 20 CV followed by a step with 100% AIEX buffer B for 5 CV. Samples of all purification steps were analyzed by SDS-PAGE (as described on the following page) and fractions containing the target protein were pooled.

Preparative size exclusion chromatography

SEC buffer I: 20 mM PIPES, 200 mM NaCl, 5% (w/v) glycerol, 5 mM DTT or 1 mM TCEP, pH 7.5
SEC buffer II: 20 mM PIPES, 500 mM NaCl, 1 mM TCEP, pH 7.5

As final purification step, SEC was used for all proteins. SEC buffer I was used for standard purifications of all proteins, while SEC buffer II was exclusively used for the final SEC step prior to crystallization of $Mkln_{1-735}$ N144R and $Mkln_{12-735}$ N144R. The protein samples obtained in the earlier steps of purification were concentrated in a suitable spin concentrator to a volume below 5% of the column volume of the respective SEC column. The concentrated samples were centrifuged for 20 min at 25,000 x g and 4 °C, applied to the SEC column equilibrated in SEC buffer using an FPLC system, and separated over 1.2 CV. Samples of the fractions were analyzed by SDS-PAGE (as described on the next page) and fractions containing the pure target protein were pooled, concentrated in a spin concentrator, and either directly used for analysis or flash frozen in small aliquots and stored at -80 °C until further use.

2.2.3 Biochemical and biophysical analyses

UV/Vis spectrophotometry

The concentrations of purified DNA and proteins were determined by measuring the UV absorbance using a spectrophotometer. Absorbance spectra in the wavelength range of 220 - 350 nm were recorded and corrected for the reference spectrum of the buffer solution devoid of DNA or protein. For DNA, the concentration was calculated based on the absorbance at 260 nm and the ratio of the absorbance at 260 nm and 280 nm was used to control for the purity of the DNA. For proteins, the concentration was calculated based on the absorbance at 280 nm and the specific extinction coefficient of the respective protein, which was calculated based on the amino acid sequence using *ProtParam*.

Separation of proteins by SDS-polyacrylamide gel electrophoresis

Separating gel:	10% - 18% acrylamide/bisacrylamide mix, 250 mM Tris pH 8.8, 0.1% SDS, freshly added: 0.1% APS and 0.4‰ TEMED
Stacking gel:	5% acrylamide/bisacrylamide mix, 125 mM Tris pH 6.8, 0.1% SDS, freshly added 0.1% APS and 1‰ TEMED
SDS-PAGE sample buffer:	250 mM Tris pH 6.8, 0.5 M DTT, 0.5% (w/v) bromphenol blue, 10% (w/v) SDS, 50% glycerol
SDS-PAGE running buffer:	192 mM glycine, 25 mM Tris, 0.1% SDS
R-250 staining solution:	1‰ - 5‰ Coomassie R-250, 50% methanol, 10% acetic acid
Destaining solution:	10% ethanol, 5% acetic acid
G-250 staining solution:	0.08‰ Coomassie G-250, 36 mM HCl

Proteins were denatured and separated by their molecular weight in SDS-PAGE following the protocol described by Laemmli (1970). Gels were prepared by sequentially casting the separating and stacking gel using the Mini Protean II system. Samples were mixed 5:1 with sample buffer and incubated for 5 min at 95 °C. Protein samples and a molecular-weight size marker were loaded to the gel installed in the electrophoresis chamber filled with running buffer. Electrophoresis was performed at 100 - 200 V until the dye had nearly traversed the gel.

For staining, gels were either first stained in R-250 staining solution for 5 - 15 min and afterwards destained in destaining solution, or washed three times with hot water, then stained for 5 min in G-250 staining solution and afterwards destained in water.

Separation of proteins by native agarose gel electrophoresis

NAGE sample buffer:	60% glycerol, 0.1‰ Orange G
NAGE running buffer:	25 mM Tris, 200 mM glycine
R-250 staining solution:	1‰ - 5‰ Coomassie R-250, 50% methanol, 10% acetic acid
Destaining solution:	10% ethanol, 5% acetic acid

Native proteins were separated via electrophoresis in agarose gels. Agarose gels were prepared by dissolving 0.75 - 1% (w/v) HEEO ultra quality agarose in NAGE running buffer and were cast using the Vari-Gel™ horizontal maxi gel system. Samples were combined in the intended concentrations (typically between 20 μM and 150 μM) and incubated for 30 - 60 min on ice. After incubation, the samples were mixed 6:1 with NAGE sample buffer and loaded onto a gel submerged in NAGE running buffer in the electrophoresis chamber. Electrophoresis was performed at 100 V at 4 °C until the dye had nearly traversed the gel. For staining, gels were first stained in R-250 staining solution for 10 - 20 min and afterwards destained overnight in destaining solution.

Determination of thermal stability using ThermoFluor

To assess the thermal stability of proteins in a ThermoFluor assay, first an array of buffer conditions was placed in the wells of a 96-well thin wall PCR plate, with at least triplicates for a reference buffer and water control. The protein sample was mixed with Sypro Orange and added to the wells containing buffer, while Sypro Orange at the same concentration alone was added to the control. The plates were sealed with optical quality sealing foil and heated in a real time PCR cycler from 25 °C to 95 °C in 1 °C steps of 1 min duration while recording the fluorescence (excitation at 490 nm, emission at 575 nm) in the wells.

Circular dichroism spectroscopy

CD buffer: 10 mM NaH_2PO_4, 50 mM Na_2SO_4, 1 mM TCEP, pH 7.5

For CD spectroscopy measurements, the protein samples were first dialyzed against CD buffer at 4 °C overnight. Dialyzed samples were centrifuged for 10 min at 25,000 x g and 4 °C and diluted in dialysis buffer to obtain an absorbance at 280 nm (A_{280}) of 0.15 AU. CD spectra were recorded for each protein and the buffer employing a spectropolarimeter at wavelengths from 260 - 195 nm at room temperature. Ten spectra were accumulated to optimize the signal to noise ratio. For analysis, the buffer spectrum was subtracted as reference from the protein spectra.

Isothermal titration calorimetry

ITC buffer: 20 mM PIPES, 200 mM NaCl, 5% (w/v) glycerol, 1 mM TCEP, pH 7.5

All proteins to be used in ITC analyses were first dialyzed overnight against ITC buffer at 4 °C. After dialysis, the samples were centrifuged for 20 min at 25,000 x g and 4 °C, and the dialysis buffer was filtered and degassed to serve as reference. The concentrations of the protein samples were determined as described above, then samples were diluted to the intended concentrations (see below) in filtered dialysis buffer and optionally degassed. Peptides to be used in the analysis were dissolved and diluted to the intended concentrations (see below) in the dialysis buffer. The isothermal titration experiments were performed at 20 °C in either the VP-ITC or ITC_{200} instrument. For each ligand, a ligand to buffer titration was carried out and subtracted from the measured data prior to curve fitting. For dissociation ITC, a buffer to buffer titration was used as reference and subtracted prior to curve fitting. All data were analyzed using the *Origin* software, assuming a one site-binding model or a simple dimer dissociation model for curve fitting.

To set protein concentrations in ITC experiments, pilot experiments giving an estimate of the expected dissociation constant (K_D) were conducted and the concentrations were adjusted to achieve a compromise between protein solubility and optimal data quality. Generally, the aim was to reach a value of the Wiseman parameter or c-value (defined as $c = K_a M_t$ with K_a being the association constant of the measured binding and M_t the concentration of the receptor, that is the macromolecule in the cell Wiseman et al., 1989) above 10 and an at least 2.5-fold molar excess of the ligand. For example, the measurements of the head-to-tail interaction (K_D of wild type: 5.5 μM) were performed using a receptor (Mkln$_{205-735}$) concentration of 140 μM, thus reaching a c-value above 25. If the necessary protein concentrations could not be reached, experiments were still performed and parameters were determined for measurements with c-values above 1. The analysis of the experiments at low c-values is likely to yield a good estimate of the K_D, while the results for enthalpy (and thus entropy) are not reliable (Turnbull and Daranas, 2003).

Analytical size exclusion chromatography

aSEC buffer: 20 mM PIPES, 200 mM NaCl, 5% (w/v) glycerol, 1 mM TCEP, pH 7.5

Prior to analysis in analytical SEC, samples were prepared at suitable concentrations (if applicable, approximately 10 times the estimated K_D) and combinations and incubated for 1 h on ice. Then, the samples were centrifuged for 20 min at 25,000 x g and 4 °C, applied to the SEC column, which was equilibrated in SEC buffer, using an FPLC system, and separated over 1.2 CV. Samples of the fractions were analyzed by SDS-PAGE (as described on page 46).

Chemical crosslinking

Cross-linking buffer: 20 mM PIPES, 200 mM NaCl, 5% (w/v) glycerol, 1 mM TCEP, pH 7.5

Each protein was prepared as duplicate in a suitable concentration to reach 20 μM upon addition of the crosslinker in cross-linking buffer and samples were incubated for 30 - 60 min on ice. BS³ was prepared freshly at a concentration of 10 mM in water. The crosslinker was added at a 20-fold molar excess to the preincubated protein samples, while simultaneously the same volume of buffer was added to the control. After 15 min incubation on ice, the reaction was quenched by addition of Tris pH 8.0 to a final concentration of 50 mM and directly mixed with SDS-PAGE sample buffer and analyzed by SDS-PAGE (as described on page 46).

Pull-down using biotinylated peptides

PBSX: 8.1 mM Na_2HPO_4, 1.6 mM KH_2PO_4, 150 mM NaCl, 2.7 mM KCl, 0.1% Triton® X-100

All peptides used within this work were N-terminally coupled to a biotin group. The peptides were resuspended, diluted in PBSX to a final concentration of 12.5 μM and incubated with streptavidin-coated agarose beads for 3 h at 4 °C under constant shaking on a rocker. Beads were washed three times with PBSX, then $Mkln_{1-156}$ (M1) diluted to a final concentration of 50 μM in PBSX was added and incubated with the beads for 3 h at 4 °C under constant shaking on a rocker. A sample of the unbound protein was taken, and the beads were washed three times with PBSX. After the last washing step, the beads were boiled in SDS-PAGE sample buffer, and samples of the loaded protein, the unbound fraction and the beads with the bound proteins were analyzed by SDS-PAGE (as described on page 46).

Molecular weight determination using multi-angle light scattering

MALS buffer: 20 mM PIPES, 200 mM NaCl, 1 mM TCEP, 5% glycerol, pH 7.5

MALS analyses were performed in conjunction with SEC using a combination of an ÄKTApurifier 10 with a suitable SEC column coupled in-line to a MALS detector and refractive index monitor. Protein samples were diluted to the desired concentration in MALS buffer, and centrifuged for at least 15 min at 25,000 x g and 4 °C. 100 μl of the sample were injected onto the column pre-equilibrated in freshly prepared MALS buffer and separated over 1.2 CV. Light scattering data were recorded in-line, processed and analyzed with the *AstraVI* software using a differential index of refraction value of 0.185 ml/g.

Molecular weight determination using analytical ultracentrifugation

AUC buffer: 20 mM PIPES, 200 mM NaCl, pH 7.5

Prior to AUC analysis, the proteins were diluted to reach an A_{280} of approximately 4 AU and dialyzed against AUC buffer at room temperature (22 °C) for at least 6 h. For analyses under reducing conditions, the AUC buffer was supplemented with 1 mM TCEP. For analyses under oxidizing conditions, the samples were first dialyzed against AUC buffer supplemented with 0.5 mM H_2O_2 for at least 4 h at room temperature (22 °C), then transferred and further dialyzed for at least 4 h against AUC buffer without additives. After dialysis, the concentration of the dialyzed proteins was adjusted to meet an A_{280} of 0.4 AU and AUC cells were loaded with 400 μl of protein and dialysis buffer

as reference. Sedimentation velocity runs were performed in a Beckman Optima XL-I analytical ultracentrifuge with an eight-hole An-50 Ti rotor at 20 °C and 40,000 rpm. Data were collected in continuous mode at a step size of 30 μm using absorption optical detection at a wavelength of 280 nm. Data were analyzed using *SEDFIT* to determine continuous distributions for solutions to the Lamm equation c(s) as described previously (Schuck, 2000). Analysis was performed with regularization at confidence levels of 0.95 and floating frictional ratio, time-independent noise, baseline, and meniscus position.

Small-angle X-ray scattering analyses

Samples for SAXS were prepared by a new SEC (as described on page 45, buffer contained 5 mM DTT) of which the purest fractions were pooled, concentrated and kept on ice until further analysis. Immediately before the SAXS experiments, several dilutions were prepared for each protein and the diluted samples were centrifuged at least 15 min at 20,000 x g and 4 °C. Synchrotron X-ray scattering data were collected at 5 °C at the European synchroton radiation facility (ESRF) beamline ID14-3, in Grenoble, France, using a PILATUS 1M pixel detector at a sample-detector distance of 2.43 m and a wavelength of 0.931 Å, in ten individual frames with 10 s exposure time for each sample.

The data were normalized to the intensity of the transmitted beam, and the scattering of the buffer was subtracted. The difference curves were scaled for concentration, and the data were extrapolated to infinite dilution using *PRIMUS*. The forward scattering $I(0)$, the radius of gyration R_g, the hydrated particle volume (calculated using the Porod invariant) and the maximum particle size D_{max} (determined from the pair distribution function) were computed by *GNOM*. The molecular mass was estimated by dividing the Porod volume by the scaling factor, which was assumed to be 1.6, unless otherwise stated. *Ab initio* models of Mkln$_{205-725}$ were created with *GASBOR* using data in the range of $0.0074 < s < 0.61$ Å^{-1}.

To explore the possibility to derive a model from the SAXS data of Mkln$_{1-205}$, a theoretical scattering curve was generated with *CRYSOL* based on a model encompassing the Mkln$_{12-205}$ crystal structure and an additional helix modeled to account for the invisible parts. Statistical noise was added to the generated scattering curve, before it used for *ab initio* modeling with *DAMMIF*. The derived models significantly differed from the original, hence it was concluded that modeling of experimental data would not yield a reliable result.

2.2.4 X-ray crystallography

Crystallization

Prior to crystallization, all proteins were again subjected to SEC (as described on page 45) and only the fractions containing the purest protein were used, concentrated or diluted to the intended concentration and centrifuged for at least 20 min at 25,000 x g and 4 °C. Initial screens for crystallization conditions were set up using the liquid handling robot HoneyBee 963 as sitting drop vapor diffusion experiments with the drops consisting of 0.3 μl protein solution mixed with 0.3 μl mother liquor and a reservoir of 40 μl mother liquor in a 96-well crystallization plate and sealed with adhesive sealing film. For optimization screens in 96-well format (detergent and additive screen), either a premix of the additives and the basis condition was prepared in a 96-well deep well block and crystallization plates were set up as described or, by hand, the additives were put to the reservoir of a crystallization plate first, then mother liquor was added, mixed, and 0.3 μl of the respective reservoir solution were transferred to the drop shelf, before protein solution was dispensed using the liquid handling robot. Follow-up grid screens were set up as hanging drop vapor diffusion experiments by hand using 24-well plates with 1 ml reservoir volume and 1 - 3 μl drops of differing composition on a cover slide sealed with silicon grease. Micro-seeding was performed by collecting crystals obtained in an earlier crystallization experiment, transferring them to fresh mother liquor and crushing the crystals by sonication in ultrasonic bath sonicator for 2 min and vortexing at high intensity for 1 min. Several dilutions (1:100 - 1:10000) of crystal seeds in mother liquor were prepared and added to the drop of the crystallization experiment.

Crystals for $Mkln_{1-205}$ and $Mkln_{12-205}$ were obtained at 20 °C by hanging drop vapor diffusion against a reservoir solution containing 100 mM BisTris (pH 5.25 - 6.0), 200 mM NaCl, and 20 - 35% PEG 3350 at a protein concentration of 15 - 20 mg/ml.

Data collection and processing

For cryoprotection during data collection, crystals were first transferred using a mounted cryoloop into their respective mother liquor containing 15% ethylene glycol, then into mother liquor containing 30% ethylene glycol, and finally flash frozen in liquid nitrogen. Single diffraction images were collected for an initial estimation of crystal quality either in house (using a MicroMax-HF 007 generator at a wavelength of 1.54 Å as X-ray source and an R-AXIS HTC imaging plate detector) or at the synchrotron facilities ESRF (Grenoble, France) or Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung (BESSY, Berlin, Germany). Parameters for data collection were adjusted to the purpose of collection and to the estimations of crystal mosaicity, space

group and unit cell dimension based on initial images. The datasets described in this work were recorded using the following facilities and parameters:

	Mkln$_{1-205}$		Mkln$_{12-205}$	
	SAD dataset	Native dataset	2.1 Å dataset	1.8 Å dataset
Beamline	BESSY BL14.1	BESSY BL14.1	BESSY BL14.1	ESRF BM14
Detector	Rayonix MX225	Rayonix MX225	Rayonix MX225	MAR 225 CCD
Wavelength	1.9	0.91841	0.9184	0.95373
Detector distance [mm]	97.01	170.69	224.74	187.39
Number of images	1150	100	100	110
Φ angle	0.5	0.5	1	1
Exposure time	10 s	10 s	5 s	7 s

Diffraction data were integrated and scaled using *iMOSFLM* or *XDS* and *Scala* (*CCP4i suite*), the relevant parameters and statistics are summarized in the results section on page 71.

Structure solution and model refinement

The structure of Mkln$_{1-205}$ was solved via phasing with sulphur single-wavelength anomalous diffraction (SAD) using the data recorded at 1.9 Å. The anomalous differences were calculated using *SHELXC* and heavy atom substructure searches were performed in parallel using *SHELXD* and the *HySS* submodule of the *Phenix* package, with varied settings for the maximum resolution (2.5-3.3 Å) and the number of heavy atom sites (7 to 11). The substructures were used to calculate initial phases using *Phaser*, and the phases were refined using *SHARP* in several cycles, until the quality of the phases allowed for automated building of partial models using the *Buccaneer* software. The partial model was completed manually and further refined in repeated maximum likelihood refinement cycles including translation, libration, screw-motion (TLS) refinement using the model building software *Coot* and refinement program *REFMAC*, resulting in a final model that encompassed residues 8 to 187 of the protein. Despite extensive rebuilding efforts, the structure could not be refined to an R$_{free}$ below 34%.

The structure of Mkln$_{12-205}$ was solved by molecular replacement with *Phaser* using the Mkln$_{1-205}$ model as search probe and improved with the model building software *Coot* with intermediary maximum likelihood refinement cycles including TLS refinement in *Phenix*.

RESULTS AND DISCUSSION

The main objective of the project presented in this thesis was to characterize muskelin as a GABA$_A$ receptor binding protein. A prerequisite for sound biochemical and biophysical analyses and crystallization experiments is the availability of the isolated, pure and intact protein. Thus, the first task was to find suitable conditions to recombinantly express and purify muskelin, the full-length protein as well as shortened variants thereof.

3.1 Construct design, expression and purification

At the beginning of this work, a strategy for the expression and purification had only been described for the full-length protein (Kiedzierska et al., 2008). This strategy was considered, partially adopted and optimized in the course of this work as described later in this section. Although shortened variants of muskelin were used in several earlier studies[1], none of them was designed for crystallization and, apart from the study by Ledee et al. (2005), no purifications had been described. For the present work, a set of shortened muskelin constructs was set up not only for mapping studies but also to increase the chances to find a construct amenable to crystallization. Therefore, the constructs were designed from scratch with the major aim to preserve secondary structure elements and the fold of the protein and its parts. To more precisely identify the potential folding units of muskelin, the domain annotation was put to the test and the results of further bioinformatic analyses - secondary structure predictions, conservation analyses and homology modeling - were considered to identify favorable positions for the boundaries.

---

[1] That is: a simple division in an N-terminal and C-terminal portion by Hasegawa et al. (2000) and Prag et al. (2004), a set of sequential truncations of the kelch motifs by Ledee et al. (2005) and a comprehensive set of truncations at the annotated domain boundaries by Valiyaveettil et al. (2008).

Table 3.1 – Results of domain and motif database searches with the rat muskelin sequence. For each kelch repeat motif in the muskelin sequence only the most significant hit is listed.

| Database | Identified domain | | Alignment | | E-value |
	ID	Name	Start	End	
PFAM	PF00754	F5_F8_type_C	19	148	$8.5 \cdot 10^{-7}$
	PF07738	Sad1_UNC	31	155	$3.5 \cdot 10^{-2}$
SMART	SM000667	LisH	172	204	$4.7 \cdot 10^{-3}$
	SM000668	CTLH	206	258	$5.3 \cdot 10^{-2}$
PFAM	PF13415	Kelch_3	282	334	$3.6 \cdot 10^{-11}$
	PF13418	Kelch_4	325	381	$7.7 \cdot 10^{-2}$
	PF13854	Kelch_5	387	418	8.6
	PF13418	Kelch_4	455	503	$1.7 \cdot 10^{-5}$
	PF01344	Kelch_1	520	561	0.6
	PF13854	Kelch_5	580	621	1.3

3.1.1 Validation of the domain annotation

To revise the annotation of the domains of muskelin, two major collections of known motifs and domains, the *SMART* and the *PFAM* databases, were scanned with the rat muskelin sequence. The *PFAM* database was searched using the *hmmscan* program of the *HMMER3* package, for the *SMART* database the provided search function of the database was used. The search results are summarized in table 3.1, sorted by the position in the muskelin sequence. For the kelch repeat motif, the *PFAM* database includes a set of six closely related motifs. A single repeat in the protein sequence can match several of these motifs and thereby be redundantly identified. For simplification, only the most significant hit for each kelch repeat motif in muskelin is listed in table 3.1.

The domain architecture resulting from the scans is in accordance with previous reports (Adams et al., 1998; Emes and Ponting, 2001; Prag et al., 2004). The findings of the database searches are complementary - domains identified in the *PFAM* database are not identified in the *SMART* database and vice versa. For the N-terminal part, two overlapping hits are found, a F5/F8 type C domain, also known and referred to as discoidin domain in this thesis, and a Sad1/UNC-like domain (SUN domain). The SUN domain is structurally highly similar to the discoidin domain, suggesting that both families are ancestrally connected (Sosa et al., 2012), which explains the double identification. Based on the expectation value (E-value), the discoidin domain is clearly

the more accurate assignment. Following the discoidin domain, the LisH and CTLH motif are identified, both with rather low significance. For the last domain, the kelch repeat domain, all kelch repeats are identified, only two of them with high significance, the first and the fourth one, one with a moderate E-value, the second repeat, and finally the third, fifth and sixth repeats are identified only with a very low reliability. This overall low reliability of the identification is probably due to the low sequence conservation of the kelch repeats in general and their inconsistent assignment in the databases (see also Introduction, section 1.2.4).

3.1.2 Analysis of the HMM profile alignments

For a more detailed analysis, hidden Markov model (HMM) profiles of the individual domains were generated based on the sets of homologous sequences provided by the databases (manually corrected for redundancies) and aligned to the muskelin sequence using *hmmalign* (figure 3.1 on the next page). The comparison of the sequence with the HMM profiles helps to identify the position of well-conserved and therefore supposedly structurally or functionally relevant residues, and the exact borders of the domains. In the analysis of the alignments, it must be considered that the *hmmalign* program has a certain weakness in extending a local alignment to the full length of the target sequence, leaving the termini often unmatched (Eddy, 2010, page 25). If possible, alignments of unmatched termini were completed manually as described in the figure legend.

For the discoidin domain, residues 24 to 136 are readily matched to the HMM profile. The two by far most conserved residues are tryptophanes at positions 29 and 80 of the motif, which are located in loop regions and probably stabilize the arrangement of the loops by hydrophobic interactions (see Introduction, section 1.2.4 on page 17). The first tryptophane is conserved in muskelin, the second is replaced by a methionine, preserving a hydrophobic residue at this position. Of the further conserved residues, the majority coincides with the β-strands that constitute the core fold of the domain and are also well conserved in the discoidin domain of muskelin. The alignment to the profile ends within the second to last β-strand, still it is obvious that the last two residues of the HMM profile are also conserved in muskelin.

The LisH motif is fully aligned to the HMM profile and generally well-conserved, with the prominent exception of the glutamate at position 28 of the motif, which is exchanged to a lysine in muskelin. This glutamate is of special relevance for the architecture of the LisH motif (see also Introduction, section 1.2.4 on page 18), as it stabilizes the arrangement of the second helices in the 4-helix-bundle formed upon dimerization of two LisH motifs.

Discoidin domain


```
¹⁹ALHKWSSFSSTYL--PENILVDKPNDQSSRWSSESNYPPQYLILKLERPAIVQNITFGKYEKTHVCNL
```



```
KKFKVFGGMNEENMTELLSSGLKNDYNKETFTLKHKIDEQMFPCRFIKIVPLLSWGPSFNFSIWYVEL¹⁵²
```

LisH motif


```
¹⁷²REQEAIRLCLKHFRQHNYTEAFESLQKKTKIAL²⁰⁴
```

CTLH motif


```
²⁰⁷PMLTDMHDKLVLKGDFDACEELIEKAVNDG-----LFNQYISQQEYKPRWSQIIPKS²⁵⁸
```

Kelch repeat domain


```
²⁷⁰MRGGHQMVIDVQTETVYLFGGWDG-T QDLADFWAYSVKENQWTCISRDT..13..³³⁰
³²⁵ARSCHKMCIDIQRRQIYTLGRYLDSS*SLKSDFYRYDIDTNTWMLLSEDT..13..³⁹¹
³⁸⁶LVFDHQMCMDSEKHMIYTFGGRIL-T*PQFSGLFAFNCQCQTWKLLREDS..13..⁴⁵⁸
⁴⁵⁵SRIGHCMLFHSKNRCLYVFGGQRSK- TYLNDFFSYDVDSDHVDIISDGT..13..⁵¹⁵
⁵¹²TGFTQRATIDPELNEIHVLSGLSKDK*NVRNSFWIYDIVRNSWSCVYKND..13..⁵⁷⁸
⁵⁸³PRFAHQLVYDELHKVHYLFGGNPGKS*MRLDDFWSLKLCRPSKDYLLRHC..16..⁶⁵¹
                                        ²⁴⁸KPRWSQIIPKS²⁵⁸
```

* KR2: VRNSK   **KR3:** CNGSVDDSRASE   **KR5:** EKREE   **KR6:** CSPK

For the CTLH domain, only the N-terminal part (residues 211 to 231) is aligned to the profile. Manually adjusting the unmatched part unveils a slight correspondence (positions 40 to 45), but generally the similarity of the C-terminal half is low. Notably, the secondary structure is predicted to consist of four β-strands, in contrast to the consensus of CTLH domains that is predicted to be α-helical. Since there is no structural information about CTLH domains available so far, one cannot judge the implications of these observations for this domain's architecture in muskelin, and it is even conceivable that it forms an additional atypical blade of the ensuing kelch domain β-propeller, similar to structure of the Krp1 protein (Gray et al., 2009).

The alignment of the kelch repeats reflects the confidence of their identification noted earlier - the first, second and fourth repeat are largely matched to the HMM profile, whereas for the other repeats substantial parts were not aligned. However, in the secondary structure prediction, the number and positions of the β-strands is well-conserved, with the exception of the last repeat lacking the fourth β-strand. The missing last β-strand indicates an N-terminal closure mechanism of the kelch repeat domain (see Introduction, section 1.2.4 on page 20). Indeed, when searching the region N-terminal to the kelch repeat domain, the last β-strand predicted prior to the first kelch repeat is matched to the fourth β-strand in the kelch motif.

The kelch motif is characterized by a glycine doublet at the end of the second β-strand, that allows for unusual flexibility of the peptide chain, and a tyrosine/tryptophane pair in the third and fourth β-strand that both stabilize the fold of the domain. In muskelin, the glycine doublet is conserved in four of the six repeats, and one single glycine is found in the others. The tyrosine is also found in four of the six repeats, and notably in each of these repeats either an asparagine or aspartate is found five positions later in the sequence as partner for side-chain-side-chain hydrogen bonds

---

Figure 3.1 *(facing page)* – **Alignment of the muskelin sequence to the HMM-profiles of its individual domains.** The coloring of the sequence denotes the secondary structure (blue: β-sheet, red: α-helix) and whether the *hmmalign* programm aligned the residue to the HMM profile (dark: aligned, pale: not aligned). The secondary structure assignment is based either on structural data (Discoidin domain, first helix of the LisH motif) or on predictions by *Phyre*[2] (second helix of the LisH motif) or PredictProtein (CTLH motif and Kelch repeat domain). Small superscript numbers indicate the numbering of the first and last residue in the sequence segment. The profiles were generated based on the family alignments provided in either the *PFAM* or the *SMART* database using the Skylign server (Discoidin domain: PF00754, LisH motif: SM00667, CTLH motif: PF10607, Kelch repeat: sequences of PF01344, PF07646, PF13418, PF13854 and PF13964, aligned with Clustal Omega). Red lines indicate common positions of insertions, with the intensity of the color increasing with the probability of an insertion. The row below the HMM profile displays the values of occupancy for each position in white for highest occupancy to blue for the lowest occupancy. The sequences were aligned to the profile using *hmmalign* with minor adjustments, CTLH: an insertion in the unaligned region was introduced to match the alignment to the HMM profile derived for SM00668, Kelch repeat domain: the unaligned part of the third repeat was adapted to align the secondary structure.

as previously described (Li, 2004). Finally, the prominent tryptophane, contributing to the hydrophobic packing, is found in all but the fourth repeat (if the N-terminal closing strand is included), where it is replaced by a valine. Overall, the kelch domain of muskelin matches the relevant traits and thus is likely to adopt the canonical β-propeller-fold.

Taking into account that the C-terminal part of the CTLH motif in muskelin does not match the consensus, it therefore seems rather likely that not the whole CTLH motif, as speculated above, but only this C-terminal part is an integral part of the kelch domain and thereby a repositioning of the domain boundaries and kelch repeat annotation as outlined in figure 3.2 seems appropriate. Note that the annotation is shifted more N-terminally for all kelch repeat motifs. This is a consequence of including the first β-strand of the repeat within the motif. It was not part of the motif used in the preceding annotation since this motif was defined based on a kelch repeat domain that uses a C-terminal closure (see Introduction section 1.2.4 on page 20).

3.1.3 Design of constructs
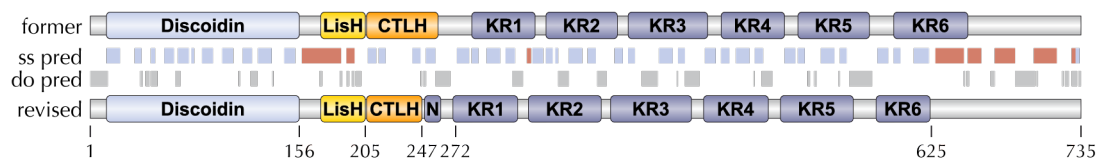
Figure 3.2 – Domain architecture of muskelin and boundaries of constructs. Initial domain architecture and domain architecture with revised position of the kelch repeat motifs (KR) and the N-terminal closing β-strand (N) is depicted below and above the prediction of secondary structure (ss pred) with β-strands in blue and α-helices in red, and disorder (do pred, disordered regions in grey) by *Phyre²*.

Based on the revised domain positions combined with predictions of secondary structure and disordered regions, the boundaries of the truncations were chosen (figure 3.2). The boundary after the discoidin domain was positioned right after the domain, before a long α-helix is predicted, that directly passes into the LisH motif. The LisH motif and CTLH motif are in close succession and the boundary is defined by a single intermediate residue. The boundary between the CTLH motif and kelch repeat domain was chosen twice, once preserving the CTLH motif as previously annotated and once based on the analyses outlined above, positioning the border within the previously annotated CTLH motif. The C-terminal border of the kelch repeat domain was again chosen based on the precedent analyses rather than the initial annotation, assuming that the C-terminal part forms another self-contained module.

3.1.4 Generation of pETM-SUMO

Following up on the construct design, the muskelin full-length sequence and the trun-cations were cloned into different expression vectors. The screening for a suitable fusion partner was planned using the pETM-vector system (Dümmler et al., 2005). For human muskelin, expression in *E. coli* was described to benefit from the fusion of a SUMO-tag (Kiedzierska et al., 2008), which was not available in the pETM-system. To include the SUMO-tag, a vector termed pETM-SUMO (figure 3.3) was generated as described in the methods section (section 2.2.1 on page 39). It shares the architecture and multiple cloning site of the pETM-vectors but includes the SUMO sequence of *S. cerevisiae* derived from the pETHSUL vector (Weeks et al., 2007). The shortened vari-ants and full-length muskelin were cloned into the generated pETM-SUMO vector as well as the pETM-11 and the pETM-30 vectors, which provide an N-terminal 6xHis- or GST-tag, respectively.
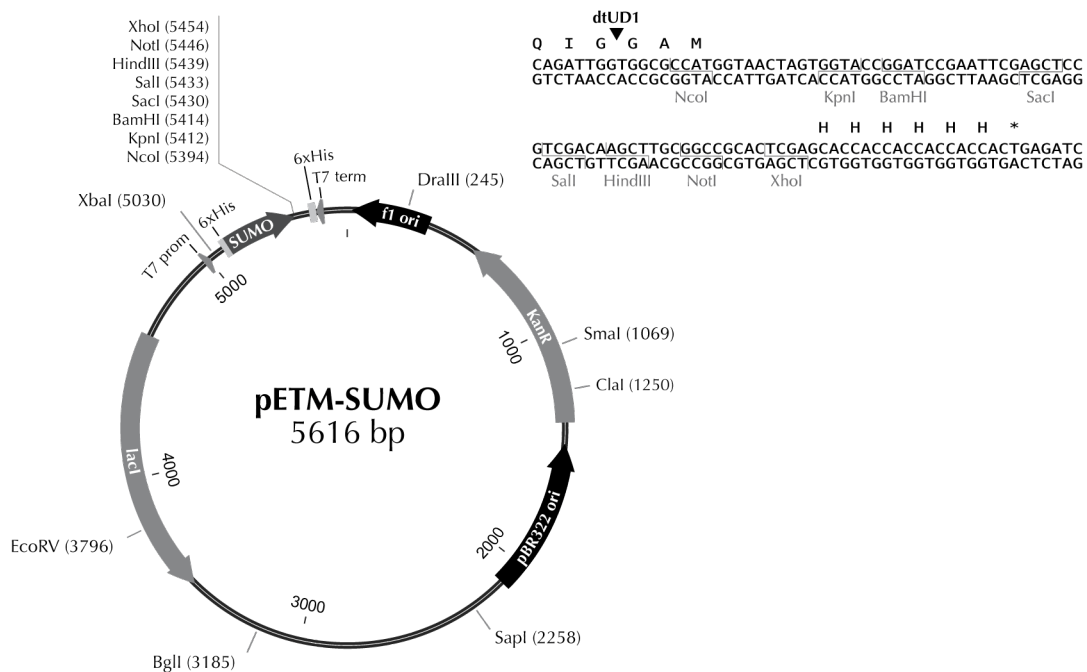
Figure 3.3 – Graphical map and details of the multiple cloning site for the pETM-SUMO vector showing the topology of the encoded sequences (T7 promoter: T7 prom, 6xHis-tag: 6xHis, SUMO-tag: SUMO, T7 Terminator: T7 term, f1 origin of replication: f1 ori, Kanamycin resistance gene: KanR, pBR322 origin of replication: pBR322 ori, lactose repressor gene: lacI) The sequence of the multiple cloning site is shown on the top right with the cleavage sites of restriction enzymes indicated. Translation of the last amino acids of the SUMO-tag and the C-terminal 6xHis-tag is given in single letter code and the cleavage position of the SUMO-specific protease (dtUD1) is marked by the arrowhead.

59
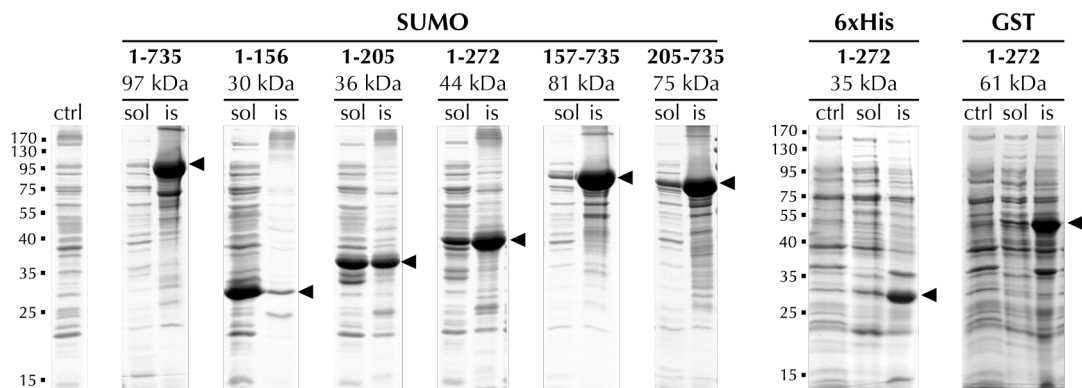
Figure 3.4 – Test expression of muskelin constructs at 20 °C in BL21-CodonPlus® (DE3)-RIL cells. Coomassie-stained SDS-PAGE with samples of cells before induction (ctrl) and of soluble (sol) and insoluble (is) fractions after expression (fused tag, muskelin sequence segment and size of fusion protein are indicated). Arrowheads mark the position of expressed protein. (left to right: M9, M1, M2, M3, M4, M5, M13, M23)

3.1.5 Cloning and Expression

A comprehensive list of all muskelin expression constructs generated in the course of this work can be found in the appendix section 6.4 on page 155. In the following, the used constructs are specified by their ID in this list at the end of each figure or table caption.

For first screenings of expression and purification conditions, several shortened variants and full-length muskelin were investigated in several expression tests (as described in methods, section 2.2.1 on page 41). 11% of the muskelin sequence from rat contains codons that are classified as rare for *E. coli* (less than 30% usage of the codon for the encoded amino acid) and thus potentially hamper expression in *E. coli*. To compensate for that, the BL21-CodonPlus® (DE3)-RIL strain was used in the first expression tests. The influence of the fused tags and temperature during expression on the level of solubly expressed protein was also examined. The result for several SUMO-tagged variants together with the 6xHis- and GST-tagged $Mkln_{1-272}$ construct expressed at 20 °C is shown in figure 3.4, the results of all tests were classified and are summarized in table 3.2.

For most SUMO-tagged proteins, a prominent band on the expected height could be observed upon expression of the fusion protein (see figure 3.4). The only exception is $Mkln_{1-272}$, for which a prominent band is observed at a slightly lower position than expected based on the molecular weight of the fusion proteins, independent of the fused tag. As the nucleotide sequence for all constructs was verified, possible explanations are either premature translation termination, co-translational degradation or an abnormal protein migration in SDS-PAGE.

	20 °C			15 °C		
	H	G	S	H	G	S
$Mkln_{1-156}$		n.d.			n.d.	
$Mkln_{1-205}$	n.d.					
$Mkln_{1-272}$						
$Mkln_{157-735}$						
$Mkln_{205-735}$		n.d.			n.d.	
$Mkln_{1-735}$						

Table 3.2 – Overview of results from expression tests using different temperatures (20 °C/ 15 °C) and fusion tags (H: 6xHis-tag, G: GST-tag, S: SUMO-tag). For each condition, the expression was classified by the ratio of expressed protein in the soluble and the insoluble fraction in five steps from soluble (white) to insoluble (dark red) or not determined (n.d.). (by row: M11, M21, M1; M12, M22, M2; M13, M23, M3; M14, M24, M4; M15, M25, M5; M19, M29, M9)

The solubility varied considerably between the different constructs - the shortest variant comprising only the discoidin domain is found almost completely in the soluble fraction; the solubility decreases for the longer variants until the full-length protein is finally found only in the insoluble fraction (table 3.2). This tendency is also observed for the 6xHis- and GST-tagged constructs. Comparing the solubility of the same construct for the different fusion tags, the majority of muskelin constructs showed a slightly higher soluble expression when fused to the SUMO-tag as compared to the 6xHis- or GST-tag. This is in line with earliers observation for full-length human muskelin (Kiedzierska et al., 2008). Finally, the tested temperatures during cultivation did not have a significant effect. It has to be noted that in comparison to the earlier study of full-length human muskelin the level of over-expression observed here was markedly higher and the use of Rosetta-gami™ (DE3) cells as expression host was not necessary for high expression levels (Kiedzierska et al., 2008). The use of Rosetta-gami™ (DE3) cells was thus abandoned, especially because muskelin is a cytoplasmic protein (see Introduction, section 1.2.2 on page 12) and its native environment is reducing rather than non-reducing, as provided by this strain.

As initial purifications indicated that the solubility of the expressed proteins was underestimated in the expression tests, and strong over-expression was already achieved in the tested conditions, further optimization was dispensable and expression at 20 °C in BL21-CodonPlus® (DE3)-RIL cells was applied for later expressions.

3.1.6 Purification

Initial purifications succeeded for the SUMO-$Mkln_{1-156}$ and the SUMO-$Mkln_{1-272}$ construct, but were heavily impaired by degradation and aggregation for full-length muskelin. To improve the purification, the buffer conditions were optimized using a ThermoFluor (Ericsson et al., 2006) screen of buffer solutions, pH values, salt concentrations and additives (figure 3.5 on the following page).

61
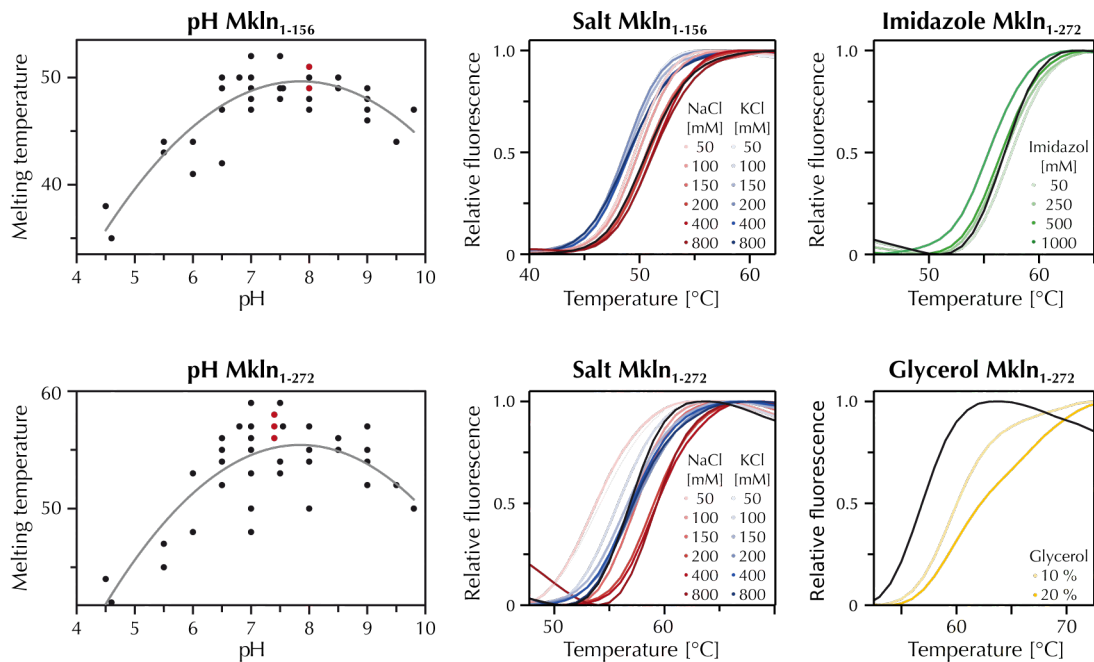
Figure 3.5 – ThermoFluor analysis of Mkln₁₋₁₅₆ and Mkln₁₋₂₇₂. *Left:* Melting temperatures plotted versus the pH of the buffer condition (red: reference measurements). *Middle:* Normalized fluorescence curves of salt conditions (red: NaCl, blue: KCl, black: reference). *Right:* Normalized fluorescence curves of additive conditions (top: imidazole, bottom: glycerol, reference in black). (M1, M3)

Despite considerable variation among the buffer substances at a given pH, an overall trend of stabilization by pH values in the neutral range was observable (figure 3.5, left). Describing this dependency with a simple polynomial function, the maximum of stabilization was found at a pH of 7.8 for both proteins. The two buffer conditions with the highest melting temperatures were PIPES pH 7.5 and ADA pH 7.0. As PIPES has a slightly higher pKa than ADA, and therefore was useful in pH ranges closer to the optimum, PIPES pH 7.5 was subsequently used as buffer condition for protein purification. Both constructs showed a preference for sodium chloride over potassium chloride (figure 3.5, middle), and stabilization by sodium chloride was constant for concentrations higher than 200 mM. Out of the tested additives, imidazole in the highest concentration slightly destabilized Mkln₁₋₂₇₂, whereas the addition of glycerol showed a strong stabilizing effect (figure 3.5, right). Neither EDTA nor DTT showed a significant effect (not shown). Following up on these results, sodium chloride at a concentration of 500 mM during affinity purification and 200 mM during size exclusion chromatography (SEC) was chosen and buffers were supplemented with 5% glycerol and 5 mM DTT in subsequent protein purifications. This markedly improved purification; moreover,

Table 3.3 – Purification of Mkln constructs. Domain scheme is depicted with the revised positioning as introduced in Figure 3.2. (M9, M1, M2, M3, M52, M53, M6, M8, M54, M55, M110)

Domain scheme	Construct	Purification steps	Average yield [mg][1]	[nmol][1]
	$Mkln_{1-735}$	IMAC - cleavage - AIEX - SEC	1	13
	$Mkln_{1-156}$	IMAC - cleavage[2]- SEC	15	830
	$Mkln_{1-205}$	IMAC - cleavage[3]- (AIEX)[3]- SEC	6.5	270
	$Mkln_{1-272}$	IMAC - cleavage - SEC	5	150
	$Mkln_{157-735}$	IMAC - cleavage - AIEX - SEC	6.5	100
	$Mkln_{205-735}$	IMAC - cleavage - AIEX - SEC	7	110
	$Mkln_{273-735}$	failed	n.d.	
	$Mkln_{247-625}$	IMAC - cleavage - SEC	8.5	200

[1] obtained from 1 l of *E. coli* culture
[2] performed on-column
[3] alternatively on-column cleavage or removal of SUMO-tag by AIEX

full-length muskelin and the other truncated constructs (except for $Mkln_{273-735}$) could readily be purified using the optimized buffer conditions, too (table 3.3). The final purification protocol comprises a starting nickel-ion affinity chromatography followed by cleavage of the SUMO-tag, an optional intermediate anion exchange chromatography step for improved purity and accelerated concentrating of proteins (as compared to a spin concentrator), and a final size exclusion chromatography.

With an average yield of 1 mg >95% pure full-length protein per liter of expression culture, the yield was comparable to the yield (1.5 mg of >90% pure protein) reported by Kiedzierska et al. (2008). For the shortened constructs, the molar yield was considerably higher than for the full-length protein and roughly proportional to the solubility upon over-expression as judged based on the expression tests (table 3.2 on page 61): $Mkln_{1-156}$ could be purified with the by far highest yield, followed in descending order by $Mkln_{1-205}$, $Mkln_{1-272}$, $Mkln_{205-735}$ and $Mkln_{157-735}$.

Notably, of all shortened constructs tested, only $Mkln_{273-735}$ and $Mkln_{157-272}$[2] were prone to aggregation and could not be purified. As the tendency to aggregate is a sign of impaired protein folding, this indicates that the kelch domain is not stable in $Mkln_{273-735}$, in contrast to $Mkln_{205-735}$ and $Mkln_{247-625}$. This supports the earlier notion that the kelch domain extends N-terminally into the annotated CTLH domain and the N-terminal closing is needed for a stable fold of the kelch repeat domain.

[2] Purification of both constructs was tested as part of the bachelor thesis of Daniel Krause (Krause, 2011).

3.2 Structural characterization of muskelin

At the beginning of this work, no structural information about muskelin was available. Although successful crystallization of human full-length muskelin had been reported earlier (Kiedzierska et al., 2008), this did not result in the determination of a crystal structure. As any structure of muskelin promised to provide valuable insights as basis for subsequent analyses, an extensive search for crystallization conditions of muskelin and shortened variants thereof was pursued following up on the successful establishment of purification protocols described in the previous section.

3.2.1 Crystallization

As initial step in the search for suitable crystallization conditions, sparse matrix crystallization screens were performed for several shortened muskelin constructs as well as the full-length protein and mutated variants thereof. A small number of crystals or useful lead conditions were observed (see figure 3.6), including:

- clusters of needles of $Mkln_{1-156}$ growing in condition H4 of the Protein complex suite,

- irregular shaped microcrystals of $Mkln_{157-735}$ growing in various conditions containing PEG 3350

- a single crystal of $Mkln_{1-205}$ growing in condition F10 of the Index screen,

- medium-sized, rounded crystals of $Mkln_{1-735}$ N144R in conditions C3, F3 and F10 (all contaning 1 M sodium malonate pH 6.0 as precipitant) of the Optmix™ 3 screen,

- tiny rods of $Mkln_{1-735}$ N144R in condition A10 of the Protein Complex Suite and needle clusters in the related condition F2 of the Index Screen.

No useful lead conditions were observed in crystallization screens of the constructs $Mkln_{1-272}$, $Mkln_{205-735}$, $Mkln_{205-625}$, $Mkln_{247-625}$, $Mkln_{1-735}$, $Mkln_{1-735}$ (F184E/L196Q) and $Mkln_{1-735}$ (N144R/F184E/L196Q), as well as in the cocrystallization attempts of $Mkln_{1-735}$ with $GABA_AR$ $\alpha1_{334-420}$ and $Mkln_{1-156}$ with $Mkln_{205-735}$.

Figure 3.6 *(facing page)* – **Crystal leads and optimized crystals** of the shortened muskelin constructs $Mkln_{1-156}$ (*top row*), $Mkln_{157-735}$ (*middle row*), $Mkln_{1-205}$ and $Mkln_{12-205}$ (*bottom row*). Composition of the mother liquor, the concentration of the protein and their mixing ratio is given below each image. (M1, M52, M2, M57)

Mkln$_{1-156}$
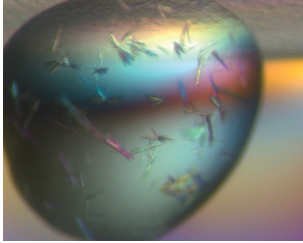
Initial condition
1.4 M sodium malonate pH 6.0
16 mg/ml protein
1:1
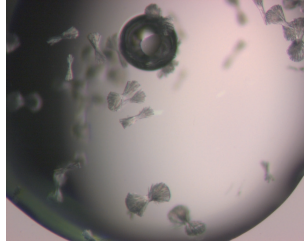
Finescreen
1.4 M sodium malonate pH 6.0
5% Glycerol
16 mg/ml protein
1:1
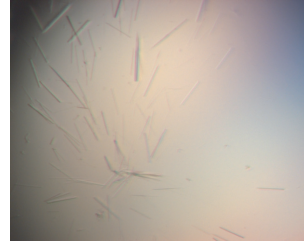
Finescreen with microseeding
1.4 M sodium malonate pH 6.2
16 mg/ml protein
1:1

Mkln$_{157-735}$
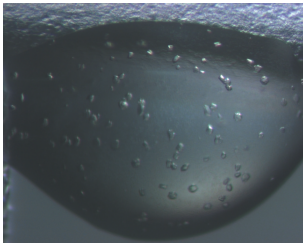
Initial condition
0.2 M ammonium citrate
pH 7.0
20% PEG 3350
15.9 mg/ml protein
1:1
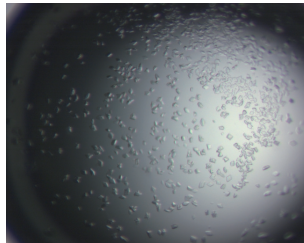
Finescreen
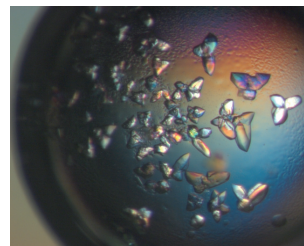0.25 M ammonium citrate tribasic
pH 7.0
20% PEG 3350
7 mg/ml protein
1:1
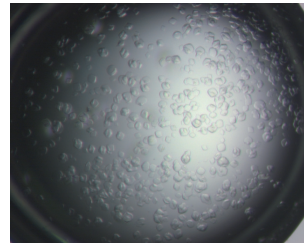
Finescreen
0.25 M ammonium citrate tribasic
pH 7.0
18% PEG 3350
7 mg/ml protein
1:1

Mkln$_{1-205}$ **Mkln$_{12-205}$**
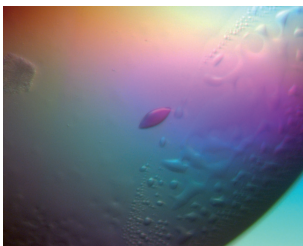
Initial condition
0.2 M sodium chloride
0.1 M BisTris pH 5.5
25% PEG 3350
16.6 mg/ml protein
1:1
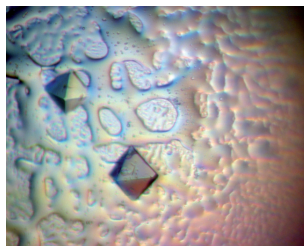
Finescreen
0.2 M sodium chloride
0.1 M BisTris pH 5.5
30% PEG 3350
16.6 mg/ml protein
1:1

Finescreen with microseeding
0.2 M sodium chloride
0.1 M BisTris pH 6.0
30% PEG 3350
16.2 mg/ml protein
1:1

The lead condition for crystallization of $Mkln_{1-156}$ (figure 3.6, top row, left) was followed up on as part of the bachelor thesis of Laura Wamprecht (Wamprecht, 2012). The needle-like crystals were reproducible in hanging drop vapor diffusion experiments, but formed tight bundles (figure 3.6, top row, middle). Using microseeding, single needles were obtained (figure 3.6, top row, right). They were tested for, but did not show diffraction at the ESRF. No further steps in optimization were undertaken and crystallization of $Mkln_{1-156}$ was abandoned upon gaining diffracting crystals of $Mkln_{1-205}$ (described below). Recently, the crystal structure of the closely related $Mkln_{1-174}$ was solved (Kim et al., 2014), and it is noteworthy that neither the precipitant, nor the buffer of the reported crystallization condition (25 % PEG 1500, 100 mM succinate-phosphate-glycine buffer pH 5.5) was contained in one of the initial sparse-matrix screens used in this work.

$Mkln_{157-735}$ formed microcrystals (figure 3.6, middle row, left), quasicrystals and spherulites in many different conditions containing PEG 3350. In the reproduction in a hanging drop vapor diffusion setup, $Mkln_{157-735}$ formed both irregularly shaped very small crystals in solution and thin plates at the surface of the drop (figure 3.6, middle row, middle and right, respectively). Neither seeding nor additive screening (Additive Screen HT™, Hampton Research) changed crystal appearance and growth. Crystals from different conditions were tested for diffraction at the ESRF, but did not diffract to a resolution of better than 20 Å. Consistently, in the recent publication by Kim et al. (2014), crystals of $Mkln_{151-735}$ were reported to grow in a related condition (0.1 M sodium acetate trihydrate, 12 % PEG 3350) and to show only poor diffraction.

For $Mkln_{1-205}$, the single crystal grown in the initial screen (figure 3.6, bottom row, left) already diffracted to a resolution of 3.3 Å at the in-house X-ray generator. In a grid screen varying the pH and the precipitant concentration, single crystals grew occasionally, but overall crystal formation was very rare, indicating a low rate of nucleation. To circumvent nucleation, microseeding was tested. This resulted in a reliable and increased crystal formation and was therefore included in all follow-up crystallization experiments. The $Mkln_{1-205}$ crystals (figure 3.6, bottom row, middle) grew in the tetragonal space group $P4_32_12$ and diffracted up to 1.7 Å. Although phasing succeeded using Sulphur SAD as described in section 3.2.2, several problems emerged in the process of structure solution and model refinement, hinting at an inherent problem of the crystal form. To improve the crystals, different strategies were pursued. While screening of additives did not lead to an improvement, and removal of the unresolved C-terminal portion inhibited crystal growth, removal of the flexible N-terminus altered the crystal form. The crystals of $Mkln_{12-205}$ (figure 3.6, bottom row, right) grew in space group $P2_12_12_1$ and diffracted up to 1.8 Å. The data derived from these crystals enabled the determination of the $Mkln_{12-205}$ crystal structure, as will be discussed in section 3.2.2.

Malonate condition
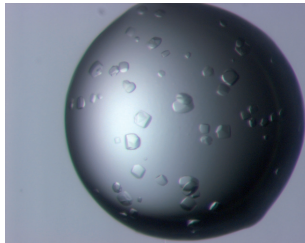
PEG condition

Initial condition
1 M sodium malonate pH 6.0
0.1 M sodium acetate pH 4.6
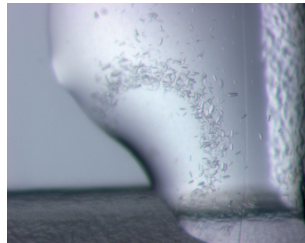
Initial condition
0.1 M potassium chloride
0.1 M Tris pH 8.0
20% PEG 2000 MME

Refined condition
0.1 M sodium citrate
0.1 M Tris pH 8.0
20% PEG 2000 MME
7.1 mM C8E5

Figure 3.7 – Crystal leads and optimized crystals of Mkln$_{1-735}$ N144R. Composition of the mother liquor is given below each image. (M96, M112)

Among the tested full-length variants, only Mkln$_{1-735}$ N144R (see section 3.3.3 and 3.3.4 for details) gave rise to crystals. It crystallized in two sets of conditions; single medium-sized, rounded crystals grew in conditions containing 1 M sodium malonate at pH 6.0 as precipitant, and small rods or needles formed in two conditions containing PEG 2000 MME as precipitant at a pH of 8.0 and 8.5, respectively.

The crystals grown in the malonate containing conditions could readily be reproduced in a hanging drop vapor diffusion setup, but diffracted only to low resolution (25 Å). In combination, additives (Additive screen HT™, Hampton Research) and different crystallization temperatures (4°C, 12°C and 20°C) were screened, but neither led to changes in the crystal form nor improved the diffraction considerably.

The PEG-based condition was followed up on together with Susanne Sauer as part of her bachelor thesis (Sauer, 2014). Variation of the pH and the concentration of PEG, as well as the substitution of the salt and screening of additives (Additive Screen HT™, Hampton Research) predominantly yielded small, thin needles, and plates in few conditions, which diffracted up to 25 Å. Notably, both in the screen of substitute salts and of additives, conditions containing citrate yielded plates. Exchanging potassium chloride in the basis condition to sodium citrate and again screening additives yielded predominantly plates (often layered or clustered), but also three dimensional crystals in three conditions, containing either 2,2,2-trifluoroethanol, *n*-octyl-β-D-glucoside and *n*-dodecyl-β-D-maltoside, of which the latter diffracted up to 12 Å. Thus, two out of three non-ionic detergents contained in the additive screen improved crystal growth and one significantly improved diffraction, prompting a follow-up screen of detergents as additives.
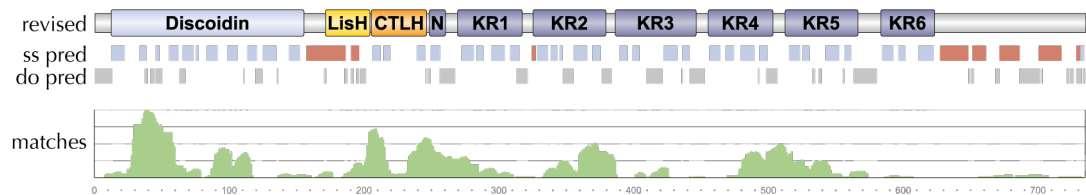
Figure 3.8 – Mass spectrometry analysis of Mkln$_{1-735}$ N144R after limited proteolysis. Revised domain scheme as introduced in figure 3.2 aligned with the predictions of secondary structure (ss pred, β-strands in blue, α-helices in red) and disorder (do pred, disordered regions in grey). Match frequency in mass spectrometric analysis (matches) is plotted below in light green. (M96)

Moreover, since both crystal forms showed poor diffraction, it seemed likely that inherent properties of the protein hamper uniform crystal growth. It was therefore tested whether removal of flexible parts would improve crystal quality. In prior experiments, limited proteolysis was performed to find a suitable protease and identify proteolysis-resistant parts of the protein. Using trypsin, a large proteolysis-resistant fragment was observed and analyzed using mass spectrometry (figure 3.8). Judged by the frequency of identification, a short N-terminal and a large C-terminal portion including the last kelch repeat motif was cleaved of during proteolysis. The resulting fragment Mkln$_{22-576}$ would have a size of 64 kDa, which also matches the position of the observed band in SDS-PAGE analysis. Based on these results, cloning of two shortened constructs, Mkln$_{12-735}$ N144R that only lacks the disordered N-terminal stretch and Mkln$_{12-625}$ N144R that lacks both the N-terminus and the whole C-terminal module, and *in situ* proteolysis during crystallization were pursued.

While *in situ* proteolysis almost completely prevented crystal growth, three-dimensional crystals were obtained for Mkln$_{12-725}$ N144R in many conditions of the detergent screen (Detergent screen™, Hampton Research). Still, crystals predominantly grew clustered or twinned and diffracted only to very low resolution, with the exception of crystals grown in the presence of C8E5 that diffracted up to 12 Å.

Overall, this is a considerable improvement to earlier reports of small, flawed or unstable crystals of full-length muskelin (Kiedzierska et al., 2008), and a significant step closer to its structural characterization, as pursuing this optimization is promising to yield crystals diffracting up to useful resolutions. A crystal structure following up on this groundwork would be very valuable, as upon completion of this thesis, there is still no structural information available about muskelin's unique C-terminus, the parts of muskelin that are decisive in GABA$_A$ receptor binding (see section 3.4), the arrangement of the muskelin oligomer (see section 3.3.4) or about any other protein that contains a CTLH motif.

68
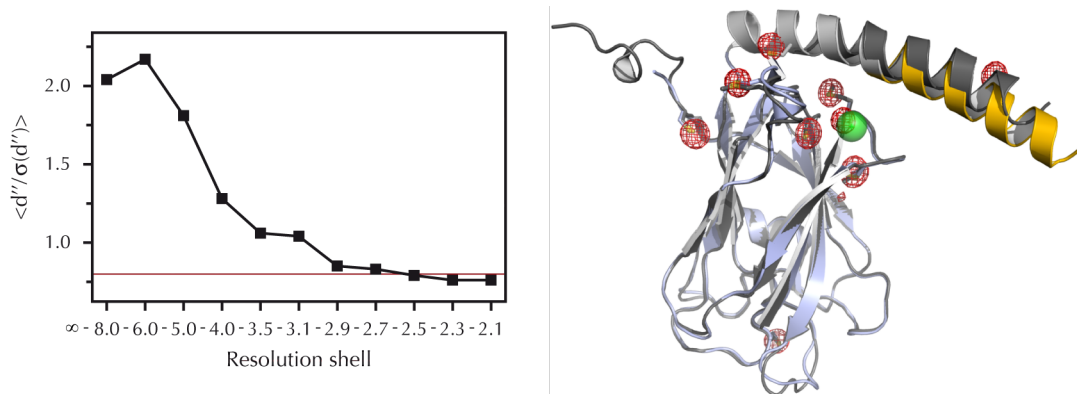
Figure 3.9 – Structure solution of Mkln$_{1-205}$ and Mkln$_{12-205}$. *Left:* The signal-to-noise ratio for the anomalous differences is plotted versus resolution for the dataset used for SAD phasing. The threshold of 0.8, which is considered to correspond to no significant anomalous signal, is indicated by the red line. *Right:* The search model of Mkln$_{1-205}$ (dark grey) and aligned final model of the Mkln$_{12-205}$ monomer (colored according to domain annotation) in ribbon representation with all sulphur-containing sidechains displayed as sticks and the modeled chloride as sphere are shown overlaid with the final anomalous map (red, contour-level 5 σ).

3.2.2 Structure solution of Mkln$_{1-205}$ and Mkln$_{12-205}$

First datasets recorded of Mkln$_{1-205}$ crystals revealed that the crystals diffracted up to resolutions of about 2 Å and were neither particularly mosaic nor prone to radiation damage, thus meeting important criteria for gaining sound crystallographic data. Since the discoidin domain covers more than 70% of the Mkln$_{1-205}$ sequence, it was initially tried to solve the phase problem by molecular replacement using structures of homologous discoidin domains (such as that of human neuropilin and of a hyaluronidase from *C. perfringens* with 23% sequence identity, and 43%/54% sequence similarity, respectively) as search models, but all attempts failed to give a solution.

In the next step, as Mkln$_{1-205}$ contains nine sulphur-atoms (five cysteines and four methionines) and the crystals diffracted to adequate resolution, it was tested to use the anomalous scattering of the intrinsic sulphur atoms for phasing by single-wavelength anomalous diffraction (SAD). The data for phasing were collected at a wavelength of 1.9 Å to enhance the anomalous scattering (the anomalous scattering contribution f" is 0.88 electrons at this wavelength). Using these data, together with an additional native high resolution dataset extending to 1.7 Å, Dr. Jochen Kuper and Dr. Hermann Schindelin were jointly able to solve the phase problem. Data up to a resolution of 2.7 Å were used for the substructure search, where the signal-to-noise ratio was 0.83 in that shell (figure 3.9). The final substructure used for phase determination contained nine anomalous sites, corresponding to eight out of the nine intrinsic sulfur atoms of the protein. The starting methionine located in the flexible N-terminus of the protein

was not observed, yet an additional anomalous peak was used in phasing, which was later interpreted as chloride atom caught in a hydrophobic pocket of the protein. In the process of substructure refinement and phase determination, multiple rounds of reprocessing and phase refinement were needed before the determined phases were accurate and the derived electron density map detailed enough to allow for automated building of a partial model. The model was completed manually and refined, resulting in a final model that encompassed residues 8 to 187 of the protein. Yet, refinement of the model yielded a free R-factor of 34% which could not be improved despite extensive rebuilding efforts. This indicated an inherent problem with the data derived from the tetragonal space group, and necessitated optimization of the crystal form.

As described in section 3.2.1 on page 66, removal of the flexible N-terminal residues resulted in a new crystal form: $Mkln_{12\text{-}205}$ crystallized in space group $P2_12_12_1$ with two copies of $Mkln_{12\text{-}205}$ in the asymmetric unit. For this orthorhombic crystal form phasing was accomplished by molecular replacement with the model obtained for $Mkln_{1\text{-}205}$, and refinement of the model resulted in an R-factor of 15.62% (R_{free}=19.31%) with good stereochemistry. All statistics of data collection, phasing and refinement are summarized in table 3.4 (facing page). Notably, at the position corresponding to the ambiguous peak in the anomalous map observed for the earlier crystal form, no density was found in the electron density maps derived for crystals of $Mkln_{12\text{-}205}$, thus no bound chloride was modeled.

In the final model of $Mkln_{12\text{-}205}$ residues 12 to 187 were visible while the last 18 residues could not be resolved, as before for $Mkln_{1\text{-}205}$. Deviations of the final model for $Mkln_{12\text{-}205}$ to the search model were minor, which is reflected by a root mean square deviation (r.m.s.d.) of 0.39 Å when structurally aligning the two models. Upon comparison of the two crystal forms, a relationship between the unit cell dimensions is obvious. In addition, the total number of molecules in the unit cell is the same with the eight asymmetric units of the tetragonal crystal form containing a single $Mkln_{1\text{-}205}$ molecule and the four asymmetric units of the orthorhombic crystals containing two copies of $Mkln_{12\text{-}205}$. Yet, reprocessing of the tetragonal data in space group $P2_12_12_1$ did not improve the refinement, ruling out an incorrectly assigned space group as explanation for the problems encountered in solving the $Mkln_{1\text{-}205}$ structure.

3.2.3 Structural features of $Mkln_{12\text{-}205}$

$Mkln_{12\text{-}205}$ crystallized as dimer, with one dimer per asymmetric unit (figure 3.10 on page 72). As expected, the N-terminal discoidin domain folds into a β-barrel, which is followed by an extended α-helix. The α-helices of the two molecules in the dimer are arranged in a roughly antiparallel fashion and are slightly twisted. The C-terminal parts of both helices, corresponding to the first helix of the LisH motif, form the dimer-

Table 3.4 – Data collection, phasing and refinement statistics. Statistics for the highest-resolution shell are shown in parentheses.

	Mkln$_{12\text{-}205}$	Mkln$_{1\text{-}205}$	
		Native	Sulphur SAD
Data collection statistics			
Wavelength (Å)	0.95373	0.91841	1.9
Resolution range (Å)	33.8 - 1.8 (1.9 - 1.8)	65.1-1.7 (1.79 - 1.7)	34.1 - 2.28 (2.4 - 2.28)
Space group	P2$_1$2$_1$2$_1$	P4$_3$2$_1$2	P4$_3$2$_1$2
Unit cell (Å)	a = 63.06, b = 65.30, c = 101.39	a = b = 65.15, c = 101.53	a = b = 64.99, c = 101.56
Total reflections	176855	106685	447829
Unique reflections	39509 (5686)	24751 (3535)	10549 (1498)
Multiplicity	4.5 (4.5)	4.3 (3.9)	42.5 (36.7)
Completeness (%)	100.00 (100.00)	100.00 (100.00)	100.00 (100.00)
Mean I/σ(I)	17.0 (4.7)	9.6 (1.8)	30.0 (6.5)
Wilson B-factor (Å^2)	26.4	25.0	43.9
R$_{merge}$[a]	4.9 (30.1)	6.9 (67.6)	9.6 (90.9)
Phasing statistics			
Figure of merit			0.324
Number of sites			9
Refinement statistics			
R-factor[b]	0.1562 (0.1974)		
R-free[c]	0.1931 (0.2313)		
Number of atoms	3398		
macromolecules	3019		
ligands	102		
water	277		
Protein residues	357		
RMS deviations in			
bond lengths (Å)	0.014		
bond angles (°)	1.42		
Ramachandran favoured[d] (%)	96		
Ramachandran outliers[d] (%)	0		
Clashscore	13.94		
Average B-factor (Å^2)	41.2		
macromolecules	40.2		
solvent	47.1		

[a] $R_{merge} = \sum_{hkl} \sum_j |I(j) - [I]| / \sum_{hkl} \sum_j I(j)$,
where $I(j)$ is the value of the jth measurement of the intensity of a reflection, $[I]$ is the mean value of the intensity of that reflection, and the summation is over all measurements.

[b] $R = \sum_{hkl} |F_{obs} - F_{calc}| / \sum_{hkl} |F_{obs}|$,
where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively, for all data (no σ cutoff).

[c] R-free = R calculated with 5% of the reflection data chosen randomly and omitted prior to the start of refinement.

[d] Ramachandran statistics have been determined with MolProbity (Davis et al., 2004).
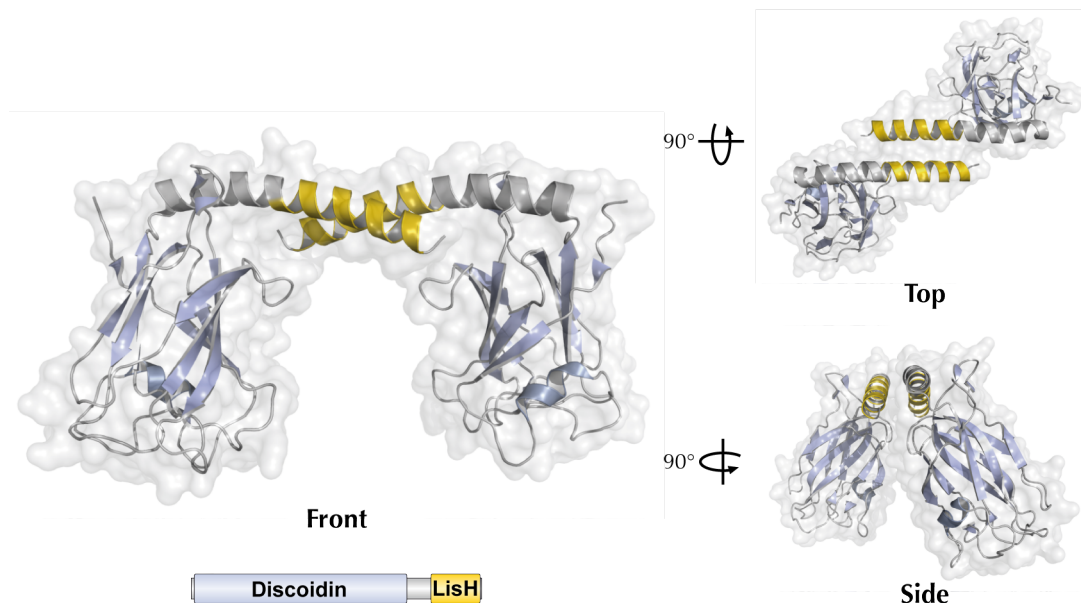
Figure 3.10 – Overall structure of Mkln$_{12\text{-}205}$ shown in ribbon and surface representation in front, top and side views. Secondary structure elements are colored according to the domain scheme depicted below the front view and the surface is rendered semi-transparent in light grey.

ization interface. The ensuing residues that, based on homologous structures, are predicted to form the second helix of the LisH motif were not visible.

The discoidin domain of Mkln$_{12\text{-}205}$ adopts the typical tertiary structure associated with this fold: a distorted β-barrel formed by two anti-parallel β-sheets consisting of three (β3, β6 and β8) and five β-strands (β1, β2, β4, β5 and β7), respectively (figure 3.11). The N- and the C-termini lie in close proximity on the top of the barrel (best visible in front view, figure 3.10), however, it does not feature the typical disulphide bridging of the termini found in other discoidin domains (Kiedzierska et al., 2007). Also, no other pair of cysteines in the discoidin domain was found in suitable distance for formation of a disulphide bond, contradictory to earlier suggestions (Kiedzierska et al., 2008). A putative disulphide bridge in the LisH motif will be discussed later in section 3.3.2.

The core between the β-sheets is packed with interdigitating hydrophobic residues and the strands are connected by long, flexible loops (three on the top and four on the bottom when viewed from the front). Additional secondary structure elements are located in these loops: one short β-strand, βL1, in the N-terminal loop, a 3$_{10}$-helix in the β1-β2-loop, and two very short two-stranded antiparallel β-sheets, one formed by βL2 and βL5 in the β1-β2-loop and the β7-β8-loop, respectively, and one β-hairpin
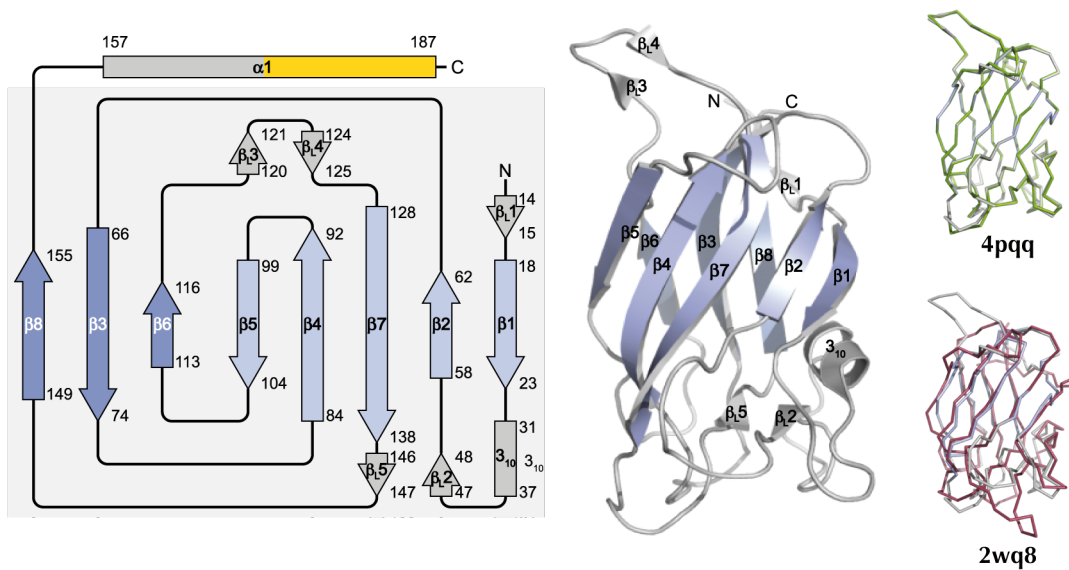
Figure 3.11 – Architecture and topology of muskelin's discoidin domain. *Left:* The topology diagram of Mkln₁₂₋₂₀₅ is shown with the discoidin domain underlaid in light grey. *Center:* The discoidin domain part of the Mkln₁₂₋₂₀₅ structure (residues 12 to 157) is shown in ribbon representation, color coded and labeled according to the topology diagram. *Right:* The discoidin domain part of the Mkln₁₂₋₂₀₅ structure in light blue is shown in Cα representation structurally aligned with the recently published structure of muskelin's discoidin domain (PDB code 4pqq, Kim et al., 2014)) in green (*right, top*) and with its closest structural homolog obtained by 3D-blast (Yang and Tung, 2006; Tung et al., 2007), the discoidin domain of the *Gibberella zeae* galactose oxidase (PDB code 2wq8, Rannes et al., 2011), in purple (*right, bottom*).

in the β6-β7-loop formed by βL2 and βL3. The small βL2-βL5 sheet forms a lid to the bottom of the barrel. Notably, it contains Trp47, the most conserved residue in discoidin domains (see figure 3.1 on page 56), which, together with Ile147, faces the hydrophobic inside of the barrel.

Overall, the core fold of the discoidin domain (residues 12 to 156) agrees well with known structures, as indicated by an r.m.s.d. value of 1.8 Å for all atoms and 1.7 Å for the main chain only, upon structural alignment with the closest structural homolog (figure 3.11). The fold of the β-sheet core is well aligned, the largest deviations are found in the arrangement of the loops. Superposition of the discoidin domain with that of the recently published Mkln₁₋₁₇₄ structure (Kim et al., 2014) results in r.m.s.d. values of 0.2 Å for all atoms. As before, the core is closely superposed and the major differences are only found in the arrangement of the loops, thus underscoring their flexibility. This structural preservation pattern is in line with the function and evolution of discoidin domains as versatile binding modules with a common binding site for a large variety of ligands (see Introduction, section 1.2.4 on page 17).
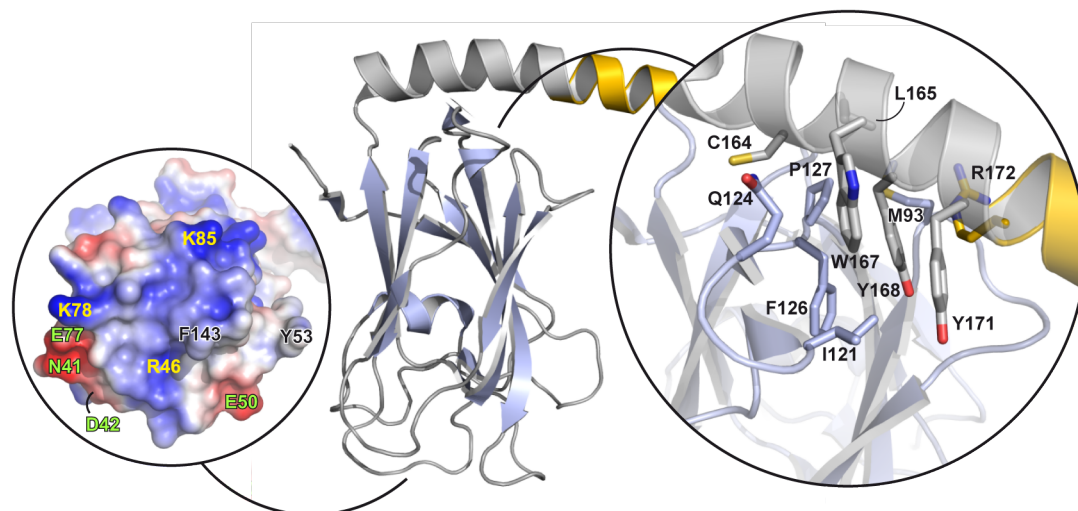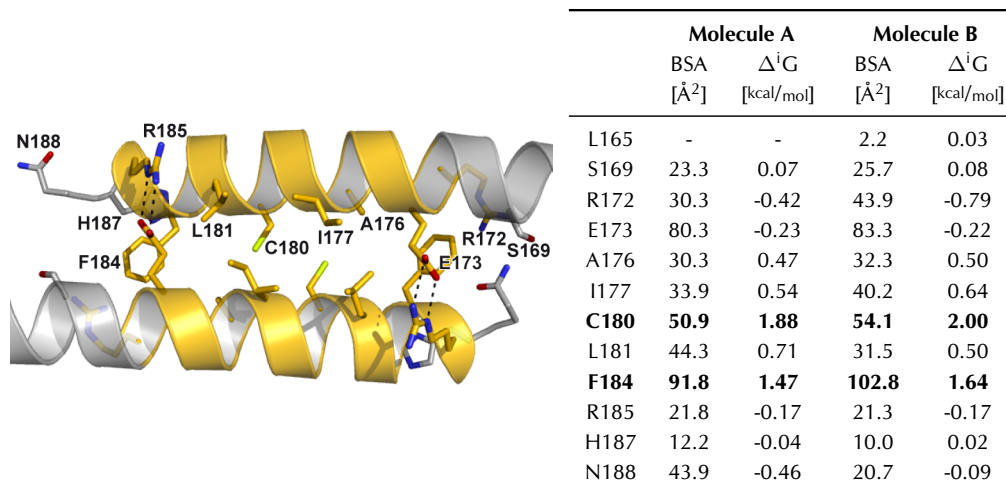
Figure 3.12 – Unique features of muskelin's discoidin domain. A surface representation colored according to electrostatic potential (positively charged patches in blue, negatively charged patches in red) of the generic binding site at the bottom (*left*) and a close-up view of the interface to the α-helix with important residues in stick representation (*right*) are shown next to the $Mkln_{12-205}$ model in ribbon representation.

The loops typically harboring the common binding site at the bottom of the β-barrel in muskelin form a rather flat surface and do not feature the prominent spike arrangement of the β1-β2 and the β3-β4 loop observed especially in membrane-binding discoidin domains. The surface has a mixed composition (figure 3.12), with two prominent hydrophobic residues, Phe143 and Tyr53, a rather positively charged surface region (Arg46, Lys78 and Lys85) surrounded by two negative patches (Glu50 as well as Glu77, Asn41 and Asp42).

At the top of the β-barrel, the arrangement of the β6-β7 loop with the protruding βL2-βL3-hairpin is unique to muskelin, and is much less prominent in related structures. The loop forms contacts with the helix that starts C-terminal to the discoidin domain, especially several residues organized in and in close vicinity to the β-hairpin located in this loop (Ile121, Gln124 and Phe126) are tightly packed with bulky residues (Trp167, Tyr168 and Tyr171) from the helix (figure 3.12). Despite the absence of these aromatic residues in the related structure by Kim et al. (2014), the β6-β7 loop adopts the same conformation. Instead, a PEG molecule is found adjacent to the loop, where it would overlap with the position of Tyr168, and contacts Phe126, thus partially replacing the contacts mediated by the terminal helix in the $Mkln_{12-205}$ structure. The binding site of this PEG molecule also coincides with the pocket in which the ambiguous chloride atom was caught in the initial $Mkln_{1-205}$ structure.

	Molecule A		Molecule B	
	BSA [Å^2]	$\Delta^i G$ [kcal/mol]	BSA [Å^2]	$\Delta^i G$ [kcal/mol]
L165	-	-	2.2	0.03
S169	23.3	0.07	25.7	0.08
R172	30.3	-0.42	43.9	-0.79
E173	80.3	-0.23	83.3	-0.22
A176	30.3	0.47	32.3	0.50
I177	33.9	0.54	40.2	0.64
C180	**50.9**	**1.88**	**54.1**	**2.00**
L181	44.3	0.71	31.5	0.50
F184	**91.8**	**1.47**	**102.8**	**1.64**
R185	21.8	-0.17	21.3	-0.17
H187	12.2	-0.04	10.0	0.02
N188	43.9	-0.46	20.7	-0.09

Figure 3.13 & Table 3.5 – The LisH-dimerization interface. *Left*: Close-up view of the dimerization interface in cartoon representation is shown with side chains in the interface in stick representation. The hydrogen bonds between Glu173 and Arg185 are indicated as dashed lines. *Right*: The table summarizes the results of the PISA analysis (Krissinel and Henrick, 2007) for the dimerization interface. Residues that contribute to the interface are listed with their buried surface area (BSA) and solvation energy effect ($\Delta^i G$). The residues with the strongest contributions, Phe184 and Cys180, are printed in bold.

Notably, on the part of the helix the contacts to the discoidin domain are almost exclusively mediated by the N-terminal portion preceding the LisH motif. The continuous α-helix can thereby be separated into two parts - the N-terminal part that attaches to the discoidin domain, and the C-terminal portion, that mediates dimerization - and this separation perfectly matches the domain annotation (see section 3.1.1 and figure 3.2).

Within the LisH motif portion, side chains of residues interdigitate and thereby form a compact dimer interface (figure 3.13). The interface is mainly composed of hydrophobic residues (Ala176, Ile177, Cys180, Leu181 and Phe184), with the side chains of Cys180 marking the center, and it is embraced by the hydrogen-bonded side chains of Glu173 and Arg185. The central Cys180 residues are in suitable distance to form a disulphide bond and indeed were observed bridged in initial structures (discussed in section 3.3.2).

As mentioned above, the ensuing residues (residues 189 to 205) corresponding to the second half of the LisH motif, including the second helix, were not resolved. This is a deviation to the generic architecture of LisH motifs, and may be due to either of three reasons: it could be an effect of crystal packing, being mutually exclusive with the generic arrangement, it could be a specific artifact of the construct used for crystallization and finally these residues could be intrinsically flexible. Interestingly,

the presence of the unresolved residues was a prerequisite for crystallization, since the shortened construct Mkln$_{1-187}$, which was generated during crystal optimization (see section 3.2.1), did not crystallize. As these residues significantly contribute to dimerization (see section 3.3.2) and Mkln$_{12-205}$ crystallized in its dimeric form, they most likely promote crystallization by enhancing dimerization.

Based on the homologous structures of LisH motifs in the proteins LIS1, FOP and TBL1 (Kim et al., 2004; Mikolajka et al., 2006; Oberoi et al., 2011), the second helix would be predicted to fold back onto the first helix, forming a 4-helix bundle upon dimerization. It cannot be ruled out that for Mkln$_{12-205}$ this arrangement is prevented by crystal contacts, as symmetry mates occupy this position. In all other known structures, the crystallized proteins continue C-terminally to the LisH motif, and the C-terminal parts constrain movements of the second helix. In contrast, Mkln$_{12-205}$ stops right after the LisH motif, thus provides no such constraints, which might facilitate displacement of the second helix.

Notably, the most conserved residue in LisH motifs, the glutamate at the end of the second helix (compare figure 3.1 on page 56), which is also found in all other known LisH motif structures, is replaced by a lysine in muskelin. The conserved glutamate holds a vital role in the known structures. Across the dimer interface, it forms hydrogen bonds with the backbone amides of the opposing helix and moreover a favorable charged interaction with the helix dipole (see Introduction, section 1.2.4 on page 18, figure 1.9), thus stabilizing the arrangement of the second helices. The lysine in muskelin cannot fulfill these important functions, and this loss of stabilization might be an important factor for deviations of muskelin's LisH motif to known structures.

Comparison to the homologous structures reveals a further fundamental difference in the arrangement of the LisH motif in muskelin: the overall rotation of the helices with respect to each other is in opposite direction (figure 3.14). Despite this markable difference, the basic principle, a dimerization via anti-parallel helices with a predominantly hydrophobic interface, is the same and residues contributing to the interface are found at homologous positions. To further probe the reversed arrangement of the first LisH helices, a normal mode analysis (NMA) of both, the Mkln$_{12-205}$ monomer and the dimer was performed to assess possible large-scale shape-changing motions. For the monomer, few restraints limit the movements and the largest motions are predicted for the LisH part of the extended helix. In the 7th normal mode[3], this portion oscillates between being bent upwards (as observed in other LisH motif structures) and downwards (as found in the Mkln$_{12-205}$ crystal structure). In contrast, movements of the helices in the dimer are more restrained, and no oscillation between the opposing rotation angles was predicted.

---

[3] The modes one to six represent the three basic translations and rotations.
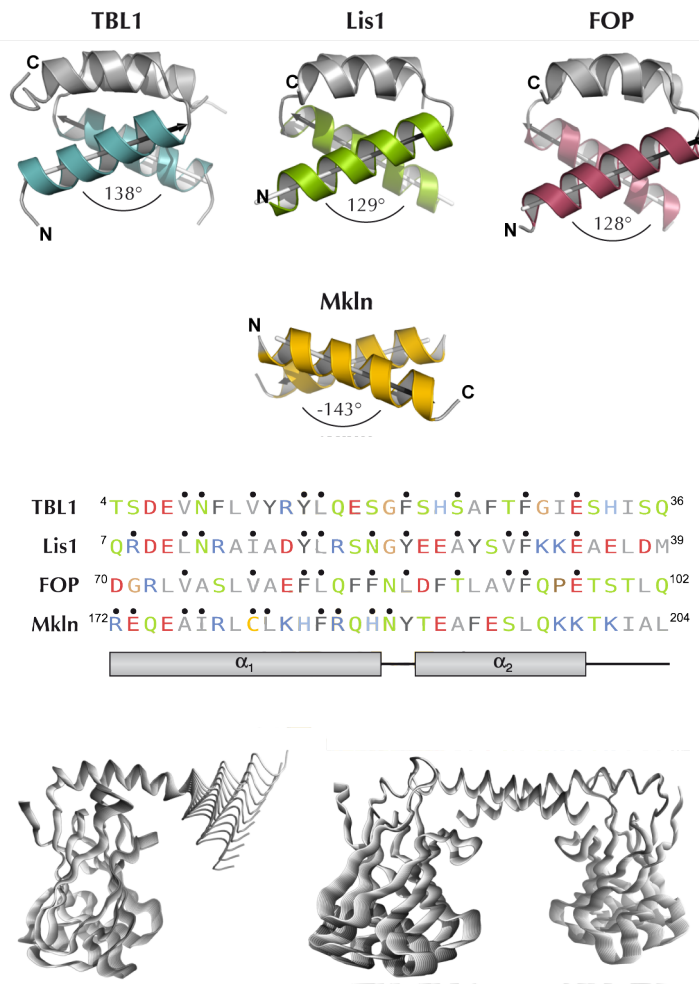
Figure 3.14 – Comparison of LisH motif architecture between muskelin, TBL1, LIS1 and FOP. *Top*: Ribbon representations of the respective LisH motifs are depicted with the first helix in each case colored differently. The angles between the helix axes (represented by vectors) are indicated. *Center:* A sequence alignment of these four LisH motifs is shown with amino acids color coded by properties (positively charged in blue; negatively charged in red; polar in green; non-polar aliphatic in light grey; non-polar aromatic in dark grey; G in light brown; P in dark brown; C in yellow). Residues contributing to the dimerization interface are marked with dark circles. *Bottom:* Normal mode analysis for the $Mkln_{12-205}$ monomer (left) and dimer (right), with prediction of vibration in the 7th mode depicted as an array of backbone traces colored from dark to light grey. PDB codes: 2xtd (TBL1), 1uuj (Lis1), 2d68 (FOP).

So while in isolation, the helix would be able to adopt the canonical arrangement, it is fixed in the reversed position within the dimer, and a conversion of the arrangement is, if at all, likely to be possible only upon dissociation of the dimer. It would be very interesting to investigate whether both arrangements are possible for $Mkln_{12-205}$ (and only the reversed was favored in crystallization) and how the helices are arranged when embedded in the full protein. A crystal structure of either $Mkln_{1-735}$ or the shortened $Mkln_{157-735}$ would reveal the arrangement of the LisH motif in the context of the full-length protein. This reinforces the interest to follow up on the achievements in the crystallization of the N144R variant of $Mkln_{1-735}$ described before (section 3.2.1).

Investigating the arrangement of the helices in $Mkln_{12-205}$ in solution using other structural techniques such as SAXS and nuclear magnetic resonance (NMR) spec-

77

troscopy, would be very interesting, but this is a challenging endeavor. A SAXS analysis was hampered by the fact that the available modeling tools had proven incapable of deriving the architecture of Mkln$_{12-205}$ when applied to theoretical scattering curves calculated for the Mkln$_{12-205}$ crystal structure (see Methods, section 2.2.3), and therefore were likely not suitable to gain reliable information about possible arrangements. For NMR analysis the biggest obstacle is presumably the size of Mkln$_{1-205}$. The monomer dimer equilibrium only necessitates further experimental steps to distinguish intra- and intermolecular bondings, but does not prevent analysis. But already the size of the monomer (24 kDa) is close to the limits of an analysis in conventional NMR experiments, with the size of the dimer ultra-high field NMR instruments and more advanced analyses need to be used.

3.2.4 SAXS analyses

In parallel to the X-ray crystallographic studies, SAXS analysis was used as a complementary technique to derive structural information. While SAXS analysis only provides low spatial resolution (up to 10 Å), it is performed with proteins in solution and hence does not require the availability of diffracting crystals. SAXS data were recorded for the shortened constructs Mkln$_{1-156}$, Mkln$_{1-205}$, Mkln$_{1-272}$, Mkln$_{157-735}$ and Mkln$_{205-735}$ as well as the full-length protein Mkln$_{1-735}$, and analyzed in cooperation with Dr. Bodo Sander. Modeling was only pursued for Mkln$_{205-735}$ for different reasons. The structure of Mkln$_{1-156}$ and of the major portion of Mkln$_{1-205}$ was known from the crystal structure, and SAXS analysis proved to be unsuitable to derive information about the arrangement of the portion that could not be resolved in the crystal structure (see Methods, section 2.2.3). For Mkln$_{157-735}$ and Mkln$_{1-735}$, data analysis was for one thing complicated by oligomerization and potential aggregation, as judged based on the concentration-dependent differences of the scaled scattering curves (see Appendix, figure 6.1 on page 158) and the estimated radii of gyration and Porod volumina (see Results, figure 3.16 on page 83). Moreover, for Mkln$_{1-735}$, the pair distance distribution function derived for the lowest protein concentration (Appendix, figure 6.2 on page 158) indicated that the protein adopts likely an elongated shape, as judged based on the slightly asymmetric shape and the large maximum diameter, but did not display prominent features. Meaningful modeling of the Mkln$_{1-735}$ oligomer would require prior knowledge on substructures and spatial restraints, which were partially available but not to the extent necessary. Taken together, the problems in the SAXS analysis of Mkln$_{157-735}$ and Mkln$_{1-735}$ may be ameliorated by a more elaborate sample preparation to ensure monodispersity and by further experiments providing information on intra- and intermolecular distances in the tetramer (as will be discussed in section 3.3.4).

10 Å

GASBOR model
for Mkln$_{205-735}$

iTasser model for
kelch repeat domain
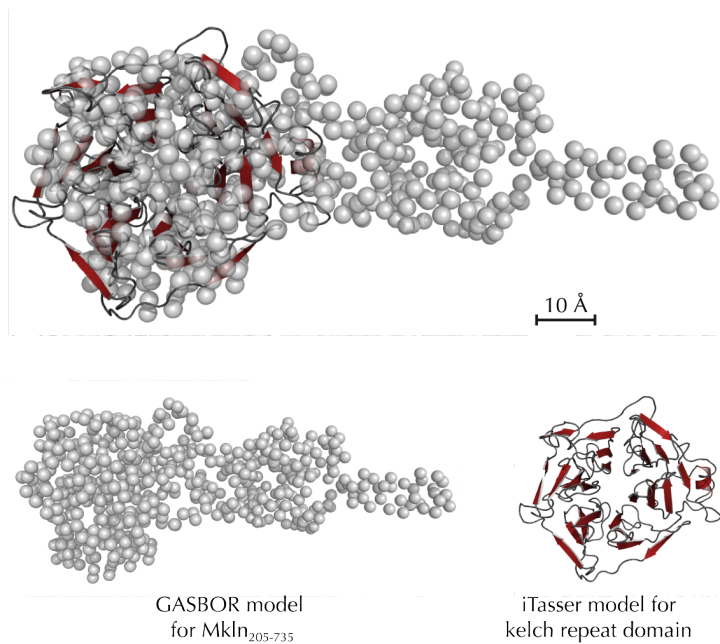
Figure 3.15 – SAXS-derived model of Mkln$_{205-735}$ obtained with *GASBOR*. The model of Mkln$_{205-735}$ represented as spheres (grey) is shown superimposed (*top*) and next to (*bottom*) a homology model for the muskelin kelch repeat domain generated using the iTasser software (Zhang, 2008; Roy et al., 2010) in ribbon representation (red). (M5)

The model derived for Mkln$_{205-735}$ (figure 3.15 and Appendix, figure 6.3 on page 159) had an elongated drumstick-like shape with a rounded, wider head that has a diameter of roughly 40 Å, a more narrow, roughly 50 Å long and 20 - 30 Å wide stalk, with a 28 Å long and 9 Å wide extension. This model could only be compared to homology models of the kelch repeat domain (residues 247 to 625), since no structural information about the other parts contained in Mkln$_{205-735}$, the CTLH motif (residues 205 to 247) and the C-terminal module (residues 625 to 735), were available. Considering the dimensions, the kelch repeat domain closely matches the head of the SAXS-derived model. The remaining parts of the protein, the N-terminal CTLH motif and C-terminal module, would then reside close by in the stem. This fits to the fact, that the N- and C-terminus in the kelch repeat domain fold are in close proximity, and hence connect to the preceding and ensuing domains in close spatial proximity (at the junction between the stalk and the head). The largest predicted secondary structure element in these parts, an α-helix encompassing 21-amino acids corresponding to an approximate length of 32 Å, could easily be accommodated in the stem of the model. The small extension likely corresponds to the flexible C-terminus of the protein, including the last ten amino acids of the protein that are predicted to be disordered, a four amino acid linker and the C-terminal 6xHis-tag. The interpretation of the SAXS model implies that the CTLH motif and the C-terminal module together form a structural unit, and this notion could be exploited in future structural and biochemical studies.

79

3.2.5 Options for future structural characterizations

This idea of a structural unit encompassing the CTLH motif and the C-terminal module as described above could be further explored by creating fusion constructs of the two putative constituents, that is a deletion of the kelch repeat domain (Δ247-625). For several reasons, this type of construct would be an attractive candidate for structural characterization. Above all, it would encompass the very two parts of muskelin for which no structural information is available, since no structure of a homologous CTLH motif has so far been solved and the C-terminal module is unique to muskelin. As they seem to form a coherent unit, constructs that do not disrupt this unit are more likely to be stable and natively folded than constructs that contain the two parts in isolation. Moreover, extending the constructs by the LisH motif is promising to provide valuable insights into the architecture of the LisH motif, which was observed in an unusual conformation in the crystal structure of $Mkln_{12-205}$. This is underscored by the observation that the C-terminal module seems to stabilize the LisH-dimerization, as revealed in the analysis of muskelin oligomerization (section 3.3.1 on page 85). Both, a construct containing only the fused CTLH motif and C-terminal module and a construct that in addition includes the LisH motif would be of a molecular weight (18 kDa and 24 kDa, respectively) that is well within the range suitable for NMR analysis, hence could be characterized in detail without the need for crystallization. Finally, omitting the kelch repeat domain likely removes a major factor interfering with crystallization. The generic architecture of kelch repeat domains features a β-propeller with long, flexible loops extending from one face (see Introduction, section 1.2.4), and, as to be expected, the kelch repeat domain in muskelin contains the majority of residues that are predicted to be disordered (see figure 3.2 on page 58). Additionally, when exploring the possibility to improve the crystallization of muskelin by surface entropy reduction, three out of four clusters that are proposed for mutation[4] reside in the kelch repeat domain. Taken together, this underscores that the kelch repeat domain might be a major source for structural variations and thereby microheterogeneity in muskelin.

Notwithstanding the obvious obstacles in crystallization of the kelch repeat domain, a high-resolution structure of it would be of high interest, particularly with respect to its putative function as interaction partner binding module in muskelin (see Introduction, section 1.2.4). This, on the other hand, opens one putative way to overcome the difficulties in crystallization by finding a suitable ligand to the kelch repeat domain for co-crystallization. The rationale here is to lock the flexible loops into a more

---

[4] Cluster proposed for mutation by the *SERp Server*: E318-K320 in between the 1st and 2nd kelch repeat, K537-K539 in the β2-β3 loop of the 5th kelch repeat, E578-E580 in between the 5th and 6th kelch repeatand E645-K647 in a predicted helix in the C-terminal module.

rigid and homogeneous conformation via the interactions with the ligand. This approach has been tested for combinations of the constructs $Mkln_{1\text{-}156}$ and $Mkln_{205\text{-}735}$, which engage in the head-to-tail interaction (see section 3.3.3 on page 89), but has not succeeded so far. The characterization of the interaction of muskelin and the $GABA_A$ receptor (presented in section 3.4 on page 116) brings up a further potential crystallization partner that moreover would provide valuable insights in a functionally relevant interaction interface. First attempts to cocrystallize muskelin with the full-length intracellular loop of the $\alpha 1$ subunit did not succeed, but were not optimized (in particular with respect to the ligand as the intracellular domain is largely disordered and hence an additional source of heterogeneity) and surely did not tap the full potential of this approach. As discussed in section 3.4, the binding site for muskelin in the $GABA_A$ receptor seems to encompass a short α-helical stretch. This is a promising starting point to design a corresponding peptide ligand that binds with reasonable affinity while exhibiting a rather defined fold and not introducing additional structural heterogeneity. Alternatively to natural binding partners of the protein, a specifically engineered high-affinity ligand could help to reduce the conformational freedom of the kelch repeat domain and additionally provide a surface for the formation of crystal contacts, thereby acting as crystallization chaperon (Bukowska and Grütter, 2013). However, this approach does not provide a simple ready-to-use solution, but rather is an elaborate task that clearly went beyond the scope of this thesis.

An alternative approach to reduce microheterogeneity could be the modification of the protein. One option is to replace or partially delete the regions that are predicted to be disordered. However, this carries the risk of compromising either the stability of the fold or interfering with ligand binding (and thereby loosing information about the ligand binding interface). Another option is the mutation of clusters of amino acids that are predicted to exhibit high conformational entropy, of which four are predicted for muskelin[4]. Replacing the amino acids in these clusters by alanines generates patches with low conformational entropy and thereby conducive to the formation of crystal contacts (Derewenda and Vekilov, 2005). Albeit representing a smaller encroachment on the native structure of the protein, potential detrimental effects on ligand binding still should be carefully considered. Apart from genetic engineering, the chemical modification of surface residues, with one commonly used strategy being the methylation of lysines (Walter et al., 2006), enable a direct manipulation of the protein surface. All of these modifications enhance the chance of crystallization, naturally at the cost of protein solubility. For that reason, an attractive candidate for these experiments is the $Mkln_{247\text{-}625}$ construct that encompasses the kelch repeat domain only. It proved to be particularly soluble[5]and hence might tolerate these manipulations.

---

[5] The highest protein concentration tested was $50\,^{\text{mg}}/_{\text{ml}}$ (≈ 1 mM), and showed no aggregation.

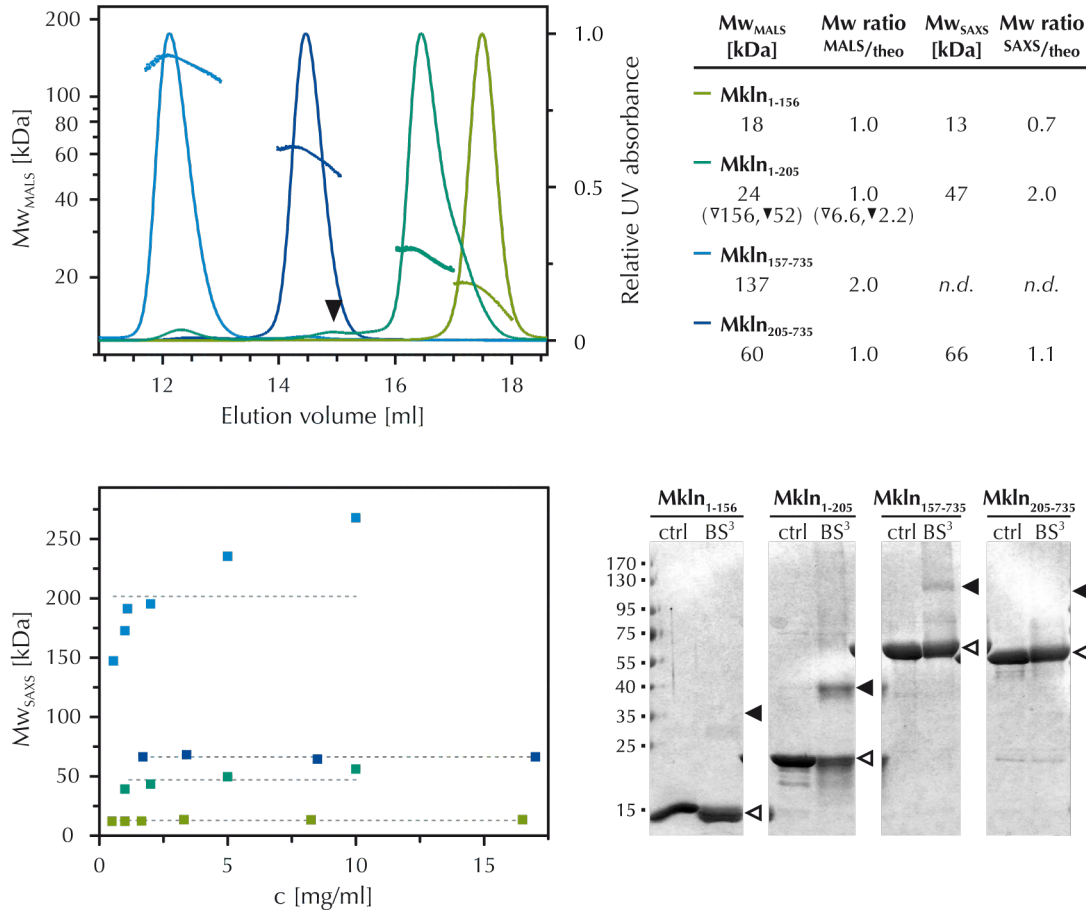## 3.3 Analyses of muskelin oligomerization

An important basic characteristic of muskelin, which has already been described in earlier studies (Prag et al., 2004, 2007; Kiedzierska et al., 2008), is its ability to form oligomers. The oligomerization was suggested to be an important factor for muskelin function and a likely target for its regulation. It was shown that the N-terminal discoidin domain is able to bind to the C-terminal portion of muskelin in a head-to-tail interaction (Prag et al., 2004). Additionally, the crystal structure of $Mkln_{12\text{-}205}$ as described in the previous section provides strong evidence that the LisH motif in muskelin can serve as a dimerization interface. Although being a general feature of LisH motifs, it has so far neither been discussed nor investigated in the context of muskelin oligomerization. This implies that muskelin oligomerization involves several of its domains, with, at a minimum, two possible underlying interactions: the head-to-tail interaction and the LisH-dimerization.

To obtain a deeper understanding of muskelin oligomerization, the following questions were addressed in the course of this work: (1) the putative involvement of further domains was probed; (2) previous results, both from earlier studies and as obtained in this work, were confirmed and the involved interactions were characterized in more detail; (3) it was investigated whether the head-to-tail interaction was an intra- or an intermolecular interaction and (4) whether the different binding events (the head-to-tail interaction and the LisH-mediated dimerization) are independent of each other. For the latter aim, it was intended to identify mutations in separate analyses of the LisH-dimerization and the head-to-tail interaction that were suitable to specifically abolish the respective interaction. These could subsequently be used for the analysis of the individual contribution of the interactions to the oligomerization of the full-length protein.

3.3.1 Determination of domains relevant for oligomerization

The first step in the analysis of muskelin oligomerization was to identify domains that are capable of self-association and thus potentially contribute to oligomerization. Therefore, the oligomeric state was assessed for a selected set of shortened constructs (see also section 3.1.3 and figure 3.2). Size exclusion chromatography (SEC) in conjunction with multi-angle light scattering (MALS) was used to assess the molecular weight and thereby determine the oligomeric state (figure 3.16, top). $Mkln_{1\text{-}156}$ and $Mkln_{205\text{-}735}$ both eluted in a single peak and the determined molecular weights (18 kDa and 60 kDa) corresponded to those calculated for the respective monomers (18 kDa and 61 kDa). In contrast, $Mkln_{157\text{-}735}$ not only eluted significantly earlier than the similarly sized $Mkln_{205\text{-}735}$ but also the experimentally derived molecular

	Mw$_{MALS}$ [kDa]	Mw ratio ^{MALS}/$_{theo}$	Mw$_{SAXS}$ [kDa]	Mw ratio ^{SAXS}/$_{theo}$
Mkln$_{1-156}$				
	18	1.0	13	0.7
Mkln$_{1-205}$				
	24 ($\triangledown$156,▼52)	1.0 ($\triangledown$6.6,▼2.2)	47	2.0
Mkln$_{157-735}$				
	137	2.0	*n.d.*	*n.d.*
Mkln$_{205-735}$				
	60	1.0	66	1.1

Figure 3.16 – Analysis of the oligomeric state for shortened muskelin constructs by SEC-MALS (*top left*), SAXS (*bottom left*) and chemical crosslinking (*bottom right*). *Top left:* The molar masses obtained by MALS plotted as a function of elution volume are overlaid with the normalized SEC A$_{280}$ absorbance profiles. For Mkln$_{1-205}$, two small peaks preceding the major peak were considered in the analysis (values in the table in brackets). Column used: Superdex™ 75 10/300 GL. *Bottom left:* The molecular weights estimated based on the Porod volumes obtained by analyses of SAXS data are plotted versus the protein concentration. The mean value of the determined molecular weights is indicated as dashed line for each shortened construct. *Top right:* The table summarizes the molecular weights determined by SEC-MALS and SAXS (Mw$_{MALS/SAXS}$) and the respective ratio to the theoretical molecular weight ($^{MALS/SAXS}$/$_{theo}$). *Bottom right:* a Coomassie-stained gel of BS^3 crosslinking products of shortened muskelin constructs separated by SDS-PAGE is shown with arrowheads indicating the expected position of the monomers (open) and dimers (filled). (SEC-MALS and crosslinking: M1, M2, M52, M53. SAXS analyses: M1, M2, M4, M5.)

weight of 137 kDa closely matched that calculated for a $Mkln_{157-735}$ dimer (135 kDa). For $Mkln_{1-205}$, a major peak preceded by two smaller peaks was observed. Based on the determined molecular weight of 24 kDa, the dominant species in the main peak can clearly be assigned as being monomeric (calculated molecular weight of 24 kDa). While the molecular weight of the particles in the peak eluting earliest varied considerably in repetitions of the experiment and could not readily be related to a distinct oligomeric state, the mass of the particles in the second peak (52 kDa) was repeatedly found in reasonable agreement with the molecular weight of a $Mkln_{1-205}$ dimer. Based on the integrated UV-absorbance of the peak, the mean concentrations were calculated to be 5 µM and 0.15 µM in the monomer and dimer peak of the present experiment, respectively. This indicates a very weak dimerization (with a K_D in the higher micromolar range) of $Mkln_{1-205}$, which is shifted towards the monomeric form at the concentrations used in the MALS experiments.

As alternative method, SAXS data (see section 3.2.4 on page 78) were analyzed with respect to the oligomeric state of the proteins in solution. The molecular weights of the proteins was estimated based on the determined Porod volumes for several protein concentrations. For $Mkln_{1-156}$ and $Mkln_{205-735}$ the derived molecular weights were independent of the protein concentration, and were in reasonable agreement with the size of the monomer. Both $Mkln_{1-205}$ and $Mkln_{157-735}$ showed a concentration dependency of the determined molecular weight, which was only slight for $Mkln_{1-205}$ but very strong for $Mkln_{157-735}$. The mean value of the determined molecular weights for $Mkln_{1-205}$ (47 kDa) closely matched that of the dimer (48 kDa). No mean value was calculated for $Mkln_{157-735}$, as the analyses were impeded by a potential aggregation of the sample, which is reflected by the strong concentration dependency.

To further probe the oligomeric state, chemical crosslinking with the amine-sensitive crosslinker BS^3 and subsequent analysis of the crosslinked products by SDS-PAGE was applied. Consistent with the results obtained by SEC-MALS and SAXS, only the monomer was detected for $Mkln_{1-156}$ and $Mkln_{205-735}$ after chemical crosslinking, while for both $Mkln_{157-735}$ and $Mkln_{1-205}$ a band for the crosslinked dimeric form could be detected (figure 3.16 on the previous page, bottom left). Especially $Mkln_{1-205}$, which was found to be predominantly monomeric in SEC-MALS, could be crosslinked very efficiently into the dimeric form.

Taken together, these results clearly showed that constructs including the LisH motif were able to dimerize while constructs without were not. This observation confirms that the LisH motif can serve as dimerization motif in muskelin. The pronounced difference in the strength of dimerization via the LisH motif in the presence and absence of the C-terminal domains demonstrated by the SEC-MALS and SAXS experiments suggested a stabilization of LisH-dimerization by the C-terminal part of muskelin. This is consistent with findings for other LisH-containing proteins, in which the LisH motif
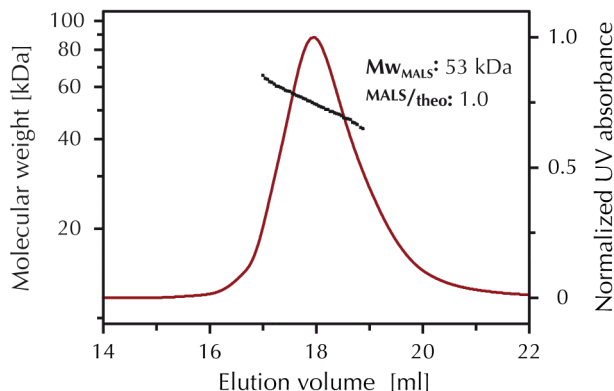
Figure 3.17 – SEC-MALS analysis of Mkln$_{157-625}$. The molar masses obtained by MALS as a function of elution volume are overlaid with the SEC A$_{280}$ profile. The obtained molecular weight (Mw) and the ratio to the theoretical molecular weight (^{MALS}/$_{theo}$) are indicated. Column used: Superose^{TM} 6 10/300 GL. (M54)

is the primary dimerization interface and is necessary, but not sufficient to mediate dimerization (Mikolajka et al., 2006; Kim et al., 2004).

To further test the hypothesis of a stabilization of the LisH-mediated dimerization by a region in the C-terminal part of muskelin, a construct that includes the LisH motif, but lacks the C-terminal module (Mkln$_{157-625}$) was analyzed using SEC-MALS (figure 3.17). Mkln$_{157-625}$ eluted as a single peak exhibiting a molecular weight that precisely matches that of a monomer. This is in contrast to what was observed for Mkln$_{157-735}$ (figure 3.16) and corroborates a stabilizing contribution of the C-terminus to the LisH-mediated dimerization. Notably, no trace amounts of dimer were found for Mkln$_{157-625}$, unlike for Mkln$_{1-205}$ (figure 3.16), indicating that the dimerization via the LisH motif is even weaker or completely prevented in the Mkln$_{157-625}$ construct.

3.3.2 Dimerization via the LisH motif

The analysis of the self-association capability of the individual domains underscored that the LisH motif in muskelin is able to dimerize, confirming the observations in the Mkln$_{12-205}$ crystal structure. Subsequently, the interface found in the crystal structure was investigated in more detail by mutational analyses with two major aims: first, to give an estimate of the relevance of individual residues to LisH-dimerization by introducing single substitutions, and second, to find a mutation to specifically and efficiently prevent LisH-dimerization for subsequent analyses of the contribution of LisH-dimerization to the assembly of the oligomer formed by full-length muskelin. For the former aim, the residues Arg172, Cys180, Phe184, Arg185 and Asn188 were mutated individually, for the latter, a double mutation was designed, exchanging a major contributor to the LisH-mediated dimer, Phe184, to a charged residue, a glutamate, to disrupt the hydrophobic core, and the conserved Leu196 in the second helix to a glutamine, to break up a second point in the four helix bundle.

85
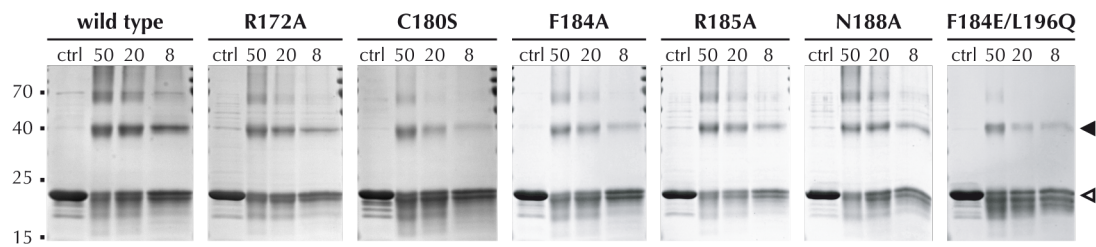
Figure 3.18 – Crosslinking of LisH interface mutations. Coomassie stained gels of crosslinking products separated by SDS-PAGE. For each variant (indicated on top) crosslinking was performed at three protein concentrations (50 µM, 20 µM, 8 µM), a 20 µM sample without crosslinker was loaded as control (ctrl). Bands of the crosslinked dimer (filled arrowhead) and the monomer (open arrowhead) were densitometrically analyzed, the ratio of the dimer at 20 µM to the monomer in the control is: wild type 75%, R172A 60%, C180S 30%, F184A 35%, R185A 55%, N188A 60% and F184E/L196Q 20%. (M2, M88, M85, M98, M90, M91, M30)

Since chemical crosslinking had proven to be suitable to display the dimerization of $Mkln_{1-205}$ in the preceding analyses, it was used to assess the effect of the mutations (figure 3.18). In order to cover the concentration range in which the most pronounced changes can be observed, crosslinking was performed for several protein concentrations. As before, wild type $Mkln_{1-205}$ was crosslinked very efficiently into its dimeric form (compare figure 3.16) in all protein concentrations tested. At higher protein concentrations additional products of crosslinking, at a position corresponding to a molecular weight of approximately 70 kDa, thus corresponding to the covalent linkage of three $Mkln_{1-205}$ molecules, were observed. Based on the comparison of the dimer fraction for the samples with 20 µM protein concentration, the impairment in dimerization can be put in the following order: N188A ≈ R172 < R185A < F184A < C180S < F184E/L196Q. This is in good agreement with the contributions of the respective residue to the LisH interface judged by the buried surface area and solvation energy effect (figure 3.13 on page 75). Particularly the F184E/L196Q variant was crosslinked into its dimeric form only very inefficiently at protein concentrations of 8 µM and 20 µM, demonstrating the most drastically weakened dimerization of this variant. As this was the intended effect, the F184E/L196Q double mutant was used in subsequent analyses as LisH-dimerization impairing mutation.

The most detrimental single mutation was the exchange of Cys180 to a serine, however, the contribution of Cys180 to LisH-dimerization merits more description. Upon dimerization, the two central Cys180 residues are in a suitable distance to form a disulphide bridge and thereby a covalent bond across the LisH interface. This disulphide bond was observed in a preliminary 2.1 Å structure of $Mkln_{12-205}$, in which both the reduced and the disulphide-linked oxidized states of Cys180 were partially occupied (figure 3.19 on the next page). The absence of the disulphide bond in the final $Mkln_{12-205}$ structure could be due to photoreduction in the higher intensity beam
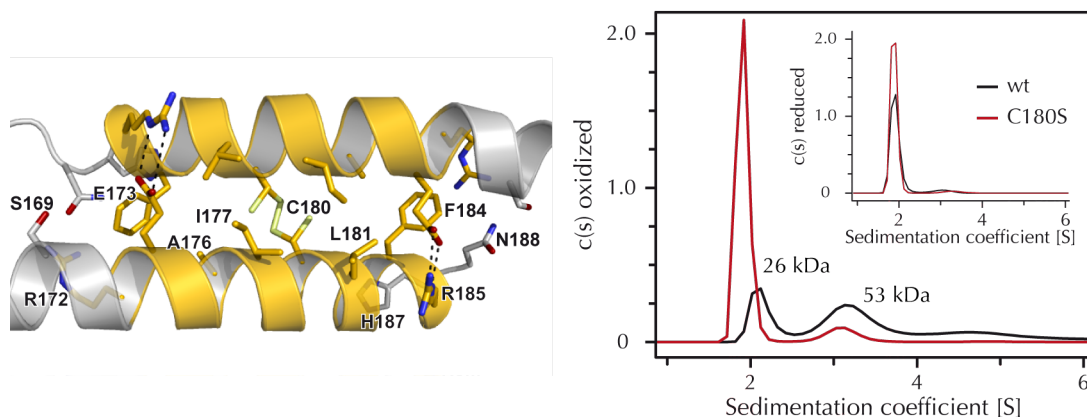
Figure 3.19 – Disulphide bridging of Cys180 across the LisH interface as observed with partial occupancy in a preliminary 2.1 Å Mkln$_{12-205}$ structure, is shown in a close-up view of the dimerization interface (*left*) in cartoon representation with residues contributing to dimerization in stick representation. *Right*: sedimentation coefficient distributions obtained in sedimentation velocity experiments for Mkln$_{1-205}$ wild type and the C180S variant under oxidizing and reducing (inset) conditions. The mean molecular weight calculated for the peaks at 2 S and at 3 S are indicated. Data were fitted as continuos c(s) distribution at 0.95 confidence level, resulting in r.m.s.d values of 0.0048/0.0050 (wild type/C180S variant). (M2, M85)

and higher overall photon dose (3.5×10^{14} photons versus 1.0×10^{14} photons) that the respective crystals were exposed to during data collection.

To further validate this putative disulphide-bridge in the LisH dimer, dimerization of Mkln$_{1-205}$ was observed under reducing and oxidizing conditions, and compared to the C180S variant, which is not able to form this bond. Dimerization was assessed by determining the molecular weight of the proteins by sedimentation velocity analytical ultracentrifugation (SV-AUC) (figure 3.19).

In agreement with earlier observations (see figure 3.16), dimer formation under reducing conditions was very weak and Mkln$_{1-205}$ as well as the C180S variant were observed predominantly with sedimentation coefficients corresponding to the molecular weight of the monomer. This also applies to the C180S variant under oxidizing conditions, while for wild type Mkln$_{1-205}$ a large fraction was found with sedimentation coefficients corresponding to the molecular weight of the dimer. This corroborates that under oxidizing conditions Cys180 can form a disulphide bridge and thereby fix Mkln$_{1-205}$ in its dimeric form. The *in vivo* relevance of this disulphide bridge is questionable, since muskelin as a nucleocytoplasmic protein is natively located in a reducing environment. However, it is consistent with a previous report by Kiedzierska et al. (2008) that stated the occurrence of two disulphide bonds in isolated muskelin based on results obtained in an Ellman's assay.
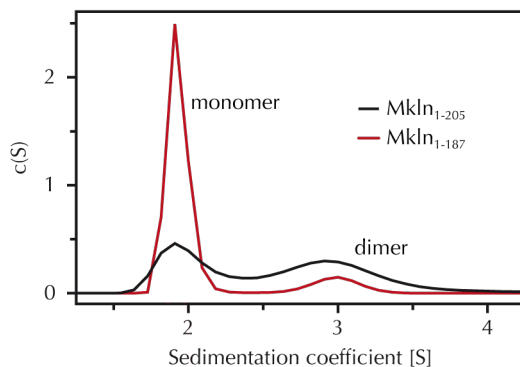
Figure 3.20 – The second helix of the LisH motif contributes to dimerization. Sedimentation coefficient distributions obtained for Mkln$_{1\text{-}205}$ and Mkln$_{1\text{-}187}$ under oxidizing conditions are shown with the peak of the monomer at 2 S and the dimer at 3 S indicated. Data were fitted as continuos c(s) distribution at 0.95 confidence level, resulting in an r.m.s.d of 0.0056. Note that one dimer corresponds to one sedimenting particle, thus at equal abundance of both sedimenting species, actually two thirds of the protein are bound in the dimer. (M2, M58)

The second helix of the LisH motif was not visible in the Mkln$_{12\text{-}205}$ crystal structure, thus no molecular details on possible contributions to the dimerization by residues in the second helix were available. The mutational analyses described above only probed residues in the first helix, with the exception of residue Leu196 in the F184E/L196Q variant. In order to assess the overall contribution by residues in the second helix, dimerization was compared between Mkln$_{1\text{-}205}$ that contains the entire LisH motif and Mkln$_{1\text{-}187}$, which lacks the second predicted helix. In these analyses, the disulphide bond formed by Cys180 was used to enhance dimerization by analyzing the proteins under oxidizing conditions. Dimerization was assessed by measuring the molecular weight distribution in SV-AUC (figure 3.20). As before, Mkln$_{1\text{-}205}$ was found to a large extent in its dimeric form under oxidizing conditions. In contrast, for Mkln$_{1\text{-}187}$ the predominant species under oxidizing conditions was detected with sedimentation coefficients corresponding to the molecular weight of the monomer, and only a small fraction of particles was found with sedimentation coefficients corresponding to the molecular weight of the dimer. The loss of the second helix thus significantly impaired the dimerization of Mkln$_{1\text{-}205}$ via the LisH motif, which demonstrates that residues in the second predicted helix significantly contribute to LisH-dimerization.

Finally, the strength and parameters of the isolated LisH-dimerization, in the absence of putative stabilizing interactions by the C-terminal part of muskelin, were assessed in dilution ITC experiments (figure 3.21). In the analysis of the observed dilution heats for Mkln$_{1\text{-}205}$, the K_D of the dimer was determined to be about 0.58 mM. Although this very weak dimerization is in line with estimates based on the ratio of monomer and dimer concentrations in the experiments assessing the oligomeric state of Mkln$_{1\text{-}205}$ described above[6], it has to be critically analyzed. The measurements and

---

[6] In the SEC-MALS experiments (figure 3.16), the monomer and dimer were found in a concentration of 5 μM and 0.15 μM, respectively, indicating a K_D of about 330 μM. In repeated AUC analyses at 200 μM under reducing conditions (see also figure 3.19) only 25% to 40% of the protein was found to be dimeric, indicating a K_D in the range of 400 - 900 μM.
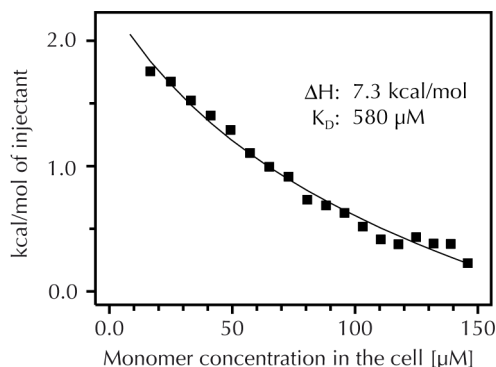
Figure 3.21 – Dissociation of Mkln$_{1\text{-}205}$. Heat release upon dilution measured in ITC for a syringe protein concentration of 851 µM is plotted versus the resulting protein concentration in the cell. (M2)

the analyses were impaired both by the weak binding and the additional aggregation observed using other techniques. Given the estimated K_D, the protein concentration that would be needed to reach substantial amounts of dimer is rather high, and at such high concentrations, aggregation would outweigh dimer formation. More severely, the model used to analyze the ITC experiments assumes a simple dimerization. A putative contribution to the measured dilution heat produced by dissociation of the additional aggregates is not distinguished and distorts the determined K_D. On the other hand, when using lower concentrations to avoid aggregation, the fraction of dissociating protein is small, hernce the heat signal is weak, resulting in rather poor data quality with small errors having a big impact on the determined K_D. The concentration used in the present experiment was chosen with respect to these limitations, yet the determined K_D should not be considered an exact value but rather a rough estimate.

3.3.3 Head-to-tail interaction

In contrast to the neglected role of the LisH motif, the head-to-tail binding of the discoidin domain to a C-terminal portion of muskelin has been investigated in several studies (Prag et al., 2004, 2007). This interaction was first demonstrated in pull-down experiments with different muskelin constructs by Prag et al. (2004), and it was moreover shown that the intact fold of the kelch repeat is a prerequisite for this interaction. In that study, the propensity of muskelin to form small intracellular particles upon overexpression in eukaryotic cells was suggested to reflect muskelin's ability to self-associate and was related to the head-to-tail binding. Based on changes in the formation of these particles, a later study suggested a regulation of the head-to-tail interaction by protein kinase C (PKC) via phosphorylation of two sites in the kelch repeat domain (Prag et al., 2007). Yet, these studies had significant shortcomings: they were not performed using isolated proteins, they did not provide a detailed mapping, biophysical binding parameters were not measured and they did not demonstrate that
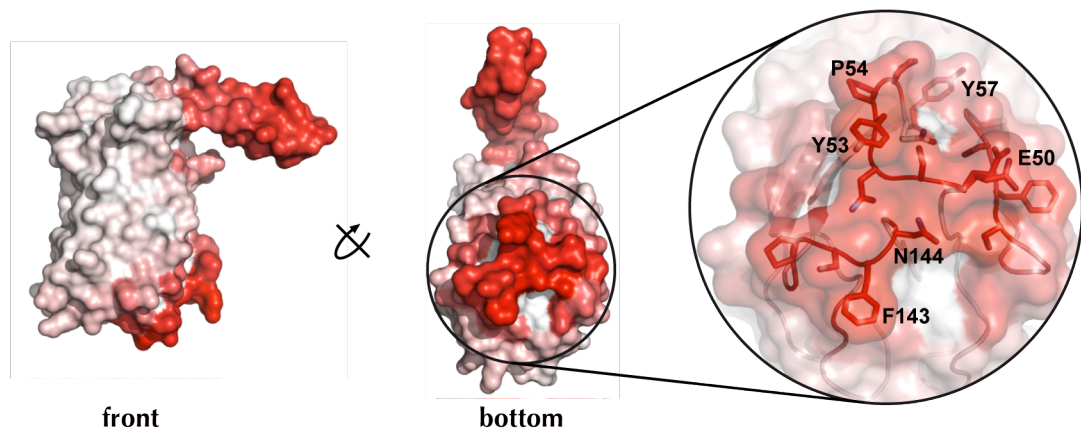
Figure 3.22 – Prediction of protein-protein interaction sites on Mkln₁₂₋₂₀₅ based on a *metaPPISP* analysis. The score attributed by *metaPPISP* is displayed as color (white = 0 to red = 0.8) on the surface of the molecule in front (*left*) and bottom view (*middle*). A close-up view into the predicted interface at the bottom of the discoidin domain is shown with a semi-transparent surface and the sidechains in the participating loops in stick representation. Residues that were chosen for mutation are labeled.

the head-to-tail interaction was needed to assemble the muskelin oligomer *in vitro*. A detailed biochemical analysis was also the prerequisite to find mutations that specifically abolish this binding and subsequently investigate the role of the interaction in muskelin oligomerization and function.

The crystal structure of Mkln$_{12\text{-}205}$ determined within this work provided valuable information about one partner of the head-to-tail interaction, the discoidin domain; most importantly which residues are surface exposed and therefore are accessible for putative interactions. In conjunction with sequence conservation and the chemical properties of the side chains, this can give an indication of probable protein-protein interaction sites. Several methods for their prediction have been developed, among them *cons-PPISP* (Chen and Zhou, 2005), *ProMate* (Neuvirth et al., 2004) and *PINUP* (Liang et al., 2006), which are combined by the *meta-PPISP* server (Qin and Zhou, 2007) to increase the accuracy of the prediction. Subjecting a monomer of Mkln$_{12\text{-}205}$ to the *meta-PPISP* analysis gave a clear result, with only two binding sites predicted; one in the LisH part of the extended α-helix and the other located at the bottom of the discoidin domain β-barrel (figure 3.22). Since the LisH motif had already been identified in this work as a dimerization interface, the positive identification was to be expected and by acting as a positive control corroborated the validity of the prediction.

The second site at the bottom of the discoidin domain is mainly formed by residues in two portions of the β1-β2 loop (Trp23 to Tyr30 and Glu50 to Pro55) and the β7-β8 loop (Trp139 to Asn144). This region overlaps with the generic binding site of

variant	T_m [°C]
wt	50
E50A	53
Y53A	52
P54E	50
Y57E	43
F143A	54
N144R	47
N144W	45
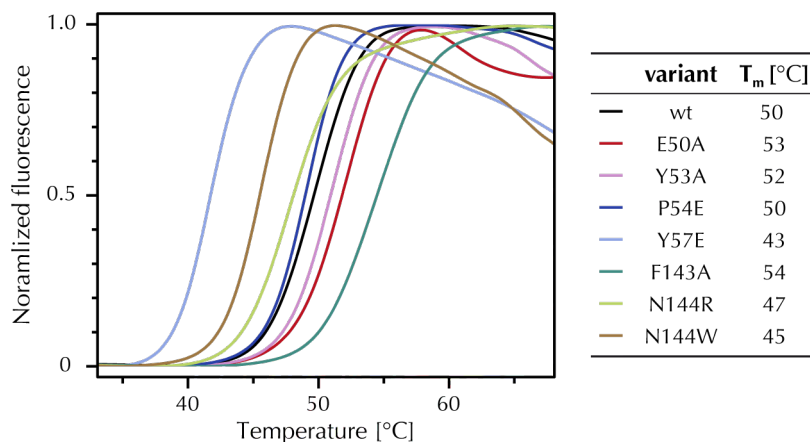
Figure 3.23 – The thermal stability of $Mkln_{1-156}$ variants was assessed in a ThermoFluor assay. The normalized fluorescence of SYPRO® orange is plotted versus the temperature. Curves represent mean values of measurements performed in triplicate. (M1, M60, M61, M62, M63, M64, M65, M66)

discoidin domains. In order to test this site as a putative interface for the head-to-tail interaction, mutations were introduced at several points. Thermal stability and CD spectra of mutated variants were compared to the wild type to ensure that the introduced mutations did not affect the fold and stability of the discoidin domain.

The thermal stability was assessed in a ThermoFluor (Ericsson et al., 2006) experiment (figure 3.23). All variants showed a clear transition in fluorescence as expected for the unfolding of a protein, indicating that all were properly folded in the first place. Overall, the melting temperatures varied by 9 °C, ranging from a slight stabilization, with the most stable protein being the F143A variant, to a marked destabilization, most pronounced for the Y57E variant, when compared to the wild type. CD spectra were recorded for the wild type, the stabilized E50A variant, and the most destabilized variant, Y57E. No significant differences could be detected. Thus, it was concluded that all variants were properly folded and artifacts due to folding defects in subsequent analyses were highly unlikely.

Next, the effect of the mutations on the head-to-tail binding was assessed by determining the binding of the variants to the C-terminal part of muskelin in analytical SEC and ITC analyses. In analytical SEC, the elution profiles of $Mkln_{205-735}$ and each $Mkln_{1-156}$ variant were compared between the single proteins and the combined samples (figure 3.24). All single proteins were detected as clear peaks in the chromatogram; $Mkln_{205-735}$ (with a small shoulder) eluted at 9.4 ml and the discoidin domain variants at 12 - 12.5 ml. These elution volumes were in reasonable agreement with the expected elution volumes of the proteins based on the respective molecular weight. Upon combination of the proteins, the changes in the elution profiles were evaluated to judge complex formation.

For the wild type discoidin domain, the observed changes were indicative of complex formation: the first peak of the chromatogram was shifted to a smaller elution volume
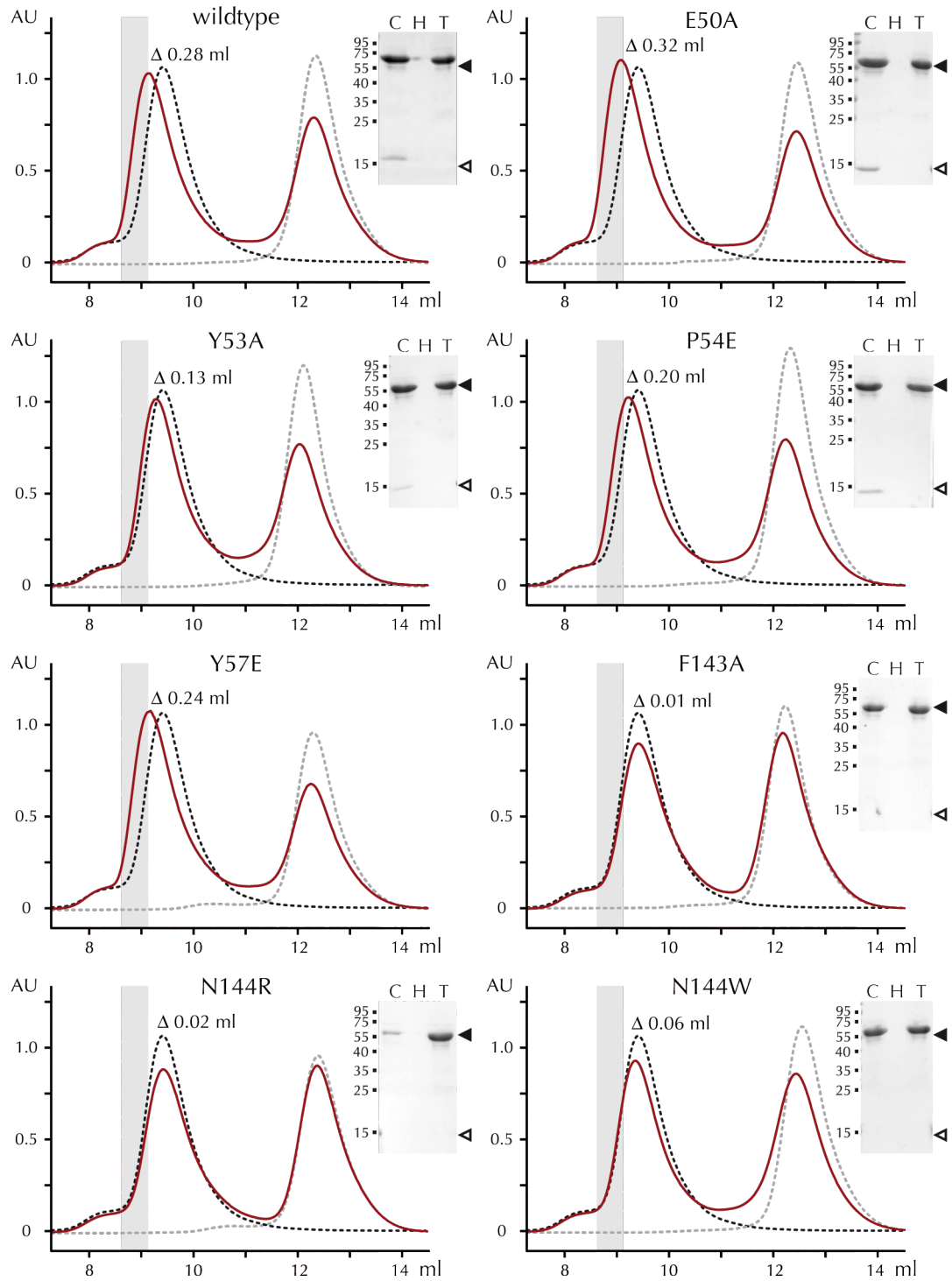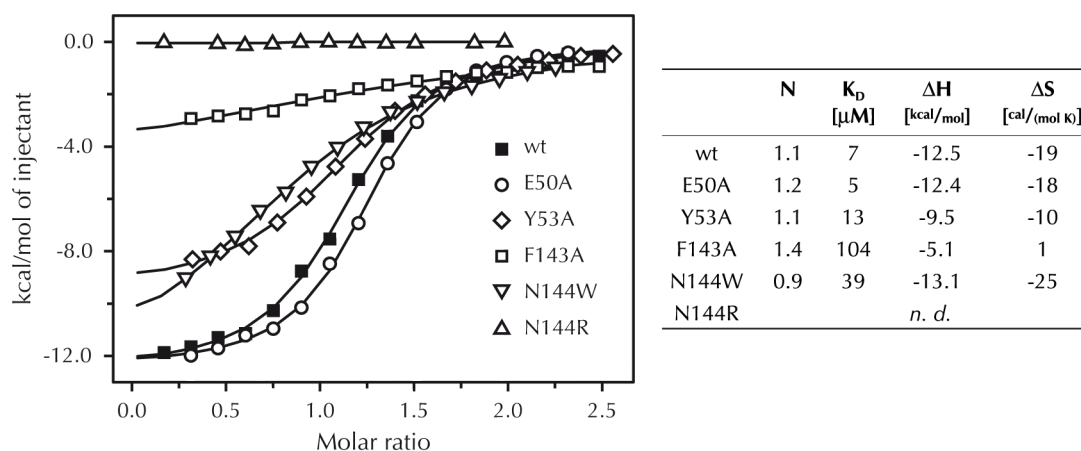
as expected for a complex with an increased size; at the same time, the height of the second peak was diminished, indicating that the discoidin domain was partially recruited to the complex. In SDS-PAGE analysis, the discoidin domain was detected in the early fractions of the first peak, while it did not elute in those fractions when subjected to SEC alone. Using the shift of the first peak and the band-intensity of the discoidin domain in SDS-PAGE analysis of the early fractions, the effect of the introduced mutations was judged as follows: the E50A variant showed an increased complex formation; the Y57E variant was very similar to the wild type; P54E showed a slightly decreased and Y53A a more significantly decreased complex formation; for the N144W variant no discoidin domain was detectable in the early fractions and the shift of the first peak was very small; finally, for both the F143A and the N144R variant, no complex formation could be detected.

The results of analytical SEC were complemented for a subset of the variants by measuring the binding in ITC experiments titrating $Mkln_{205-735}$ with the respective $Mkln_{1-156}$ variant (figure 3.25). For the wild type discoidin domain, the analysis revealed a 1:1 binding with an affinity in the low micromolar range, driven by a favorable enthalpic and an unfavorable entropic contribution. These results were confirmed in three further independent measurements, giving a mean K_D of (5.5 ± 1.5) μM, with a stoichiometry of 0.9 ± 0.2, an enthalpy change (ΔH) of (-13.4 ± 2.6) kcal/mol and a change in entropy (ΔS) of (-21.5 ± 8.7) cal/(mol K). In a recent study, binding parameters were determined for very similar constructs, $Mkln_{5-174}$ to $Mkln_{208-735}$ (Kim et al., 2014). While the reported enthalpic and entropic contributions are in good agreement (-11.0 kcal/mol and -16 cal/(mol K)) with the findings described above, the reported K_D of 10 μM was slightly higher, however, given that only single measurements were presented, this deviation is not significant.

The effect of the introduced mutations on the affinity of the binding ranged from no effect for the E50A mutation, over a weakened binding for the Y53A, N144W and F143A mutation up to a reduction of the affinity to a level that was not detectable anymore for the N144R mutation. A comparison of the associated thermodynamic

Figure 3.24 *(facing page)* – **Analytical SEC of Mkln₂₀₅₋₇₃₅ and Mkln₁₋₁₅₆ variants.** For each mutant, the A_{280} chromatograms (UV absorption in absorption units (AU) plotted versus the elution volume in ml) of the isolated $Mkln_{205-735}$ (dashed black line) and the isolated $Mkln_{1-156}$ variants (dashed grey line) are overlaid with the chromatograms of the combined samples (red line). The difference (Δ) in elution volume between the peak of $Mkln_{205-735}$ and the respective peak for the combined samples is indicated. The areas of the fractions analyzed by SDS-PAGE (shown as inserts, with the exception of the Y57E variant, for which samples were lost) are shaded in grey. Fractions of the combined samples (C) and the isolated proteins (head - $Mkln_{1-156}$: H, tail - $Mkln_{205-735}$: T) were analyzed, arrowheads indicate the expected positions in the gels for $Mkln_{205-735}$ (filled) and $Mkln_{1-156}$ variants (open). Column used: Superdex™ 75 10/300 GL (M5 and row by row: M1, M60, M61, M62, M63, M64, M65, M66)

	N	K_D [µM]	ΔH [kcal/mol]	ΔS [cal/(mol K)]
wt	1.1	7	-12.5	-19
E50A	1.2	5	-12.4	-18
Y53A	1.1	13	-9.5	-10
F143A	1.4	104	-5.1	1
N144W	0.9	39	-13.1	-25
N144R		*n. d.*		

Figure 3.25 – ITC analysis of the mutations in the predicted head-to-tail interaction interface of the discoidin domain. The released heat measured upon titration of $Mkln_{205-735}$ with the respective $Mkln_{1-156}$ variant, wild type (wt) or indicated mutation, is plotted against the molar ratio of the binding partners. The table summarizes the binding parameters derived by fitting the data to a one-site-binding model. (Titrants M1, M60, M61, M64, M66, Ligand M5)

quantities, ΔH and ΔS, to the wild type revealed the following: the E50A not only displayed a very similar affinity but also no changes in ΔH and ΔS; for the Y53A variant the decrease in affinity was driven by a loss in the enthalpic contribution, while the entropic penalty was reduced; the loss in affinity of the N144W variant was driven by an increased unfavorable entropy with no significant change in enthalpy; for the F143A variant the low c value (see Methods, section 2.2.3) of the measurement interferes with deriving accurate thermodynamic information, and changes are likely not informative.

Overall, the results of the investigations on the mutations in the putative binding site using analytical SEC and ITC were consistent. Based on the combined results, the binding affinities can be put in the following order:

wt ≈ E50A > Y57E > P54E > Y53A > N144W > F143A > N144R.

This order does not reflect the thermal stability of the domain (see figure 3.23):

F143A > E50A > Y53A > wt ≈ P54E > N144R > N144W > Y57E.

Thus, the effect of each mutation on the binding is unlikely to be caused by a decrease in protein stability, but rather by a disturbed binding interface. In their entirety, these results confirmed the predicted interaction interface found by *meta-PPISP* (figure 3.22 on page 90). As this identified interaction site overlaps with the generic binding site observed for discoidin domains and these protein domains typically bind one ligand specifically (Kiedzierska et al., 2007), this provides evidence that the natural ligand to the discoidin domain is the C-terminus of muskelin. Muskelin is thereby the first
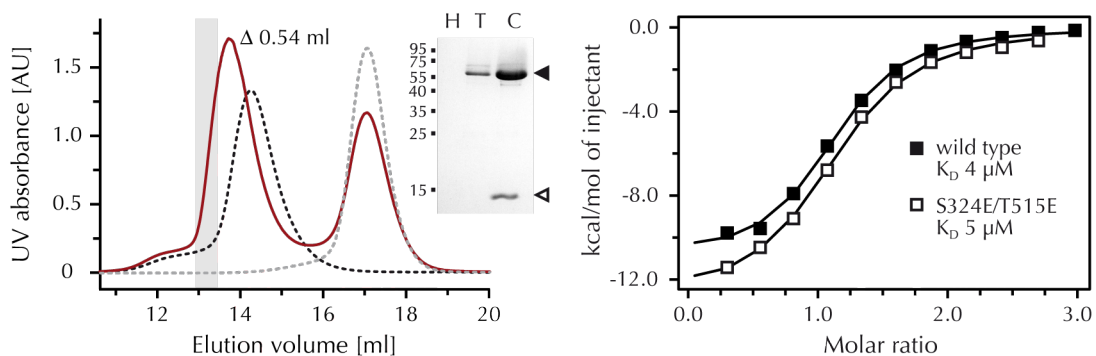
Figure 3.26 – Effect of the phosphomimetic S324E/T515E mutation on the head-to-tail interaction analyzed by analytical SEC and ITC. *Left:* The A_{280} chromatogram (UV absorption in absorption units (AU) plotted versus the elution volume in ml) of the isolated $Mkln_{1-156}$ (dashed grey line) and $Mkln_{205-735}$ S324E/T515E variants (dashed black line) are overlaid with the UV chromatogram of the combined samples (red line). The difference (Δ) in elution volume between the peak of $Mkln_{205-735}$ S324E/T515E and the first peak for the combined samples is indicated. The area of the fractions analyzed in the SDS-PAGE (shown as insert) is shaded in grey. Fractions of the combined samples (C) and the isolated proteins (head - $Mkln_{1-156}$: H, tail - $Mkln_{205-735}$: T) were analyzed, arrowheads indicate the expected positions in the gel for $Mkln_{205-735}$ S324E/T515E (filled) and $Mkln_{1-156}$ (open). Column used: Superdex™ 200 10/300 GL. *Right:* Released heat measured upon titration of either $Mkln_{205-735}$ wild type or the S324E/T515E variant with $Mkln_{1-156}$ is plotted against the molar ratio of the binding partners, with the determined K_D values given. Results of fit for wild type: K_D 4 μM, stoichiometry 1.1, ΔH -11 $^{kcal}/_{mol}$, ΔS -13 $^{cal}/_{(mol\ K)}$; and for S324E/T515E: K_D 5 μM, stoichiometry 1.1, ΔH -13 $^{kcal}/_{mol}$, ΔS -20 $^{cal}/_{(mol\ K)}$. (M1, M49, M5)

example of a discoidin domain contributing to oligomerization via a direct binding with its canonical binding site.

For a comprehensive biochemical characterization, the mapping of the head-to-tail binding on the tail, the C-terminal part of muskelin, remained to be done. The interface in the tail was especially interesting as it had been suggested to be the target of PKC, thus regulating this interaction (Prag et al., 2007). In this study, the phosphorylation of Ser324 and Thr515 was proposed to impair the head-to-tail binding. Notably, both residues are predicted to reside in the loop regions of the generic binding sites in kelch repeat domains. Yet their role was investigated only by using alanine substitutions that prevented the phosphorylation, but not by mutations that mimic the phosphorylation. This limitation was chosen as starting point to validate the findings of the aforementioned study. Phosphorylation of the proposed residues was mimicked by substitution with the charged amino acid glutamate, and it was subsequently tested wether these phosphomimetic mutations affected the head-to-tail interaction in analytical SEC and ITC experiments (figure 3.26).

In analytical SEC, the marked shift to a lower elution volume of the first peak in the elution profile of the combined samples and the recruitment of the discoidin domain to

the fractions of this peak both demonstrated that the Mkln$_{205-735}$ S342E/T515E variant was able to efficiently bind to the discoidin domain. In addition, the parameters of binding derived in ITC analysis were equal within the limits of accuracy between the wild type and the S342E/T515E variant. Thus, the phosphomimetic mutations did not influence the head-to-tail binding in the present experiments. Either the introduced mutations did not sufficiently mimic the effect of phosphorylation, or the changes observed in the previous study (Prag et al., 2007) were not mediated via a direct impairment of the head-to-tail binding.

As a role for Ser342 and Thr515 could not be confirmed, no information about determinants of the head-to-tail binding on the tail was available, and mapping needed to start from scratch. The first step was to validate the previous suggestion that the interaction was mediated by the kelch repeat domain. The constructs used in the binding studies by Prag et al. (2004) either contained the C-terminal part (the protein comprised residues 244 to 735) or included mutations, which were designed to disrupt the fold of the kelch repeat domain, and it cannot be excluded that the disruption of the fold also influences the other parts of the protein. In addition, a possible influence of the dimerization via the LisH motif had not been considered yet which was also addressed in the following preliminary mapping studies.

To probe which C-terminal parts of the protein (the kelch repeat domain or the C-terminal region) are needed and whether LisH-dimerization would affect the head-to-tail interaction, binding parameters of shortened constructs either with or without the LisH motif (starting either with residue Asp157 or Glu205) and with or without the C-terminal module (ending either at residue Leu735 or Arg625) were compared in ITC analyses (figure 3.27 on the next page). The measured binding isotherms, as well as the derived parameters of binding were very similar for all tested constructs and consistent with the previous measurements of Mkln$_{205-735}$ (compare figure 3.25). These results demonstrate that the C-terminal region does not contribute to the head-to-tail binding, and that this interaction is not dependent on LisH-dimerization.

These results are contradictory to the findings of the recent publication by Kim et al. (2014), in which binding parameters were determined for very similar combinations of muskelin constructs[7]. Based on their measurements, the authors stated that the LisH

---

[7] Binding parameters published by Kim et al. (2014) for binding of Mkln$_{5-174}$ to the indicated construct:

	N	K$_D$ [μM]	ΔH [kcal/mol]	ΔS [cal/(mol K)]
Mkln$_{151-735}$	1.0	10	-12	-18
Mkln$_{151-662}$	0.8	2	-4	11
Mkln$_{208-735}$	1.4	30	-11	-16
Mkln$_{208-662}$	1.0	19	-14	-26

|	N	K_D [µM]	ΔH [kcal/mol]	ΔS [cal/(mol K)]
$Mkln_{157-735}$	0.8	6	-13.8	-23.1
$Mkln_{157-625}$	0.7	5	-11	-13.3
$Mkln_{205-735}$	0.7	5	-13.2	-20.8
$Mkln_{205-625}$	0.9	5	-12.9	-19.6
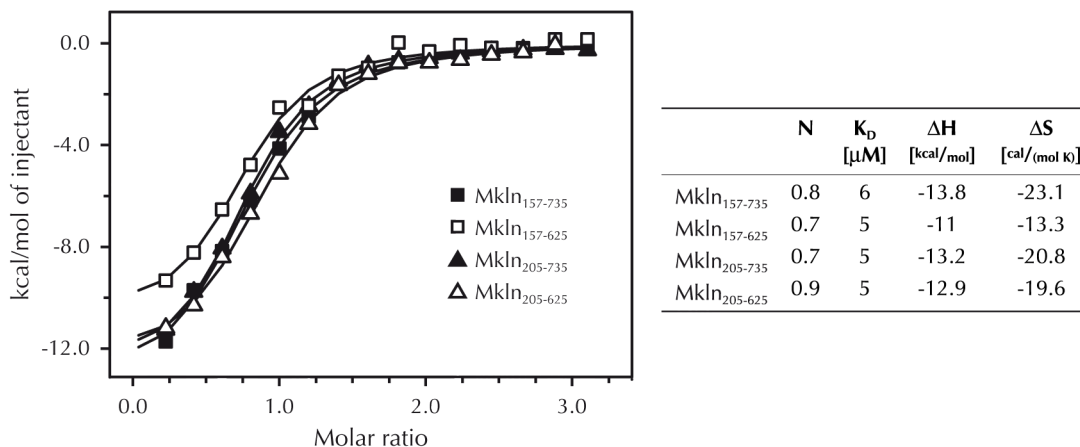
Figure 3.27 – ITC analysis of respective domain contributions in the C-terminal part to the head-to-tail interaction. The released heat measured upon titration of either $Mkln_{157-735}$, $Mkln_{157-625}$, $Mkln_{205-735}$ or $Mkln_{205-625}$ with $Mkln_{1-156}$ is plotted against the molar ratio of the binding partners. The table on the right summarizes the binding parameters derived by fitting a one-site-binding model to the data. (M1, M52, M53, M54, M55)

motif would promote the head-to-tail interaction, while the C-terminal region had a detrimental effect on the binding. Yet, the overall affinities as well as the enthalpic and entropic contributions are in reasonable agreement to the values obtained within this work, with the exception of the $Mkln_{151-662}$ construct, and the observed differences are unlikely to be significant, particularly since only single measurements were presented in the study. Nevertheless, the discrepancy of the determined entropic contribution, which was found to be unfavorable for $Mkln_{157-625}$ in this work and favorable for $Mkln_{151-662}$ in the study by Kim et al. (2014), cannot be explained based on the data obtained so far.

The experiments outlined on the previous pages confirmed that the binding partner of the discoidin domain in the head-to-tail interaction is the kelch repeat domain. To further refine the mapping, it was investigated wether the kelch repeat domain, just as the discoidin domain, mediated the head-to-tail binding via its generic binding site, which is formed by long loops on one face of the propeller (see introduction section 1.2.4 on page 21)

Lacking a high-resolution crystal structure of muskelin's kelch repeat domain, the precise arrangement and composition of this putative interface was unknown. Therefore, it was intended to start with narrowing down the contributing loops. The generic binding site is formed by the loops between the second and third β-strand within each kelch repeat, the β2-β3-loops, and the loops connecting the last β-strand of one kelch repeat to the first β-strand of the next, the β4-β1'-loops (figure 3.28).
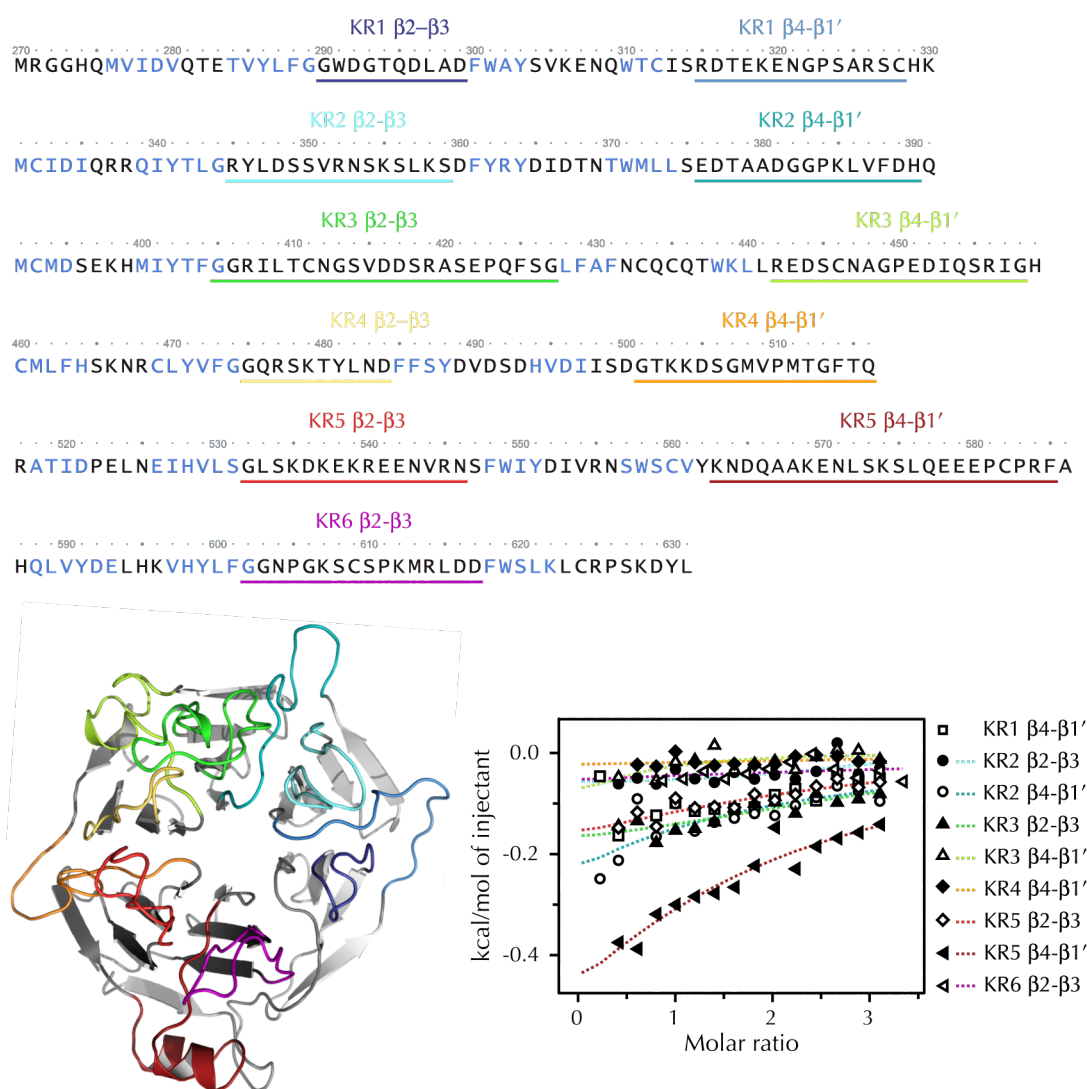
Figure 3.28 – Analysis of contributions of loops in the generic kelch repeat domain binding site to the head-to-tail interaction. *Top*: the sequence of muskelin's kelch repeat domain is shown, grouped by the repeats with residues that belong to the predicted β-strands in the β-sheets colored in blue. The loops that have either been replaced or used as peptides for the analyses are underlined with different colors. *Bottom left*: the corresponding amino acids are colored accordingly in the ribbon representation of a model of the kelch repeat domain obtained with i-TASSER (Zhang, 2008; Roy et al., 2010). *Bottom right*: heat releases upon titration of $Mkln_{1-156}$ with the corresponding peptides measured in ITC. Fits to the data are indicated by dashed lines color-coded accordingly to the rest of the figure. Note that it was not possible to derive reliable binding parameters. For KR1 β2-β3-loop and KR4 β2-β3-loop, no measurements could be obtained, for KR 1β4-β1'-loop no converging fit to the data was possible.

98

A set of substitution-mutations was generated in cooperation with Laura Wamprecht (Wamprecht, 2012), replacing the β2-β3-loops and β4-β1'-loops by three and ten alanine residues, respectively, both as single substitutions of each loop individually and double substitutions of both loops in one kelch repeat. Yet, a comprehensive analysis using this approach failed, as several of the variants were insoluble and could not be purified. Since the substitution of the loops is unlikely to disturb the basic fold of the β-propeller, it is more likely that the deletions either destabilized the loop arrangement by uncovering hydrophobic parts in neighboring loops, or the substitution itself introduced too much hydrophobic surface and thereby provoked aggregation. In future studies, a combination of serine and glycine in the substitutions should be attempted to improve the solubility of the variants.

As alternative approach for the basic identification of determinants in the kelch repeat domain, peptides corresponding to the respective loops were used in additional binding studies. Initial pulldown experiments were conducted with N-terminally biotinylated versions of the peptides coupled to a streptavidin-coated matrix as bait and the discoidin domain ($Mkln_{1-156}$) as prey. These experiments failed, as no traces of bound discoidin domain could be detected (not shown). Similarly, ITC analyses which tested the binding of the peptides to $Mkln_{1-156}$, were inconclusive. The heat signatures upon titration with the peptides were, when measurable at all, very weak, with the strongest reaching only a thirtieth of the signal measured for the interactions of the full domains (figure 3.27 on page 97), and for none of the experiments a converging fit with reasonable parameters could be obtained.

This suggests that either the generic binding site in the kelch repeat domain is not used for binding to the discoidin domain, or the methods used so far were not sensitive enough to detect binding. It is well conceivable that binding of the individual loops is too weak and therefore eluded detection, especially when taking into account that in the known complex structures of kelch repeat domains, several or all blades of the propeller contribute to ligand binding, thus the individual contributions might be small in comparison to the overall binding. Future studies would need to use an alternative method to allow for the detection of these low-affinity binding events. One such possible alternative is to perform displacement ITC experiments, in which the binding of the peptides would be characterized by determining their competitive inhibition of the binding of a stronger ligand, which would be a muskelin construct including the kelch repeat domain in this case.

Yet, it might be worthwhile to consider that the head-to-tail binding happens via an alternative site. An example for such alternative binding sites on a kelch repeat domain can be found in the architecture of the galactose oxidase enzyme (Ito et al., 1994). The enzyme consists of three domains, an N-terminal discoidin domain followed by a kelch domain and a third domain forming a β-sandwich. Although the enzyme is monomeric

and hence all interactions between the domains are intra-molecular, its structure nicely illustrates how the two β-sandwich domains (the discoidin and the C-terminal domain) are bound not via the generic binding site of the kelch repeat domain, but laterally in the case of the discoidin domain and on the opposite side of the in case of the third domain. The identification of an alternative binding site would require a systematic mapping of contributing residues in the whole kelch repeat domain. One feasible approach therefore would be an oligopeptide array with peptides spanning the whole length of the domain, which was for example used to identify the binding site of Lasp1 in the kelch repeat domain of Krp1 (Gray et al., 2009).

In summary, the binding interfaces of the head-to-tail interaction could be narrowed down very precisely to single amino acids in the discoidin domain, while for the binding site on the tail, a coarse mapping of the binding site to the kelch repeat domain was achieved. Moreover, with the N144R mutation, which completely abolished the head-to-tail binding in the experiments presented here, a suitable mutation to be used in the analyses of the role of the head-to-tail binding in muskelin oligomerization and function was identified.

3.3.4 Oligomeric assembly of full-length muskelin

The analyses presented so far on the one hand confirmed and identified two possible interactions in muskelin's discoidin domain and LisH motif, respectively, and on the other hand identified point mutations that specifically interfere with these interactions. These formed the basis for the following experiments, in which the contribution of the interactions to the oligomerization of the full-length protein was investigated. The oligomeric state of the full-length protein was assessed by SEC-MALS and SV-AUC, and compared between the wild type protein and the variants harboring mutations that either impair the head-to-tail binding, the LisH-dimerization or both (figure 3.29).

The molecular weight of the wild-type protein was detected in SV-AUC in between the molecular weight of a trimer and a tetramer, whereas it closely matched the molecular weight of a tetramer in SEC-MALS. When either the LisH-dimerization or the head-to-tail interaction were disturbed, the molecular weight rendered by SV-AUC analyses corresponded to that of a dimer. While for the N144R variant, which is impaired in the head-to-tail interaction, the molecular weight detected by MALS also closely matched that of the dimer, it was in between that of the monomer and the dimer for the F184E/L196Q variant, which is impaired in the LisH-mediated dimerization. This ambiguous molecular weight was repeatedly obtained for the F184E/L196Q variant, and the determination depended on the protein concentration (see insert in figure 3.29). This deviation is most likely caused by a fast equilibrium between the monomer and dimer mediated by the remaining head-to-tail interaction, which

|	Mw$_{MALS}$ [kDa]	Mw ratio $^{MALS}/_{theo}$	Mw$_{AUC}$ [kDa]	Mw ratio $^{AUC}/_{theo}$
— wt	344	4.1	297	3.5
— R	167	2.0	151	1.8
— EQ	130	1.5	164	1.9
— REQ	84	1.0	71	0.8
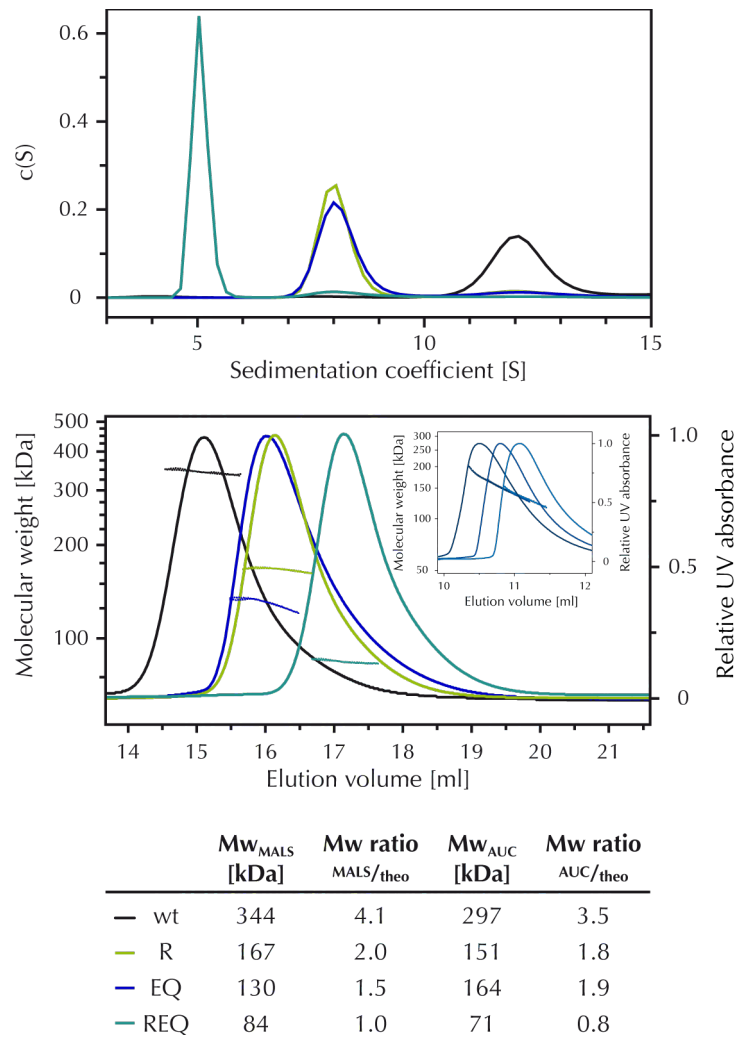
Figure 3.29 – Analysis of the oligomeric state of full-length muskelin. The wild type (wt, black), N144R (R, light green), F184E/L196Q (EQ, blue) and N144R/F184E/L196Q (REQ, teal) variants were analyzed with SV-AUC and SEC-MALS. *Top:* Sedimentation coefficient distributions obtained for the wild type and variants in SV experiments. Data were fitted as continuous c(s) distribution at a 0.95 confidence level resulting in r.m.s.d. values of 0.0095 for the wild type, 0.008 for the R, 0.0083 for the EQ and 0.0068 for the REQ variant. *Center:* The molar masses obtained by MALS as a function of elution volume are overlaid with the respective SEC A$_{280}$ chromatograms obtained when applying protein concentrations of 40-45 µM. In the inset, three SEC-MALS measurements of the F184E/L196Q variant at concentrations of 30, 60 and 120 µM (in light, middle and dark blue) are overlaid. Column used: Superose™ 6 10/300 GL. *Bottom:* The table summarizes the experimental molecular weights (Mw$_{AUC}$, Mw$_{MALS}$) and the ratios of the experimental molecular weight to the theoretical molecular weight of the monomer of 85 kDa ($^{AUC}/_{theo}$ and $^{MALS}/_{theo}$). (M9, M42, M96, M97)
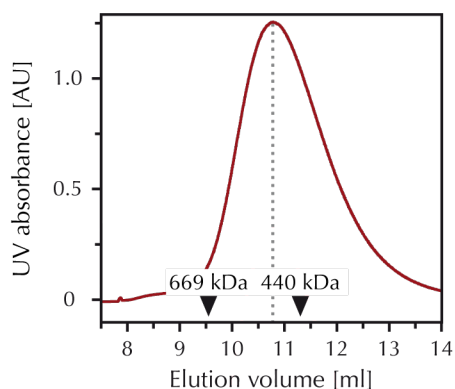
Figure 3.30 – Comparison of Mkln$_{1\text{-}735}$ elution volume to marker proteins. The A$_{280}$ chromatogramm of Mkln$_{1\text{-}735}$ is shown with arrowheads marking the positions of peak maxima for thyroglobulin (669 kDa) and ferritin (440 kDa), and a grey dashed line indicating the position of the maximum for Mkln$_{1\text{-}735}$.
Column used: Superdex™ 200 10/300 GL. (M9)
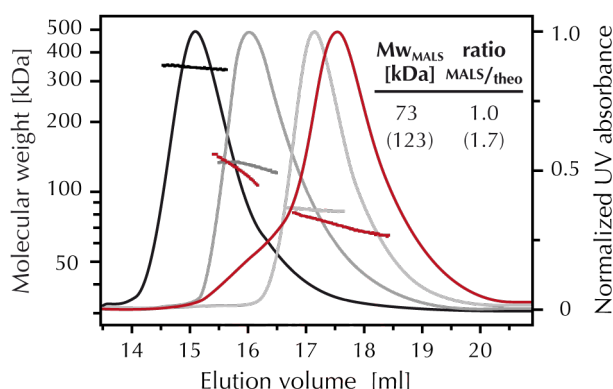
Figure 3.31 – SEC-MALS analysis of Mkln$_{1\text{-}625}$. The molar masses obtained by MALS as a function of elution volume are overlaid with A$_{280}$ chromatograms in red for Mkln$_{1\text{-}625}$, Mkln$_{1\text{-}735}$ wild type (black), F184E/L196Q (dark grey) and N144R/F184E/L196Q (light grey) are shown for comparison. The obtained molecular weight (Mw) for the main peak and the shoulder (values in brackets) and the ratio to the theoretical Mw (ratio ^{MALS}/$_{theo}$) is given.
Column used: Superose™ 6 10/300 GL. (M108)

could not be separated by SEC, and interfered with molecular weight determination in MALS. The fact that the head-to-tail impaired N144R variant was observed as dimer confirms that the head-to-tail interaction is an intermolecular, not an intramolecular binding event. Finally, when both interactions were impaired, the detected molecular weight corresponded to that of the monomeric protein in both methods used. This shows that both the head-to-tail binding and the LisH-dimerization are needed to assemble the oligomer present *in vitro*, which is a tetramer formed as a dimer of dimers.

The assignment of the oligomer present *in vitro* to a tetramer contradicts an earlier report in which muskelin was reported to assemble into a hexamer (Kiedzierska et al., 2008). This earlier study assessed the oligomeric state of muskelin based on SEC with a set of marker proteins including the 440 kDa protein ferritin. Consistent with the findings of that publication, muskelin was also observed to elute earlier than ferritin in this work (figure 3.30). The difference in the assigned molecular weight is therefore due to differences in the methods used for analysis rather than an actual difference of the analyzed protein preparations. Among the methods that were applied to determine the molecular weight of the muskelin oligomer in the previous publication and this work, SEC, SV-AUC and MALS, only the latter is independent of the shape and directly measures the molecular mass of the protein, giving the most accurate molecular weight determination for a homogenous solution. In contrast, the elution volume

in SEC is influenced by the shape of the molecule and overestimates the molecular weight for elongated molecules. Together with the molecular weight determination by MALS, this is an indication that the muskelin tetramer adopts a more elongated form than the almost perfectly spherically shaped ferritin. The preliminary SAXS data derived for $Mkln_{1-735}$ also hint towards a rather elongated shape of the tetramer (see section 3.2.4). Moreover, the deviation towards a smaller molecular weight for the muskelin tetramer in AUC experiments is also in line with this assumption, as in sedimentation velocity experiments, particles of equal molecular weight will sediment faster if spherical than if elongated in shape. The molecular weight therefore tends to be underestimated for elongated molecules.

The fact that the mutation of the two interfaces, the LisH-dimerization and the head-to-tail binding, were sufficient to convert muskelin to its monomeric form implied that these are the only interactions contributing to muskelin oligomerization. Yet the earlier observation that the LisH-dimerization was depending on the presence of the C-terminal module accounted for a more complex interplay of the domains in oligomerization. To investigate the contribution of the C-terminal module to the oligomer formed by full-length muskelin, the oligomeric state of $Mkln_{1-625}$, a construct lacking only the C-terminal module, was determined using SEC-MALS (figure 3.31).

Interestingly, two species were detected in SEC-MALS analysis of $Mkln_{1-625}$, a large fraction with a molecular weight matching that of the monomer, and a smaller fraction with a molecular weight close to that of a dimer. The ambiguity in molecular weight determination for the smaller fraction is likely due to a non-homogeneous mixture, caused either by a fast, dynamic equilibrium (as for the F184E/L196Q variant), or by the poor separation of the monomer and the dimer in SEC. Nevertheless, this result clearly demonstrates that the C-terminus not only stabilizes the LisH-dimerization, but also influences the head-to-tail interaction. The results obtained earlier for the isolated head-to-tail binding (see figure 3.27 on page 97) contradict a direct contribution to the binding, thus indicating that the C-terminus plays an indirect, but fundamental role in a more complex interplay of the domains driving oligomerization.

Combining the results described above with the conclusions drawn from the crystal structure of $Mkln_{12-205}$ and the mapping of the head-to-tail binding, a model of the muskelin tetramer can be constructed. The following criteria were taken into consideration for developing this model:

- The arrangement of the discoidin domain and LisH motif in the full-length protein corresponds to that observed in the $Mkln_{12-205}$ structure

- The head-to-tail-binding is mediated via the generic binding site in the discoidin domain.
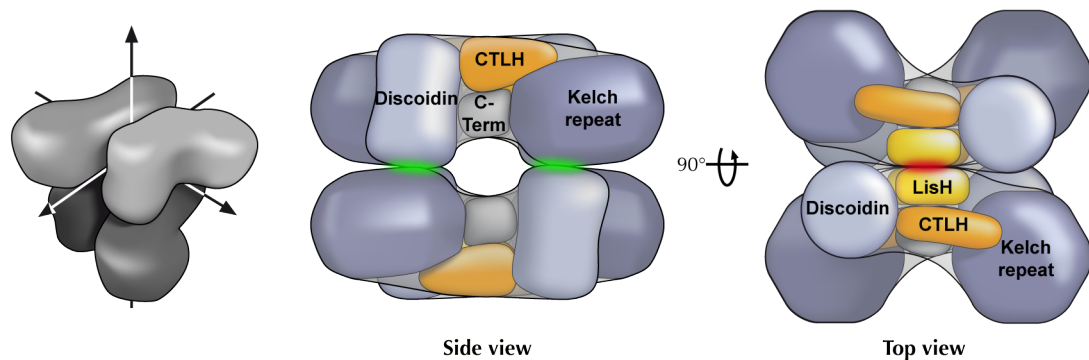
Figure 3.32 – Hypothetical model of the muskelin tetramer. An overview is shown in grey on the *left*, and more detailed models are shown in side (*middle*) and top view (*right*). The domains are indicated and are colored as introduced in figure 3.2 on page 58. The interactions underlying tetramerization are indicated in red for the LisH-dimerization and green for the head-to-tail binding. Note that predictions of the shape are neither possible for the CTLH-domain nor the C-terminal module. Still placeholder shapes are depicted for both, as assumptions about their position were considered in developing the model.

- The kelch-repeat domain adopts the generic fold.

- The N- and the C-terminus of the kelch domain are in close spatial proximity, and consequently the CTLH motif and the C-terminal are adjacent

- The arrangement of the CTLH-motif, the kelch repeat domain and the C-terminal module matches the shape derived for $Mkln_{205\text{-}735}$ in SAXS analyses (figure 3.15)

- The C-terminal module is in contact with the LisH motif to mediate its stabilizing effect on the LisH-dimerization

- The tetramer is closed and each of the interfaces is occupied

- The loss of either LisH dimerization or head-to-tail interaction results in a stable dimer which is assembled via the remaining interaction

A hypothetical model that meets all these criteria is illustrated in figure 3.32. In the absence of additional structural information, this model features only one likely arrangement of the individual domains. Further refinement of the model would require the use of additional techniques. Single-molecule förster resonance energy transfer (FRET) (Kalinin et al., 2012) and chemical crosslinking in conjunction with mass-spectrometry (Leitner et al., 2010), both could support the deduction of the architecture by mapping intramolecular distances. These informations would also facilitate the interpretation of data derived by low resolution structural characterization techniques such as SAXS, electron microscopy or atomic force microscopy, which could be used to derive an experimental model of the tetramer.
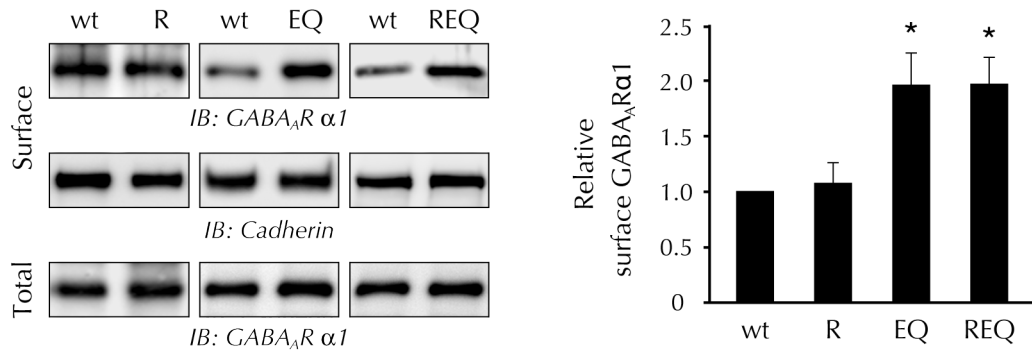
Figure 3.33 – Effect of impaired oligomerization on GABA$_A$R α1 surface levels in HEK293 cells coexpressing the α1 and β3 GABA$_A$ receptor subunits and the respective full-length variant of muskelin: wild type (wt), N144R (R), F184E/L196Q (EQ) or N144R/F184E/L196Q (REQ). After biotinylation of surface proteins, streptavidin-precipitated GABA$_A$R α1 and pan cadherin (control for equal protein amounts) and of total GABA$_A$R α1 (control for equal expression levels) were detected by western blot analysis (*left*). The quantification of signal intensities of surface GABA$_A$R α1 (*right*) compared to wild type (set to 1) in four independent experiments resulted in: R 1.07 ± 0.17 (n = 3), EQ 1.95 ± 0.28 (n = 5) and REQ 1.97 ± 0.23 (n = 4). Data are represented as mean values$\pm$SEM, and the Student's t-test (*$p < 0.05$) was used for statistical analysis. Data were kindly provided by Dr. Frank Heisler[8]. (M117, M120, M123, M126)

3.3.5 Effect of oligomerization on muskelin function

Since its discovery, muskelin's ability to oligomerize has been suggested to be a target for the regulation of its function (Prag et al., 2004). Yet, so far only possible initiations of regulation, like phosphorylation by PKC (Prag et al., 2007), but not the consequences have been investigated. In the previous section, it was shown that the mutations identified in this work were suitable tools to disassemble the muskelin oligomer and enforce different oligomeric states: the wild type tetramer, the LisH-mediated dimer, the head-to-tail-bound dimer and the monomer. Hence these mutations eventually allowed to directly address the functional consequences of alterations in muskelin oligomerization.

Dr. Frank Heisler, in the group of Dr. Matthias Kneussel[8], tested the effect of the mutations in the individual interfaces in a cellular environment. As indicator of muskelin function, the surface expression levels of the GABA$_A$ receptor α1 subunit (GABA$_A$R α1) were compared upon coexpression of wild type muskelin and the mutated variants impaired in the head-to-tail binding (N144R), the LisH-dimerization (F184E/L196Q), or both (N144R/F184E/L196Q), in HEK293 cells (figure 3.33). While no significant differences were detected between the wild type and the N144R variant, the surface levels of GABA$_A$R α1 were markedly increased (by about 95%) upon coexpression of the

---

[8] AG Kneussel, Center for Molecular Neurobiology (ZMNH) University Medical Center Hamburg-Eppendorf, D-20251 Hamburg, Germany
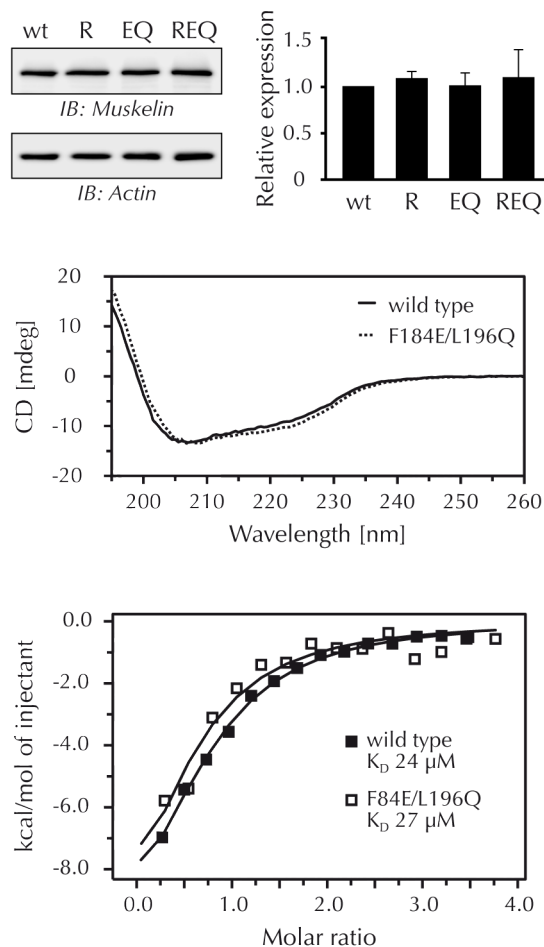
Figure 3.34 – Controls of expression, folding and GABA_A receptor binding of the Mkln_{1-735} F184E/L196Q variant. *Top:* Total expression levels of mCherry-muskelin-fusion proteins (wild type = wt, N144R = R, F184E/L196Q = EQ, N144R/F184E/L196Q = REQ) in HEK293 cells assessed by western blot analysis with actin detection as control. The quantification of western blot signal intensities of the variants compared to the wild-type (set to 1) resulted in: R 1.05±0.05; EQ 1.02±0.08; REQ 1.06±0.13; 1 (n = 5). Data are represented as mean values ± SEM. Data were kindly provided by Dr. Frank Heisler[8] *Center:* UV CD spectra of recombinant Mkln_{1-735} wild type (solid) and the F184E/L196Q variant (dashed). *Bottom:* Measured heat releases upon titration of Mkln_{1-735} wild type (filled squares) or the F184E/L196Q variant (open squares) with the intracellular domain, GABA_AR α1_{334-420}, plotted as a function of the molar ratio of the binding partners. The derived K_D values are given in the leged. (Expression: M119, M122, M125, M128, CD + ITC: M9, M42)

F184E/L196Q variant, in which the LisH-dimerization is impaired, and the monomeric N144R/F184E/L196Q variant. This indicates that the loss of LisH-dependent dimerization is detrimental to the function of muskelin in GABA_A receptor transport.

To exclude that these observations were biased by unequal expression levels, folding defects or an impaired binding of the F184E/L196Q variant to the GABA_A receptor α1 subunit, expression levels of the different muskelin variants in HEK cells were assessed by western blot analysis, and fold, thermal stability and binding properties of the recombinant wild type protein and the F814E/L196Q variant were compared in CD spectroscopy, a ThermoFluor assay and ITC measurements (figure 3.34). Differences in the total expression level in HEK cells proteins could be excluded, as all of the mCherry-fused variants were detected at equal levels. The recorded CD spectra were

almost identical between the wild type protein and the F184E/L196Q variant and the thermal stability of the F184E/L196Q variant was negligibly decreased by 2 °C (not shown); both methods thus indicated an intact fold of the F184E/L196Q variant. Likewise, the determined affinities of the binding to the GABA$_A$ receptor α1 subunit were found to be identical for both proteins within the limits of experimental uncertainty in the ITC analysis . The ITC analysis of the binding of Mkln$_{1-735}$ F184E/L196Q to GABA$_A$R α1$_{334-420}$ was repeated in three further independent experiments, resulting in a mean K$_D$ of (20 ± 10) μM, with a stoichiometry of 0.8 ± 0.2, a change in enthalpy of (- 5 ± 3) kcal/mol and a change in entropy of (-8 ± 15) cal/(mol K), which is in perfect agreement with the results obtained for the wild type protein (see section 3.4.1 on page 118). Taken together, this corroborates that the observed differences in GABA$_A$R α1 surface levels were directly caused by the impaired LisH-dimerization.

The observed impairment in the functionality of the F184E/L196Q variant necessitated further controls of the prerequisites of correct protein function. As equal expression levels were already verified, the next prerequisites to be tested were the correct processing and targeting of the protein. As no specific processing needed for muskelin maturation and activity was known, only the targeting of the protein could be examined. To this end, the intracellular distribution of the muskelin variants was tested by using fluorescent protein-fusions in different cell types (figure 3.35).

Upon overexpression in HEK 293 cells, striking differences in the intracellular distribution were observed. The wild type protein as well as the N144R variant were predominantly cytoplasmic, an observation which is in line with previous reports (Adams et al., 1998; Prag et al., 2004; Tagnaouti et al., 2007; Valiyaveettil et al., 2008). In contrast, the LisH-dimerization impaired F184E/L196Q and the monomeric N144R/F184E/L196Q variants were redistributed to the nucleus, and hardly detectable in the cytosol. This effect held true for the vast majority of cells, as revealed by quantification, and was also observed in hippocampal neurons. The intracellular distribution therefore seems to be highly dependent on the LisH-mediated dimerization. A similar impact of a disturbed LisH-dimerization has previously been reported for three other LisH motif containing proteins, LIS1, TBL1 and oral-facial-digital syndrom 1 protein (Gerlitz et al., 2005). Mutations in the LisH motif, which disrupted the dimerization, led to an altered intracellular distribution and a dramatic decrease in protein half life in all three cases. It remains to be answered in future experiments whether the effect on protein half life also applies to muskelin.

Notably, no effect on the function or localization of muskelin by the head-to-tail binding impaired N144R variant was observed. A regulation of the head-to-tail binding via phosphorylation by PKC has been proposed (Prag et al., 2007) and, although the suggested relevance of two specific residues could not be confirmed in this work (see section 3.3.3 on page 96 and figure 3.26), it is still likely that the head-to-tail
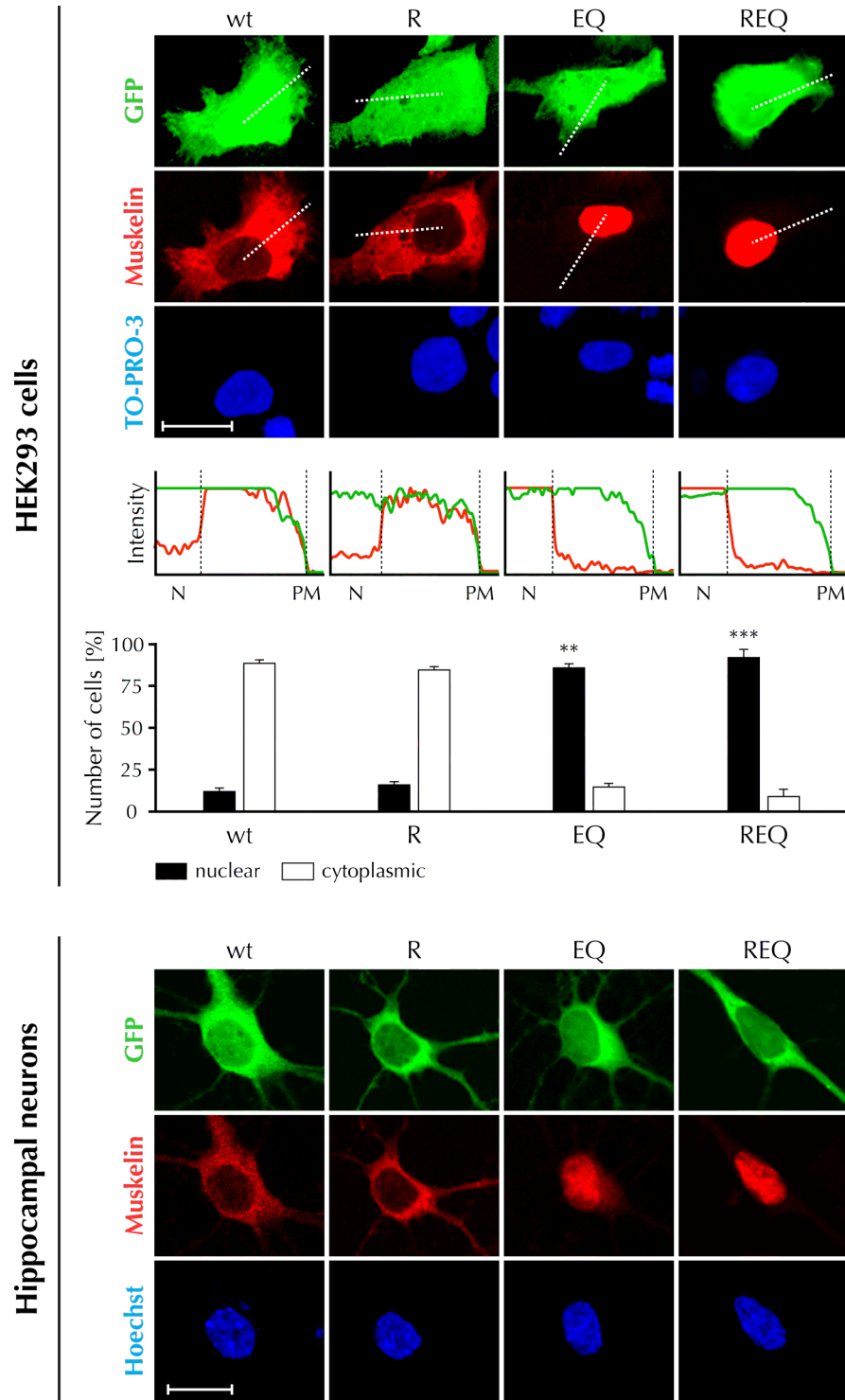
interaction also is a target for modulating the function of muskelin. It remains to be seen whether another functional readout is needed to detect the effect of an impaired head-to-tail interaction. At this point, it cannot be ruled out that in the cells used in the analysis presented in this work, the head-to-tail binding is *per se* inactivated and thereby potential effects of a disturbed interaction would be concealed.

3.3.6 Coupling of muskelin oligomerization and cellular localization

A striking redistribution of muskelin to the nucleus, similar to the findings described above, was observed in an earlier study (Valiyaveettil et al., 2008) on determinants of muskelin's subcellular localization, when the very C-terminal 35 residues of muskelin were deleted. Less pronounced changes were reported for a double mutation of two basic residues in the LisH motif (K182A/H183A) and a phosphomimetic mutation of a residue within the crucial last 35 residues (T725D), which both resulted in a uniform distribution of muskelin between cytoplasm and nucleus. Based on their results, the authors proposed an interplay between a cytoplasmic restraint signal in the C-terminal residues and a cryptic nuclear localization activity by the LisH motif.

The resemblance of the changes in subcellular localization between the oligomerization impaired mutations presented in this work and the mutations described by Valiyaveettil et al. (2008) prompted a more detailed analyses of the latter. To account for the possible effects of a cytoplasmic restraint and a nuclear localization signal on subcellular localization, the K182A/H183A double mutation in the LisH motif and the C-terminal deletion (ΔC35) were included in the analysis.

Figure 3.35 *(facing page)* – **Intracellular distribution of muskelin full-length variants.** The wild type (wt) protein and the head-to-tail binding impaired N144R (R), the LisH-dimerization impaired F184E/L196Q (EQ) and the monomeric N144R/F184E/L196Q (REQ) variants were overexpressed and their intracellular distribution was analyzed in HEK293 cells (*top*) and hippocampal neurons (*bottom*). Confocal images are shown with red fluorescence detecting muskelin variants fused to either mRFP- (HEK293 cells) or mCherry (hippocampal neurons), green fluorescence detecting green fluorescent protein (GFP) (coexpressed to identify cell boundaries) and blue fluorescence detecting nuclear stain (TO-PRO3® for HEK293 cells, Hoechst 33342 for hippocampal neurons). Scale bars correspond to 15 μm. For HEK293 cells, line scans (traces indicated by dashed lines in the confocal images) are shown below the corresponding images, with borders of the nucleus (N) and the plasma membrane (PM) indicated by dashed lines. Subcellular localization in HEK293 cells was quantified resulting in: wild type (11.7±1.2)% nuclear and (88.3±1.2)% cytoplasmic; R (15.7±1.0)% nuclear and (84.3±1.0)% cytoplasmic; EQ (85.6±1.6)% nuclear and (14.4±1.6)% cytoplasmic; REQ (92.3±2.7)% nuclear and (7.7±2.7))% cytoplasmic, in three independent experiments. In the graph, data are represented as mean values±SEM. The Student's t-test (**p < 0.01; ***p < 0.001) was used for statistical analysis. Images and data were kindly provided by Dr. Frank Heisler, AG Kneussel, Center for Molecular Neurobiology (ZMNH), University Medical Center Hamburg-Eppendorf. (mRFP-fused proteins: M118, M121, M124, M127, mCherry-fused proteins: M119, M122, M125, M128.)
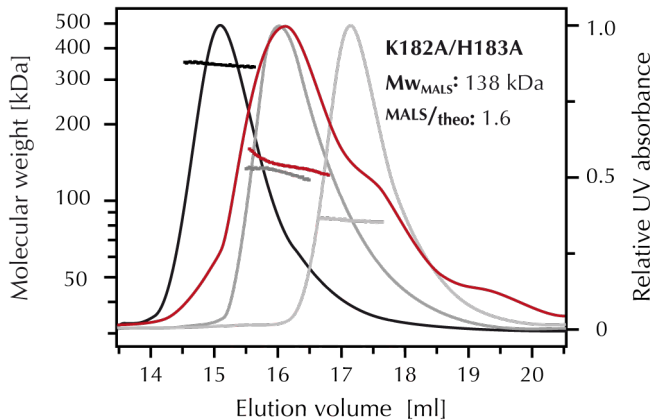
K182A/H183A   **ΔC35**

GFP   Muskelin   Hoechst   GFP   Muskelin   Hoechst
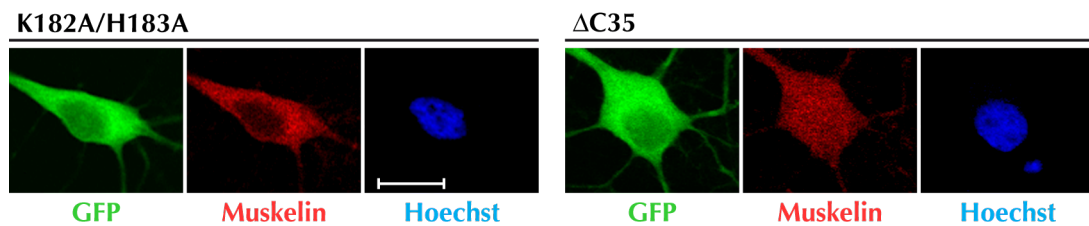
Figure 3.36 – Intracellular distribution of the muskelin K182A/H183A and ΔC35 variants in hippocampal neurons. Confocal images are shown with red fluorescence detecting mCherry-fused muskelin variants, green fluorescence detecting GFP (coexpressed to identify cell boundaries) and blue fluorescence detecting the nuclear stain Hoechst 33342. The scale bar corresponds to 15 μm. Images were kindly provided by Dr. Frank Heisler, AG Kneussel, Center for Molecular Neurobiology (ZMNH), University Medical Center Hamburg-Eppendorf. (M129, M130)

First, the localization of the K182A/H183A and the ΔC35 variant in hippocampal neurons was investigated by Dr. Frank Heisler[8] (figure 3.36). The intracelluar distribution of the $Mkln_{1\text{-}735}$ K182A/H183A variant resembled that of the wild type protein (compare figure 3.35 on page 108, bottom, with figure 3.36, left) with muskelin being predominantly localized to the cytoplasm. This distribution was altered for the ΔC35 variant, which was found to be present equally in the nucleus and cytoplasm. This deviates from the earlier description of $Mkln_{1\text{-}735}$ K182A/H183A being uniformly distributed between the nucleus and the cytoplasm and the ΔC35 variant being mainly localized in the nucleus (Valiyaveettil et al., 2008). Thus, the reported effect of the mutations was found to be less pronounced in a neuronal background. Still, a clear influence of the C-terminus on muskelin localization was observable.

In experiments presented earlier in this work, also a fundamental impact of the C-terminus on muskelin oligomerization was observed (figure 3.31 on page 102). The biochemical results on the impact of the C-terminus in this work were obtained for a construct ($Mkln_{1\text{-}625}$) with a complete deletion of the C-terminal module, while the C-terminally truncated variant that was analyzed for its subcellular localization was shortened only by the last 35 amino acids ($Mkln_{1\text{-}700}$). To bridge this gap, a biochemical analysis of the ΔC35 variant was attempted. However, this was prevented by the fact that the protein was prone to aggregation and inseparable from the bacterial chaperone protein DnaK during purification. Still, the elution volume in preparative size exclusion chromatography was close to the elution volume expected for its dimeric form, which hinted towards an impaired oligomerization similar to what was observed for $Mkln_{1\text{-}625}$. This again argues for a tight coupling of the oligomeric state of muskelin to its subcellular localization.

To further pursue this coupling, the effect of the K182A/H183A mutation on muskelin oligomerization was also investigated. Isolation of pure $Mkln_{1\text{-}735}$ K182A/H183A

Figure 3.37 – SEC-MALS analysis of Mkln$_{1-735}$ K182A/H183A. Molar masses obtained by MALS as a function of elution volume are overlaid with the respective SEC A$_{280}$ profiles for Mkln$_{1-735}$ K182A/H183A (red), wild type (black), F184E/L196Q (dark grey) and N144R/F184E/L196Q (light grey). (M99, M9, M42, M97)

and Mkln$_{1-205}$ K182A/H183A proved difficult, as the proteins showed a strong tendency towards aggregation and degradation. Still, sufficient amounts of reasonably pure protein could be obtained for subsequent biochemical analyses. The effect on LisH-dimerization was tested in crosslinking experiments comparing different variants of Mkln$_{1-205}$ (wild type, K182A/H183A and F184E/L196Q), and the influence on the oligomerization of the full-length protein was tested by assessing the oligomeric state of Mkln$_{1-735}$ K182A/H183A in SEC-MALS.

In the SEC-MALS analysis of the Mkln$_{1-735}$ K182A/H183A variant (figure 3.37), a mixed distribution was observed, with the predominant form detected with a molecular weight in between that of the monomer and the dimer. Two marked shoulders of the main peak indicate the presence of both higher and lower oligomers, the latter of which were more abundant. The position of the main peak as well as the determined molecular weight resembled that found for the LisH-dimerization impaired F184E/L196Q variant. Taken together, this demonstrated a marked destabilization of the oligomeric assembly in the K182A/H183A variant, which seemed to be primarily impaired in the LisH-mediated dimerization. Interestingly, the effect of the K182A/H183A mutation on the oligomerization of the full-length protein observed in the SEC-MALS analysis (figure 3.37) was not as prominent as for the other variants tested, as still a small amount of higher oligomers could be detected. This resembles the attenuated phenotype in subcellular distribution analyses of this variant, which was found to not be completely relocalized; it was either equally distributed between the cytoplasm and the nucleus in the previous study (Valiyaveettil et al., 2008), or was even unimpaired in the analysis presented here (figure 3.36).

Notably, when the ability to dimerize via the LisH motif was tested in isolation by crosslinking analyses of the Mkln$_{1-205}$ K182A/H183A variant (figure 3.38), it was found to be similar to that of the wild type. This seems contradictory at first sight, but

111

Figure 3.38 – Position of Lys182 and His183 in the Mkln$_{12-205}$ structure *Left*: Close-up view of the LisH-dimerization interface in cartoon representation shown with side chains in the interface and of Lys182 and His183 (red) in stick representation. *Right*: Coomassie stained gels of separated products of crosslinking of Mkln1-205 variants performed at three protein concentrations (indicated in μM) and a control (ctrl) without crosslinker. (M2, M101, M30)

is in good agreement with the position of the residues in the LisH motif. Both residues do not contribute to LisH-dimerization, instead they point away from the LisH interface (figure 3.38). Lys182 and His183 thereby can only be involved in interactions of other parts of the protein with the LisH motif, and it is well conceivable that they are important for the stabilizing effect on LisH-dimerization mediated by the C-terminal part of muskelin. When this stabilization is lost, the isolated LisH-dimerization is not disturbed, but the oligomer of the full-length protein is destabilized, thus explaining both results, that of the crosslinking analyses of the Mkln$_{1-205}$ variants and that of the SEC-MALS analyses of the Mkln$_{1-735}$ variants.

In summary, a direct effect on the oligomerization of the isolated proteins was observable for both, a deletion of the C-terminal module, and the K182A/H183A variant. These results underscore that muskelin oligomerization and intracelluar localization cannot be considered independently and shed new light on the findings of the previous study by Valiyaveettil et al. (2008). They interpreted their findings to indicate a complex interplay between a cryptic nuclear import and a cytoplasmic restraint signal. This might now be replaced by a more straightforward explanation according to which the oligomeric state of muskelin determines its localization. It would be very interesting to further follow up on this hypothesis and explore its implications. For instance, the suggested localization-determining phosphorylation of the C-terminus by a proline-directed serine/threonine kinase might be a regulatory mechanism for muskelin oligomerization.

So far, the question on how the nuclear localization of muskelin is mediated is still open. Nuclear import and export are facilitated by transport receptors[9] that usually bind to specific signals within the cargo protein termed nuclear localization signal (NLS) or nuclear export signal (NES), respectively[10]. The shuttling of a protein between nucleus and cytoplasm can be regulated by (1) modifications of a NLS or NES, (2) by masking of a NLS or NES, either by inter- or intramolecular interactions, or (3) by supply of the NLS or NES by another protein (Poon and Jans, 2005).

To explore the first and second option, the modification or masking of a signal sequence, first the presence of such signal sequences in muskelin has to be considered. Yet, predictions of NLSs and NESs suffer from both limited sensitivity and specificity, especially the prediction of classical NESs (Xu et al., 2012), and results have to be considered with caution. For muskelin, the NLS prediction programs *cNLS Mapper* and *NLStradamus* do not find any classical monopartite NLS, and only a questionable bipartite motif with an unusually long linker[11]. This bipartite motif is not identical, but overlapping with the non-canonical NLS in the LisH motif suggested by Valiyaveettil et al. (2008). When searching for NESs with the prediction programs *NetNES 1.1* and *NES Finder 0.2*, only the latter detects four potential NES sequences, out of which three can be ruled out directly[12]. The absence of a NES from the muskelin sequence would be in line with the finding by Valiyaveettil et al. (2008) that muskelin does not undergo exportin-1-mediated nuclear export. Hence, the presumably only signal sequence is the putative NLS sequence in the LisH motif. This suggests a double function of the LisH motif - as dimerization element and as NLS carrier. Notably, all findings obtained within this thesis and published by Valiyaveettil et al. (2008) can be reconciled when assuming a dual role. The intramolecular interactions that stabilize the LisH-mediated dimerization would then additionally mask the NLS. This is consistent with the contribution of Lys182 and His183 to oligomerization shown here. In partic-

---

[9] The majority of nuclear transport processes is mediated by members of the karyopherin protein family, which is subdivided in the importins (further subdivided in the α- and β-isoforms) and the exportins.

[10] The classical NLSs fall into two classes; the monopartite NLSs that have a single cluster of basic amino acid residues and the bipartite NLSs that contain two clusters of basic amino acids separated by 10 to 12 amino acid linker. Classical NLSs are recognized by importin α, classical leucine-rich NESs are recognized by the key exportin, exportin-1 (also termed chromosome region maintenance 1 protein homolog, Crm1).

[11] Motif identified by *cNLS Mapper* (score 5.6 out of 1 to 10): ^{178}RLCLKHFRQHNYTEAFESLQKKTK^{201}.

[12] The NES Sequences detected by *NES Finder 0.2* are (1) ^{81}VCNLKKFKV^{90} in the discoidin domain, where the hydrophobic residues are not surface exposed, (2) ^{520}IDPELNEIHV^{529} in the fifth kelch repeat, where the hydrophobic residues are predicted to be engaged in maintaining the overall fold and hence are not surface exposed, (3) ^{615}LDDFWSLKL^{623} in the sixth kelch repeat, where the hydrophobic residues are again predicted to be engaged in maintaining the fold and hence are not surface exposed, (4) ^{656}LKYLQNDLYI^{665} in the C-terminal module, predicted to be completely α-helical (*PROFphd*, *PredictProtein*) and therefore more likely represent a false positive hit (Xu et al., 2012).

ular, it would offer an explanation for the contradictory effects of the K182A/H183A mutation observed by Valiyaveettil et al. (2008): Introducing this mutation impaired the nuclear import activity of the LisH motif when placed in a heterologous protein, and reduced the nuclear import of $Mkln_{1-700}$, but resulted in an increased nuclear localization of full-length muskelin. Assuming that this mutation not only destabilizes the dimer but also impairs the recognition of the NLS by the transport receptor, but to a level that does not completely abolish the NLS activity, the contradictory effects simply reflect the unmasking of a partially active NLS: The NLS is not active enough to import the foreign protein, but in muskelin it is still active enough to partially import the protein when the NLS is unmasked by the loss of dimerization, however, it is not active to the full extent, as revealed when the NLS is unmasked by the deletion of the C-terminal residues in $Mkln_{1-700}$ or the mutation of Thr723.

The third option mediating changes in nuclear trafficking is the masking of a signal sequence or the supply of a foreign signal sequence by an interaction with another protein. To explain the observations described in this thesis in the literature, this interaction would need to depend on the status of LisH-dimerization in muskelin. In this context, the association of muskelin with the CTLH complex via RanBP9 (see Introduction section 1.2.1 on page 10) is an interesting candidate. For both muskelin and RanBP9, a relocalization to the nucleus upon HO-1 induction in prostate cancer cells has been reported (Gueron et al., 2014), implying that they undergo parallel changes in localization upon specific triggers. Moreover, a differential colocalization of components of the CTLH complex in the nucleus and the cytoplasm has been reported (Kobayashi et al., 2007), which suggests a compartment-specific composition of the complex. Based on a comparison with the elution of marker proteins in SEC, the size of the CTLH complex was estimated to be slightly larger than 670 kDa (Nishitani et al., 2001; Umeda et al., 2003). Single copies of all identified components in humans would add up to a mass of 353 kDa[13], thus arguing that either not all components have been identified so far, or one, several or all of the subunits are contained as oligomers. This is very likely given the fact that five of the six known subunits (all but Armc8) carry a LisH motif, and thus a dimerization element. The sum of the molecular weights increases to 633 kDa when all LisH motif containing proteins are included twice, which is close to the estimated molecular weight. Muskelin in its tetrameric form together with the other LisH motif containing proteins in their dimeric form finally add up to about 803 kDa, which is markedly larger than the estimated size of the CTLH complex. Hence a recruitment of muskelin to the complex that requires its dimeric state seems to be the most likely scenario. It remains open whether a change

---

[13] Molecular weights of the human proteins: Muskelin 84.9 kDa, RanBP9 77.9 kDa, Twa1 26.8 kDa, Rmnd5a 44 kDa, MAEA 45.3 kDa, Armc8 75.5 kDa.

in localization by an altered LisH-dimerization is incompatible with, or a a prerequisite for, or a consequence of the recruitment of muskelin to the CTLH complex.

It would be highly interesting to assess the oligomeric state of muskelin in the cell, especially in different cellular compartments and in response to specific stimuli, such as the induction of HO-1 (Gueron et al., 2014). This would furthermore allow to control wether the head-to-tail binding is active in certain cellular contexts, and possibly provide an explanation as to why no effect of an impaired head-to-tail interaction has been observed in the experiments presented here. Yet, this is a technically challenging endeavor. For an analysis of the oligomeric state within cells, microscopy techniques represent the most promising approach. Different variants of FRET-based measurements could be tested with the isolated proteins beforehand and then transferred to measurements in cells. One possibility would be single-molecule FRET (Roy et al., 2008), which relies on the introduction of a suitable pair of fluorophores at two different positions in the protein. This would allow to detect rearrangements within the molecule or the oligomer by measuring the changes in FRET-dependent emission wavelengths. As an alternative to FRET-measurements, microscopy techniques that allow for an analysis of single particles, such as a ultrahigh-resolution imaging techniques or single-molecule fluorescence microscopy (Walter et al., 2008), could be used to judge the oligomeric state based on the fluorescence intensity of single particles or the correlated movement of fluorophores. However, these measurements would not allow to distinguish between the LisH-mediated and the head-to-tail-bound dimers.

If the analysis of the oligomeric state in the cell should prove to be too challenging, a further alternative is the analysis of cell extracts, similar to the approach used in the initial identifications of the CTLH complex components (Nishitani et al., 2001; Umeda et al., 2003; Kobayashi et al., 2007). In the study by Umeda et al. (2003) GST-fused muskelin was found ranging in size between 158 and 670 kDa - hence as a dimer, as a higher oligomer and bound in the complex. However, the fusion of a protein to GST is a major drawback as GST itself dimerizes (Fabrini et al., 2009) and thereby alters the oligomeric state of the fusion partner. Still, the study provides a first indication that muskelin can adopt different oligomeric states in the cell. The analyses could be improved by either analyzing the native protein and using a specific antibody for its detection or by choosing a different fusion-tag that does not affect the oligomerization of the protein, for example small epitope tags like the hemagglutinin or FLAG tag. Furthermore, the analysis could be extended by introducing a prior cellular fractionation to separate cytoplasmic and nuclear components.

3.4 Interaction between the GABA$_A$ receptor and muskelin

The objective of this work, the characterization of muskelin as a GABA$_A$ receptor $\alpha1$ subunit (GABA$_A$R $\alpha1$) binding protein, was based on a recent study by Heisler et al. (2011), in which the interaction of muskelin and GABA$_A$R $\alpha1$ was first described. The study furthermore provided an extensive characterization of the role of muskelin in the intracellular transport of the GABA$_A$ receptor (see also Introduction, section 1.1.2 on page 7). Within this work, a more in-depth analysis of the muskelin-GABA$_A$R $\alpha1$ interaction was performed as described in the following.

3.4.1 Mapping the interaction site in muskelin

The interaction of muskelin and the GABA$_A$ receptor $\alpha1$ subunit was initially discovered in a yeast two-hybrid screen using the intracellular domain of the GABA$_A$R $\alpha1$ as bait. It was subsequently verified in GST pull-down assays and reciprocal co-immunoprecipitations of the binding partners from brain lysates. To verify and further characterize this interaction, the isolated proteins were investigated using analytical SEC, NAGE and ITC. Moreover, the binding site for GABA$_A$R $\alpha1$ on muskelin was narrowed down by comparing the binding of shortened muskelin constructs (figure 3.39 on the next page).

In SEC-analyses, the differences in the elution profiles of the respective muskelin construct and the intracellular GABA$_A$R $\alpha1$ loop (GABA$_A$R $\alpha1_{334\text{-}420}$) when subjected to SEC alone and in combination were evaluated. In the analyses of Mkln$_{1\text{-}156}$ neither a shift of the absorption maxima in the elution profiles nor a co-elution of the binding partners could be detected. For Mkln$_{1\text{-}205}$, a small shift of the maximum and an

Figure 3.39 *(facing page)* – **Mapping of the GABA$_A$ receptor $\alpha1$ binding site on muskelin.** *Top*: SEC analyses: In the graphs, the A$_{280}$ (in absorption units, AU) is plotted versus the elution volume (in ml) for the respective muskelin variant (dashed grey), GABA$_A$R $\alpha1_{334\text{-}420}$ (dashed black) and the combined samples (red). The grey shaded areas mark the area of the fractions analyzed by SDS-PAGE as shown in the inlets (Labeling of lanes: c = combined samples, m = muskelin variant, α = GABA$_A$R $\alpha1_{334\text{-}420}$). Black arrowheads indicate the expected position of the respective muskelin variant, open arrowheads the position of GABA$_A$R $\alpha1_{334\text{-}420}$. Column used: Superdex™ 75 10/300 GL. Note that, although smaller in size, GABA$_A$R $\alpha1_{334\text{-}420}$ elutes at a smaller elution volume than Mkln$_{1\text{-}156}$ and Mkln$_{1\text{-}205}$, indicating a rather elongated shape of GABA$_A$R $\alpha1_{334\text{-}420}$. The multiple bands observed in SDS-PAGE analysis of the fractions of SEC of Mkln$_{1\text{-}156}$ and GABA$_A$R $\alpha1_{334\text{-}420}$ likely correspond to products of proteolytic degradation of GABA$_A$R $\alpha1_{334\text{-}420}$, which varied considerably between different preparations. *Center*: NAGE: Coomassie-stained native agarose gel of separated muskelin variants loaded alone (-) or mixed with GABA$_A$R $\alpha1_{334\text{-}420}$ ($\alpha1$). For comparison, GABA$_A$R $\alpha1_{334\text{-}420}$ alone was loaded in the outer lanes. *Bottom*: ITC analyses: Binding enthalpies upon titration of muskelin variants with GABA$_A$R $\alpha1_{334\text{-}420}$ are plotted as a function of the molar ratio of the binding partners; the derived binding parameters are summarized in the table to the side. (Mkln$_{1\text{-}735}$: M9, Mkln$_{1\text{-}156}$: M1, Mkln$_{1\text{-}205}$: M2, Mkln$_{1\text{-}272}$: M3, Mkln$_{157\text{-}735}$: M4, Mkln$_{205\text{-}735}$: M5, GABA$_A$R $\alpha1_{334\text{-}420}$: G1)

increased intensity of the band for $Mkln_{1-205}$ in the analyzed fraction were detected, however, the almost identical elution volumes for $Mkln_{1-205}$ and $GABA_AR$ $\alpha1_{334-420}$ hampered the analysis. Notably, for the complementary constructs to $Mkln_{1-156}$ and $Mkln_{1-205}$, namely $Mkln_{157-735}$ and $Mkln_{205-735}$, a shift of the elution maxima and a co-elution of $GABA_AR$ $\alpha1_{334-420}$ with the respective muskelin construct was detected, especially in the early fractions, where $GABA_AR$ $\alpha1_{334-420}$ is much less abundant when separated on the column alone. For $Mkln_{157-735}$, the analysis was impaired by its elution close to the void volume of the column (caused by its dimerization and thereby increased size), which limited the resolution of the SEC. Still, the changes were clearly indicating a complex formation.

The findings from the NAGE analysis agreed well with the results obtained by analytical SEC. A prominent shift in the band of the muskelin construct as well as a retention of $GABA_AR$ $\alpha1_{334-420}$ were observed for full-length muskelin, $Mkln_{157-735}$ and $Mkln_{205-735}$. In contrast, only subtle changes were found for $Mkln_{1-205}$ and $Mkln_{1-272}$, and no changes for $Mkln_{1-156}$. The appearance of the sample of $Mkln_{1-205}$ when combined with $GABA_AR$ $\alpha1_{334-420}$ resembled that of $Mkln_{1-156}$ in the presence of $GABA_AR$ $\alpha1_{334-420}$. Although suspect at first sight, this effect proved to be reproducible in several experiments.

Finally, in ITC analyses, binding parameters could be determined for the binding of $Mkln_{1-735}$, $Mkln_{157-735}$ and $Mkln_{205-735}$ to $GABA_AR$ $\alpha1_{334-420}$, while for $Mkln_{1-205}$ and $Mkln_{1-272}$ there was either no binding or it was too weak to be detected by ITC. The analysis of the binding of full-length muskelin to $GABA_AR$ $\alpha1_{334-420}$ was repeated in three independent measurements, giving a mean K_D of (18 ± 5) μM with a stoichiometry of 0.8 ± 0.1, an enthalpy change ΔH of (-8 ± 4) kcal/mol and an entropy change ΔS of (-6 ± 14) cal/(mol K). It has to be noted that these ITC measurements were performed at low c-values (ranging from 1.5 to 2.9) and the derived binding parameters (the stochiometry, ΔH and ΔS) have to be interpreted with caution (see methods section, section 2.2.3 on page 47). The binding of muskelin to the $GABA_A$ receptor $\alpha1$ subunit therefore was characterized as a 1:1 binding event with a low affinity and a favorable enthalpic and a small unfavorable entropic contribution.

Taken together, these results argue against a participation of the discoidin domain, leave open the possibility that the LisH motif may contribute to binding and indicate that the primary binding site resides in one or a combination of the CTLH motif, the kelch repeat domain and the C-terminal module. Interestingly, the fragment of muskelin that was identified in the initial yeast two-hybrid screen spans residues 90 to 200, corresponding to half of the discoidin domain and nearly the whole LisH motif. Based on the analyses described above, this fragment would be predicted to have at best a rather low affinity in comparison to the full-length protein. Notably, the $Mkln_{90-200}$ fragment was sufficient to cause increased $GABA_A$ receptor surface levels upon co-

118

expression in HEK293 cells (Heisler et al., 2011). This was supposed to be caused by Mkln$_{90\text{-}200}$ outcompeting native muskelin in GABA$_A$R α1 binding. While the high overexpression might well compensate for the weaker binding, there is an additional explanation for the observed effect: it is well conceivable that Mkln$_{90\text{-}200}$ competes with the full-length protein in LisH-dimerization and thereby also influences the function and potentially the subcellular localization, especially when considering that this fragment would be expected exhibit nuclear localization activity based on the findings of Valiyaveettil et al. (2008).

In future studies, a further refinement of the mapping of the binding site would fulfill a basic requirement to address more in-depth questions. A systematic mapping could either be conducted in extensive mutational analyses or with the use of oligopeptide arrays, as discussed earlier for mapping of the head to tail interaction on the kelch repeat domain (section 3.3.3 on page 99). With respect to muskelin's role in the GABA$_A$ receptor transport, the interplay of binding to the receptor subunit and the motor complex subunit dynein intermediate chain isoform 1A (DIC1A) would be of particular interest. As the binding site of DIC1A has been mapped to the central part of muskelin (residues 172 to 258, covering the LisH and CTLH motif), a simultaneous interaction of both partners seems well conceivable, but only a more detailed analysis can reveal wether these binding events are independent or cooperative and how they are regulated.

3.4.2 Subunit specificity

In the original publication (Heisler et al., 2011), muskelin was described to bind exclusively to the GABA$_A$ receptor α1 subunit. Specifically, binding to the α2, α3, α5, β2 or γ2 subunit was ruled out based on yeast two-hybrid screenings. Yet, in GST-pulldowns, α2 was also found to retain muskelin. Surface levels of the GABA$_A$ receptor α2 subunit were slightly increased in neurons derived from muskelin knockout mice, but the effect was much less pronounced when compared to the α1 subunit, corroborating that the α1 subunit is the primary target for muskelin. In yeast two-hybrid studies, the binding site on the GABA$_A$R α1 was mapped to the very C-terminal portion of the intracelluar domain with the last 31 residues sufficient and the last 22 residues necessary to detect binding. Notably, this region of the intracellular domain is highly conserved among the α1, α2, α3 and α5 subunits, raising the question of how subunit specificity is achieved.

To address subunit specificity, the binding of muskelin to the α1, α2, α3 and α5 subunit was tested. Since NAGE had proven to be suitable to detect weaker interactions in the earlier experiments, it was used for these experiments. As negative controls for binding and to confirm the mapping of the yeast two-hybrid experiments, variants of

α1 391 SATIEPKEVKP-----------ETKPPEPKKTFNSVSKIDR 420
α2 390 SATTPEPNKKP-----------ENKPAEAKKTFNSVSKIDR 419
α3 419 AAAAPSASSTPTVIASPKTTYVQDSPAET-KTYNSVSKVDK 458
α5 400 NAVGTASIRAS-----------EEKTSESKKTYNSISKIDK 429
          .   .   : .... **:**:**:*:
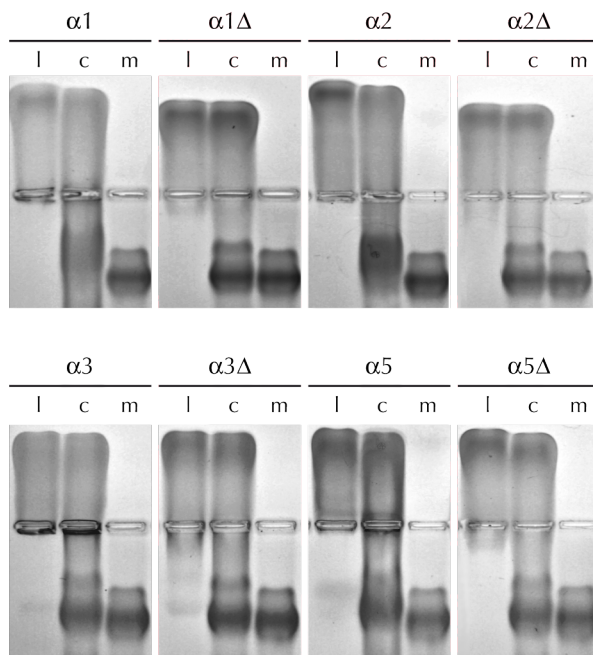
Figure 3.40 – Receptor subunit specificity of muskelin tested by native agarose gel electrophoresis. *Top*: Alignment of the last 30 amino acids of the intracelluar domains of the rat GABA$_A$ receptor α1, α2, α3 and α5 subunits with residues identified in yeast two-hybrid studies to be sufficient for binding of GABA$_A$R α1 to muskelin highlighted in red, and residues found to be necessary underlaid in dark red (Heisler et al., 2011). *Bottom*: Coomassie-stained native agarose gels onto which for each subunit the full intracellular loop (GABA$_A$R α1$_{334-420}$, GABA$_A$R α2$_{335-419}$, GABA$_A$R α3$_{360-458}$, GABA$_A$R α5$_{342-429}$) and the respective deletion variant (Δ) truncated by the last 13 amino acids (GABA$_A$R α1$_{334-407}$, GABA$_A$R α2$_{335-406}$, GABA$_A$R α3$_{360-445}$, GABA$_A$R α5$_{342-416}$) were separated either loaded alone (l) or combined with Mkln$_{1-735}$ (c) next to Mkln$_{1-735}$ alone (m) for comparison. (Mkln$_{1-735}$: M9, α1: G1, α1Δ: G5, α2: G2, α2Δ: G6, α3: G3, α3Δ: G7, α5: G4, α5Δ: G8)

the intracellular domains in which the last 13 amino acids were deleted were included in the analysis (figure 3.40).

For GABA$_A$R α1, the binding was detected as before by a clear shift of the muskelin band in the presence of the full-length intracellular GABA$_A$R α1 loop, and a faint retention of GABA$_A$R α1. As opposed to this, only a marginal shift of the muskelin band was detectable in presence of the deletion variant. A similarly clear shift in presence of the full intracellular loop was observed for the α2 and α5 subunits, while the deletion variants caused only a minor shift of the muskelin band. For the α3 subunit a very weak shift of the muskelin band was detected, with no difference between the full intracelluar loop and the deletion variant.

These results suggest that muskelin *in vitro* is able to bind not only to the α1, but also to the α2 and α5 subunit, and the vital contributions to this binding reside in the C-terminal portion of the intracellular loops. No binding of the α3 subunit was detected. The results were reproduced in independent experiments, hence the impaired binding of the α3 subunit is either an intrinsic experimental error, or it demonstrates that

subtle differences in the main binding region have a major impact or that further critical determinants of specificity reside in other parts of the intracellular loop.

In the future, further binding studies are planned that will help to quantify the differences in affinity of muskelin for the different GABA$_A$ receptor subunits. In addition, a systematic mutational analysis using a peptide array will provide valuable information about the determinants of binding. A co-crystal structure of muskelin with a GABA$_A$ receptor-derived peptide would also give important insights. The rather low affinity of the GABA$_A$ receptor subunits to muskelin and the lack of a muskelin construct that is amenable to crystallization and additionally has a high affinity to the GABA$_A$ receptor are weak points of this approach, but on the other hand, cocrystallization with a ligand might induce the conformational rigidity needed to facilitate crystallization of constructs that proved recalcitrant so far (discussed in more detail in section 3.2.5 on page 80).

The corroboration that main determinants of binding reside in the very C-terminal part of the cytoplasmic domain of the GABA$_A$ receptor allows one to draw tentative conclusions about details of the interaction with muskelin. Although for the GABA$_A$ receptor α1 subunit, neither a structure nor data on the functional relevance is available for this C-terminal part of the cytoplasmic domain (termed MA, for membrane associated, region), studies on homologous subunits and receptors provide important insights. The crystal structures of the *Torpedo marmorata* nicotinic acetylcholine receptor (Unwin, 2005) and the mouse serotonin 5-HT$_3$ receptor (Hassaine et al., 2014) show that the MA region forms a long, curved α-helix that precedes as a continuous extension of the fourth transmembrane helix and shapes the intracellular vestibule of the channel (figure 3.41). In terms of channel function, the MA region has been shown to influence gating kinetics and channel conductance in the cation-selective nicotinic acetylcholine and serotonin receptors (Peters et al., 2010) and the conductance of the anion-selective glycine receptor (Carland et al., 2009). For GABA$_A$ receptors, studies on the MA region have mainly focussed on the γ2 subunit, in which it constitutes the binding site for the GABA$_A$ receptor associated protein GABARAP and is implied to contribute to receptor clustering by self-associating (Nymann-Andersen et al., 2002) as well as to mediate the modulation of channel conductance upon clustering (Everitt et al., 2009). Secondary structure predictions for GABA$_A$ receptor subunits do not predict an extended α-helix in the MA region comparable to the MA helices in the acetylcholine and the serotonin receptor. However, NMR studies of peptides derived from the γ2 and δ subunits (corresponding to the last 22 respectively 30 residues of the intracellular loops) demonstrated that they adopt a mainly helical fold (Seymour et al., 2012). It has been hypothesized that this helical fold is stabilized when incorporated in the protein, and that the MA helix is structurally conserved in other GABA$_A$ receptor subunits, too. Notably, the α1 subunit might deviate from this to
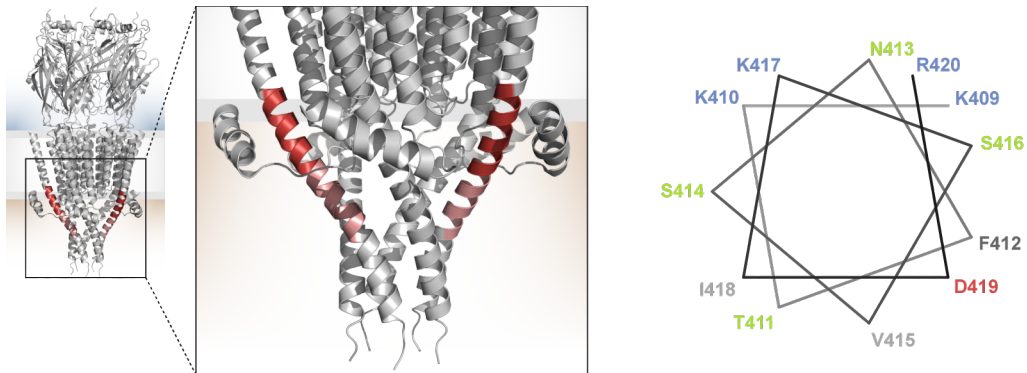
Figure 3.41 – Muskelin binding site in the GABA_A receptor illustrated by coloring the equivalent regions (residues sufficient for muskelin binding in light red, necessary residues in dark red, see figure 3.40) in the crystal structure of the mouse serotonin 5-HT₃ receptor (*left*), shown in side view and in a close-up view of the intracellular domain. *Right*: the critical binding region (residues 409-420) is shown as helical wheel projection.

some extent. While the very C-terminal residues of the MA region, which correspond to the muskelin binding motif, are predicted to form an α-helix, an upstream series of five prolines in close succession indicates that the putative MA helix is either strongly kinked, discontinuous or much shorter than in the known structures.

Hence, muskelin seems to recognize a motif with a distinct three-dimensional structure rather than a simple, unfolded linear peptide. In that view, it is interesting to note that an α-helix formed by the critical binding site would exhibit very different faces: all positively charged residues would fall on one face of the helix, while all hydrophobic residues would fall on the opposite site (figure 3.41). Moreover, the recognition of a motif with a defined structure allows one to speculate on a mechanism of subunit specificity, which is difficult to explain based on the highly conserved primary sequence. The series of prolines is not particularly conserved among the α subunits, especially the triplet of Pro405, Pro406 and Pro408, directly preceding the highly conserved motif, is found only in the α1 subunit. The distortion of the helical structure thus is predicted to be most pronounced in the α1 subunit. Therefore it is well conceivable that the parts adjacent to the critical binding interface confer subunit specificity by imposing sterical restraints. Future mutational studies on the determinants of binding need to carefully consider putative effects on the secondary and tertiary structures in order to gain sound information.

But the special location of the binding site can give some further indications. One the one hand it is an interesting coincidence that muskelin and GABARAP bind to corresponding sites on the respective receptor subunit, and both are important regulators of receptor trafficking. It might be worthwhile to consider an interplay between these

two binding partners to the GABA$_A$ receptor or further parallels in the mode of action of both. On the other hand, the binding site resides in an element that is putatively implied in channel gating movements, as emphasized by the recent mouse serotonin receptor structure, in which the intracellular vestibule formed by the MA helices is in fact closed (Hassaine et al., 2014). This raises the question if the functional state of the receptor might be a relevant factor for muskelin binding, that is, if muskelin binding is sensitive to conformational changes in the receptor. A further hint can be found in the crystal structure of the GABA$_A$R β3 homopentameric channel (Miller and Aricescu, 2014). Here, Asp424, which corresponds to the last residue in the muskelin binding motif (Asp419 in the α1 subunit), is seen as part of the hydrogen bonding network at the back of the second transmembrane helix at the narrowest constriction of the channel, the gate. The gate is compressed shut in this structure, which was suggested to reflect the desensitized state of the channel. Being part of the hydrogen bonding network at the gate, the exact position of Asp424 likely depends on the channel state, and could in turn influence the binding interface for muskelin. A conformation-sensitive binding of muskelin would offer a simple mechanism for a usage-dependent internalization of the receptor. Several studies have addressed this topic (Naylor et al., 2005; Goodkin et al., 2008; Saito et al., 2009; Chaumont et al., 2013) and, although results are somewhat conflicting, particularly the most recent study implies a direct link of receptor endocytosis and intracellular trafficking to agonist-induced conformational changes (Chaumont et al., 2013).

OUTLOOK

The aim of this thesis was the structural and biochemical characterization of muskelin and its interaction with the $GABA_A$ receptor. While the previous chapter presented the advancements that were accomplished in this work, some ends remain to be tied up in the future, and potential strategies that incorporate the findings of this work will be discussed in the following.

A crystal structure of a dimer of the muskelin fragment encompassing its discoidin domain and LisH motif could be solved and provided valuable insights, but further high resolution structural information is still missing. For the aim of completing the structural analysis, two main strategies emerge from the work accomplished in this thesis. The first is to follow up on the crystallization of the $Mkln_{1-735}$ N144R variant and to improve the diffraction quality of the crystals in further rounds of optimization. In parallel to exploring the potential of manipulations of the existing crystals, such as dehydration and soaking (Heras and Martin, 2005), it is probably worthwhile to attempt to adjust the processes of nucleation and crystal growth, too. The crystals grew rather fast (1 to 3 days), and controlling the kinetics of crystallization potentially will potentially improve the order and diffractive properties of the crystals, in particular by slowing down crystal growth. Options to consider here include placing a layer of oil over the reservoir, as barrier reducing evaporation and hence the rate of drop equilibration, or dynamically adapting the growth conditions by exchanging the reservoir condition during the period of crystal growth (Chayen, 2004). The second strategy to advance the structural characterization would be to stick to the 'divide and conquer' approach of crystallizing and solving the structure of shortened fragments. To cover the missing parts by applying the findings of this work, a dual strategy of exploiting the putative advantages of deletions of the kelch repeat domain while addressing the specific challenges in the crystallization of the kelch repeat domain separately could be pursued. If the hypothesis of a coherent unit formed by the CTLH motif and the C-terminal module holds true, and corresponding fusion constructs indeed exhibit the

stable fold needed to isolate and characterize them, they are even suitable to be analyzed by NMR in addition to X-ray crystallography. The structural characterization of the kelch repeat domain at this point seems to pose the greater challenges, which might hopefully be overcome by careful construct design and by employing measurements to reduce conformational heterogeneity, such as cocrystallization with ligands, mutational surface entropy reduction or protein modification (discussed in more detail in section 3.2.5). With respect to construct design, taking into account the N-terminal closure mechanism revealed by the detailed analyses of the domain structure presented in section 3.1.2 proved to be very beneficial and is an important detail that will improve the future design of constructs.

The biochemical analysis of muskelin oligomerization presented here has answered many of the initial questions, and added LisH-dimerization and its fundamental role in determining cellular localization as new aspect of muskelin oligomerization. What is missing to complete the picture is a better understanding of the fundamental role of the C-terminal module, and the experimental validation of the model of the tetramer architecture. For the first point, a high resolution X-ray or NMR study of a muskelin fragment harboring both the C-terminal module and the LisH-motif, as discussed above, would be highly informative. For the latter, the SAXS analysis of the full-length protein should be improved by a more careful sample preparation, if applicable complemented with experimental data of distance restraints in the tetramer, and expanded for the analyses of the oligomerization deficient variants, which would give an impression of the individual parts of the puzzle.

The foremost question left open by the analyses of oligomerization, however, is the cell biological basis of the coupling of LisH-dimerization and nuclear import. In a first step, it would be very informative to assess the oligomeric state of muskelin directly in the cell. Here, a combination of biochemical and cell biological assays could again take full effect, by validating the reporters of oligomerization (such as FRET sensors, or chemical crosslinkers) in biochemical analysis, making use of the available protein variants with defined oligomeric states, while cell biological assays and microscopic techniques would address the superordinate question of where in the cell and under which conditions muskelin adopts different oligomeric states.

Finally, the question of how the nuclear localization is mediated requires elaborate cell biological studies that systematically interfere with the individual pathways of nuclear transport. To elucidate the molecular basis of localization, biochemical studies will have to distinguish whether a localization signal in muskelin, which is modified or masked either by an interaction of muskelin with itself or with another protein, or the contribution of a localization signal by an interaction partner underly the observed phenomena. Although the data presented in this work in conjunction with earlier studies hint towards a nuclear localization activity of the LisH motif, which is

masked by muskelin oligomerization, this will be difficult to demonstrate unequivocally. The NLS activity, the masking interactions and the overall oligomerization are likely correlated directly, and cannot be addressed individually, which interferes with important controls. The alternative import pathway, a cotransport with a binding partner (the obvious candidate here being RanBP9) that only binds to either of muskelin's oligomeric states, can probably be addressed more directly, if a detailed analysis of the interaction identifies mutations that interfere with partner binding, but not LisH dimerization.

The major findings of the analysis of the interaction of muskelin and the GABA$_A$ receptor were one the one hand that muskelin is able to bind not only to the α1 subunit, but also to the α2 and α5 subunits, albeit probably with reduced affinity, and on the other hand, that critical determinants of this binding reside in the very C-terminal part of the intracellular loop of the receptor, which is likely organized in the MA helix. The interesting notion that the local architecture of the MA helix might influence the binding and thereby confer subunit specificity could be further explored by a mutational analysis of the binding site and the surrounding parts, particularly targeting the abundant prolines in the α1 subunit, and monitoring both the impact on binding to muskelin in ITC experiments and the structure of the MA helix using CD and NMR spectroscopy. Such studies would also provide valuable data for the design of a suitable ligand for cocrystallization. In the design of derivative peptide ligands, a stabilization of the helical fold should be considered, as it is to be expected that the fold of an isolated peptide is significantly less stable than that of an integral part of the protein. This stabilization could be managed for example by using stapled peptides (Walensky and Bird, 2014), and has the potential to improve the binding and reduce conformational heterogeneity of the GABA$_A$ receptor derived ligand.

Beyond the characterization of the interaction with the GABA$_A$ receptor, a further line of investigations would be to employ a similarly detailed biochemical analysis as initiated here on other interaction partners of muskelin. One could even take this approach further by not only characterizing reported partners, but also screening for new partners using suitable techniques like a tandem approach of co-immunoprecipitation and mass spectrometry, and including the new partners in the analysis. So far, the identification of an interaction to muskelin was initiated in all cases from the side of the counterpart, and a yeast two-hybrid screen reported by Valiyaveettil et al. (2008) only identified the already known interaction to RanBP9. Given the fact that the majority of the interactions of muskelin seems to be mediated, at least in parts, by its kelch repeat domain (see Introduction, section 1.2.4, figure 1.7), which was also found for the GABA$_A$ receptor in this work, one might speculate that the interaction partners might fall into two classes based on their binding site: the targets, such as the α1 subunit as cargo, binding to the kelch repeat domain, and the effectors, such as the dynein

complex via DIC1A or the E3 Ligase in the CTLH complex via RanBP9, which bind to other parts of the protein. Although this classification might prove a far too simplified concept, it would be interesting to explore it by fine-mapping of the binding sites and analyzing the interdependencies of the partners, distinguishing independent, competing and cooperative binding events. With single mutations that specifically abolish the binding of muskelin to the respective interaction partners, it would be possible to probe the connection of the individual binding events to the different functions of muskelin. In particular, the interplay between the binding to RanBP9 and that to other partners, which are putative substrates to the CTLH complex, could reveal a potential link of the participation in an E3 Ligase to the other roles of muskelin.

BIBLIOGRAPHY

D.W. Abbott, S. Hrynuik, and A.B. Boraston. Identification and characterization of a novel periplasmic polygalacturonic acid binding protein from *Yersinia enterolitica*. *J Mol Biol*, 367(4):1023–1033, April 2007.

J.C. Adams. Characterization of a *Drosophila melanogaster* orthologue of muskelin. *Gene*, 297(1-2):69–78, September 2002.

J.C. Adams, B. Seed, and J. Lawler. Muskelin, a novel intracellular mediator of cell adhesive and cytoskeletal responses to thrombospondin-1. *EMBO J*, 17(17):4964–4974, September 1998.

J.C. Adams, R. Monk, A.L. Taylor, S. Ozbek, N. Fascetti, S. Baumgartner, and J. Engel. Characterisation of *Drosophila* thrombospondin defines an early origin of pentameric thrombospondins. *J Mol Biol*, 328(2):479–494, April 2003.

P.D. Adams, R.W. Grosse-Kunstleve, L.W. Hung, T.R. Ioerger, A.J. McCoy, N.W. Moriarty, R.J. Read, J.C. Sacchettini, N.K. Sauter, and T.C. Terwilliger. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr*, 58(Pt 11):1948–1954, November 2002.

S.F. Altschul, W. Gish, W. Miller, E.W. Myers, and D.J Lipman. Basic Local Alignment Search Tool. *J Mol Biol*, 215(3):403–410, October 1990.

I.L. Arancibia-Cárcamo, E.Y. Yuen, J. Muir, M.J. Lumb, G. Michels, R.S. Saliba, T.G. Smart, Z. Yan, J.T. Kittler, and S.J. Moss. Ubiquitin-dependent lysosomal targeting of GABA$_A$ receptors regulates neuronal inhibition. *Proc Natl Acad Sci USA*, 106(41):17552–17557, October 2009.

L.C. Armstrong and P. Bornstein. Thrombospondins 1 and 2 function as inhibitors of angiogenesis. *Matrix Biol*, 22(1):63–71, March 2003.

F.A.C. Azevedo, L.R.B. Carvalho, L.T. Grinberg, J.M. Farfel, R.E.L. Ferretti, R.E.P. Leite, W. Jacob Filho, R. Lent, and S. Herculano-Houzel. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *J Comp Neurol*, 513(5):532–541, April 2009.

H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne. The Protein Data Bank. *Nucl Acids Res*, 28(1):235–242, January 2000.

N. Bocquet, H. Nury, M. Baaden, C. Le Poupon, J.-P. Changeux, M. Delarue, and J.-P. Corringer. X-ray structure of a pentameric ligand-gated ion channel in an apparently open conformation. *Nature*, 457(7225):111–114, January 2009.

I. Böhme, H. Rabe, and H. Lüddens. Four amino acids in the α subunits determine the γ-aminobutyric acid sensitivities of GABA$_A$ receptor subtypes. *J Biol Chem*, 279 (34):35193–35200, August 2004.

A. Bonnefoy, R. Moura, and M.F. Hoylaerts. Thrombospondins: from structure to therapeutics. *Cell Mol Life Sci*, 65(5):713–727, January 2008.

P. Bork and R.F. Doolittle. *Drosophila kelch* motif is derived from a common enzyme fold. *J Mol Biol*, 236(5):1277–1282, March 1994.

P. Bornstein. Diversity of function is inherent in matricellular proteins: an appraisal of thrombospondin 1. *J Cell Biol*, 130(3):503–506, August 1995.

P. Bornstein, A. Agah, and T.R. Kyriakides. The role of thrombospondins 1 and 2 in the regulation of cell-matrix interactions, collagen fibril formation, and the response to injury. *Int J Biochem Cell Biol*, 36(6):1115–1125, June 2004.

L.M. Boyd, W.J. Richardson, J. Chen, V.B. Kraus, A. Tewari, and L.A. Setton. Osmolarity regulates gene expression in intervertebral disc cells determined by gene array and real-time quantitative RT-PCR. *Ann Biomed Eng*, 33(8):1071–1077, August 2005.

N. Brandon, J. Jovanovic, and S.J. Moss. Multiple roles of protein kinases in the modulation of γ-aminobutyric acid$_A$ receptor function and cell surface expression. *Pharmacol Ther*, 94(1-2):113–122, April 2002.

B. Braun, T. Pfirrmann, R. Menssen, K. Hofmann, H. Scheel, and D.H. Wolf. Gid9, a second RING finger protein contributes to the ubiquitin ligase activity of the Gid complex required for catabolite degradation. *FEBS Lett*, 585(24):3856–3861, December 2011.

G. Bricogne, C. Vonrhein, C. Flensburg, M. Schiltz, and W Paciorek. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr D Biol Crystallogr*, 59(Pt 11): 2023–2030, November 2003.

N. Brown, J. Kerby, T.P. Bonnert, P.J. Whiting, and K.A. Wafford. Pharmacological characterization of a novel cell line expressing human $\alpha_4\beta_3\delta$ GABA$_A$ receptors. *Br J Pharmacol*, 136(7):965–974, August 2002.

M.A. Bukowska and M.G. Grütter. New concepts and aids to facilitate crystallization. *Curr Opin Struct Biol*, 23(3):409–416, June 2013.

P. Canning, C.D.O. Cooper, T. Krojer, J.W. Murray, A.C.W. Pike, A. Chaikuad, T. Keates, C. Thangaratnarajah, V. Hojzan, B.D. Marsden, O. Gileadi, S. Knapp, F von Delft, and A.N. Bullock. Structural basis for Cul3 protein assembly with the BTB-Kelch family of E3 ubiquitin ligases. *J Biol Chem*, 288(11):7803–7814, March 2013.

F. Carafoli, D. Bihan, S. Stathopoulos, A.D. Konitsiotis, M. Kvansakul, R.W. Farndale, B. Leitinger, and E. Hohenester. Crystallographic insight into collagen recognition by discoidin domain receptor 2. *Structure*, 17(12):1573–1581, December 2009.

J.E. Carland, M.A. Cooper, S. Sugiharto, H.-J. Jeong, T.M. Lewis, P.H. Barry, J.A. Peters, J.J. Lambert, and A.J. Moorhouse. Characterization of the effects of charged residues in the intracellular loop on ion permeation in α1 glycine receptor channels. *J Biol Chem*, 284(4):2023–2030, January 2009.

S. Chaumont, C. Andre, D. Perrais, E. Boue-Grabot, A. Taly, and M. Garret. Agonist-dependent endocytosis of γ-aminobutyric acid type A (GABA$_A$) receptors revealed by a γ2(R43Q) Epilepsy Mutation. *J Biol Chem*, 288(39):28254–28265, September 2013.

N.E. Chayen. Turning protein crystallisation from an art into a science. *Curr Opin Struct Biol*, 14(5):577–583, October 2004.

H. Chen, M.E. Herndon, and J. Lawler. The cell biology of thrombospondin-1. *Matrix Biol*, 19(7):597–614, December 2000.

H.L. Chen and H.X. Zhou. Prediction of interface residues in protein-protein complexes by a consensus neural network method: Test against NMR data. *Proteins*, 61(1):21–35, October 2005.

Z.-W. Chen and R.W. Olsen. GABA$_A$ receptor associated proteins: a key factor regulating GABA$_A$ receptor function. *J Neurochem*, 100(2):279–294, January 2007.

The UniProt Consortium. Ongoing and future developments at the Universal Protein Resource. *Nucl Acids Res*, 39(Database issue):D214–D219, January 2011.

K. Cowtan. The *Buccaneer* software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr*, 62(Pt 9):1002–1011, September 2006.

A. Crider, C.D. Pandya, D. Peter, A.O. Ahmed, and A. Pillai. Ubiquitin-proteasome dependent degradation of GABA$_A\alpha$1 in autism spectrum disorder. *Mol Autism*, 5 (45):1–10, September 2014.

C. Daniel and M. Öhman. RNA editing and its impact on GABA$_A$ receptor function. *Biochem Soc Trans*, 37(Pt 6):1399–1403, December 2009.

I.W. Davis, L.W. Murray, J.S. Richardson, and D.C. Richardson. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucl Acids Res*, 32(Web Server issue):W615–9, July 2004.

P. Debenedittis, C. Harmelink, Y. Chen, Q. Wang, and K. Jiao. Characterization of the novel interaction between muskelin and TBX20, a critical cardiogenic transcription factor. *Biochem Biophys Res Commun*, 409(2):338–343, June 2011.

Z.S. Derewenda and P.G. Vekilov. Entropy and surface engineering in protein crystallization. *Acta Crystallogr D Biol Crystallogr*, D62:116–124, December 2005.

R. Dhavan and L.H. Tsai. A decade of CDK5. *Nat Rev Mol Cell Biol*, 2(10):749–759, October 2001.

V.K. Dhodda, K.A. Sailor, K.K. Bowen, and R. Vemuganti. Putative endogenous mediators of preconditioning-induced ischemic tolerance in rat brain identified by genomic and proteomic analysis. *J Neurochem*, 89(1):73–89, April 2004.

D.A. Drachman. Do we have brain to spare? *Neurology*, 64(12):2004–2005, June 2005.

A. Dümmler, A.-M. Lawrence, and Ario de Marco. Simplified screening for the detection of soluble fusion constructs expressed in *E. coli* using a modular set of vectors. *Microb Cell Fact*, 4(1):34, 2005.

Sean R Eddy. *HMMER User's Guide*. Howard Hughes Medical Institute, HMMER Development Team, Janelia Farm Research Campus, 19700 Helix Drive, Ashburn VA 20147 USA, March 2010.

R.D. Emes and C.P. Ponting. A new sequence motif linking lissencephaly, Treacher Collins and oral-facial-digital type 1 syndromes, microtubule dynamics and cell migration. *Hum Mol Genet*, 10(24):2813–2820, November 2001.

P. Emsley, B. Lohkamp, W.G. Scott, and K Cowtan. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*, 66(Pt 4):486–501, April 2010.

U.B. Ericsson, B.M. Hallberg, G.T. DeTitta, N. Dekker, and P. Nordlund. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal Biochem*, 357(2):289–298, 2006.

P. Evans. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr*, 62(Pt 1):72–82, January 2006.

A.B. Everitt, V.A.L. Seymour, J. Curmi, D.R. Laver, P.W. Gage, and M.L. Tierney. Protein interactions involving the $\gamma 2$ large cytoplasmic loop of GABA$_A$ receptors modulate conductance. *FASEB J*, 23(12):4361–4369, November 2009.

R. Fabrini, A. De Luca, L. Stella, G. Mei, B. Orioni, S. Ciccone, G. Federici, M. Lo Bello, and G. Ricci. Monomer–dimer equilibrium in glutathione transferases: a critical re-examination. *Biochemistry*, 48(43):10473–10482, November 2009.

H.-J. Feng and R.L. Macdonald. Multiple actions of propofol on $\alpha\beta\gamma$ and $\alpha\beta\delta$ GABA$_A$ receptors. *Mol Pharmacol*, 66(6):1517–1524, December 2004.

R.D. Finn, J. Clements, and S.R. Eddy. HMMER web server: interactive sequence similarity searching. *Nucl Acids Res*, 39(Web Server issue):W29–W37, July 2011.

R.D. Finn, A. Bateman, J. Clements, P. Coggill, R.Y. Eberhardt, S.R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E.L.L. Sonnhammer, J. Tate, and M. Punta. Pfam: the protein families database. *Nucl Acids Res*, 42(D1):D222–D230, January 2014.

O. Francis, F. Han, and J.C. Adams. Molecular phylogeny of a RING E3 ubiquitin ligase, conserved in eukaryotic cells and dominated by homologous components, the muskelin/RanBPM/CTLH complex. *PLoS ONE*, 8(10):e75217, October 2013.

D. Franke and D.I. Svergun. *DAMMIF*, a program for rapid *ab-initio* shape determination in small-angle scattering. *J Appl Crystallogr*, 42:342–346, April 2009.

T. Fukutomi, K. Takagi, T. Mizushima, N. Ohuchi, and M. Yamamoto. Kinetic, thermodynamic, and structural characterizations of the association between Nrf2-DLGex degron and Keap1. *Mol Cell Biol*, 34(5):832–846, March 2014.

J.L. Galzi, A. Devillers-Thiéry, N. Hussy, S. Bertrand, J.P. Changeux, and D. Bertrand. Mutations in the channel domain of a neuronal nicotinic receptor convert ion selectivity from cationic to anionic. *Nature*, 359(6395):500–505, October 1992.

G. Gerlitz, E. Darhin, G. Giorgio, B. Franco, and O. Reiner. Novel functional features of the Lis-H domain: role in protein dimerization, half-life and cellular localization. *Cell Cycle*, 4(11):1632–1640, November 2005.

L. Goldschmidt, D.R. Cooper, Z.S. Derewenda, and D. Eisenberg. Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. *Protein Sci*, 16(8):1569–1576, August 2007.

H.P. Goodkin, S. Joshi, Z. Mtchedlishvili, J. Brar, and J. Kapur. Subunit-specific trafficking of GABA$_A$ receptors during status epilepticus. *J Neurosci*, 28(10):2527–2538, March 2008.

H.L. Grabenstatter, S.J. Russek, and A.R. Brooks-Kayal. Molecular pathways controlling inhibitory receptor expression. *Epilepsia*, 53(Suppl 9):71–78, December 2012.

C.H. Gray, L.C. McGarry, H.J. Spence, A. Riboldi-Tunnicliffe, and B.W. Ozanne. Novel β-propeller of the BTB-Kelch protein Krp1 provides a binding site for Lasp-1 that is necessary for pseudopodial extension. *J Biol Chem*, 284(44):30498–30507, October 2009.

G. Gueron, J. Giudice, P. Valacco, A. Paez, B. Elguero, M. Toscani, F. Jaworski, F.C. Leskow, J. Cotignola, M. Marti, M. Binaghi, N. Navone, and E. Vazquez. Heme-oxygenase-1 implications in cell morphology and the adhesive behavior of prostate cancer cells. *Oncotarget*, 5(12):4087–4102, June 2014.

N. Hájos, Z. Nusser, E.A. Rancz, T.F. Freund, and I. Mody. Cell type- and synapse-specific variability in synaptic GABA$_A$ receptor occupancy. *European J Neurosci*, 12 (3):810–818, March 2000.

M. Hämmerle, J. Bauer, M. Rose, A. Szallies, M. Thumm, S. Düsterhus, D. Mecke, K.D. Entian, and D.H. Wolf. Proteins of newly isolated mutants and the amino-terminal proline are essential for ubiquitin-proteasome-catalyzed catabolite degradation of fructose-1,6-bisphosphatase of *Saccharomyces cerevisiae*. *J Biol Chem*, 273 (39):25000–25005, September 1998.

H. Hasegawa, H. Katoh, H. Fujita, K. Mori, and M. Negishi. Receptor isoform-specific interaction of prostaglandin EP3 receptor with muskelin. *Biochem Biophys Res Commun*, 276(1):350–354, September 2000.

G. Hassaine, C. Deluz, L. Grasso, R. Wyss, M.B. Tol, R. Hovius, A. Graff, H. Stahlberg, T. Tomizaki, A. Desmyter, C. Moreau, X.-D.. Li, F. Poitevin, H. Vogel, and H. Nury. X-ray structure of the mouse serotonin 5-HT$_3$ receptor. *Nature*, 512(7514):276–281, August 2014.

F.F. Heisler, S. Loebrich, Y. Pechmann, N. Maier, A.R. Zivkovic, M. Tokito, T.J. Hausrat, M. Schweizer, R. Bähring, E.L.F. Holzbaur, D. Schmitz, and M. Kneussel. Muskelin regulates actin filament- and microtubule-based GABA$_A$ receptor transport in neurons. *Neuron*, 70(1):66–81, April 2011.

B. Heras and J.L. Martin. Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallogr D Biol Crystallogr*, 61(9):1173–1180, August 2005.

M.B. Herd, D. Belelli, and J.J. Lambert. Neurosteroid modulation of synaptic and extrasynaptic GABA$_A$ receptors. *Pharmacol Ther*, 116(1):20–34, October 2007.

R.E. Hibbs and E. Gouaux. Principles of activation and permeation in an anion-selective Cys-loop receptor. *Nature*, 474(7349):54–60, May 2011.

R. Higuchi, B. Krummel, and R.K. Saiki. A general method of *in vitro* preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. *Nucl Acids Res*, 16(15):7351–7367, August 1988.

R.J.C. Hilf and R. Dutzler. X-ray structure of a prokaryotic pentameric ligand-gated ion channel. *Nature*, 452(7185):375–379, March 2008.

S.N. Ho, H.D. Hunt, R.M. Horton, J.K. Pullen, and L.R. Pease. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, 77(1):51–59, April 1989.

Y. Ichimura, S. Waguri, Y.-S. Sou, S. Kageyama, J. Hasegawa, R. Ishimura, T. Saito, Y. Yang, T. Kouno, T. Fukutomi, T. Hoshii, A. Hirao, K. Takagi, T. Mizushima, H. Motohashi, M-S. Lee, T. Yoshimori, K. Tanaka, M. Yamamoto, and M. Komatsu. Phosphorylation of p62 activates the Keap1-Nrf2 pathway during selective autophagy. *Mol Cell*, 51(5):618–631, September 2013.

N. Ito, S.E. Phillips, C. Stevens, Z.B. Ogel, M.J. McPherson, J.N. Keen, K.D. Yadav, and P.F. Knowles. Novel thioether bond revealed by a 1.7 Å crystal structure of galactose oxidase. *Nature*, 350(6313):87–90, March 1991.

N. Ito, S.E. Phillips, K.D. Yadav, and P.F. Knowles. Crystal structure of a free radical enzyme, galactose oxidase. *J Mol Biol*, 238(5):794–814, May 1994.

T.C. Jacob, S.J. Moss, and R. Jurd. GABA$_A$ receptor trafficking and its role in the dynamic modulation of neuronal inhibition. *Nat Rev Neurosci*, 9(5):331–343, May 2008.

M.V. Jones, N.L. Harrison, D.B. Pritchett, and T.G. Hales. Modulation of the GABA$_A$ receptor by propofol is independent of the γ subunit. *J Pharmacol Exp Ther*, 274(2): 962–968, August 1995.

W Kabsch. *XDS*. *Acta Crystallogr D Biol Crystallogr*, 66(Pt 2):125–132, February 2010.

S. Kalinin, T. Peulen, S. Sindbert, P.J. Rothwell, S. Berger, T. Restle, R.S. Goody, H. Gohlke, and C.A.M. Seidel. A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nat Methods*, 9(12):1218–1225, November 2012.

L.A. Kelley and M.J.E. Sternberg. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc*, 4(3):363–371, February 2009.

A. Keramidas, A.J. Moorhouse, C.R. French, P.R. Schofield, and P.H. Barry. M2 pore mutations convert the glycine receptor channel from being anion- to cation-selective. *Biophys J*, 79(1):247–259, July 2000.

A. Kiedzierska, K. Smietana, H. Czepczynska, and J. Otlewski. Structural similarities and functional diversity of eukaryotic discoidin-like domains. *Biochim Biophys Acta*, 1774(9):1069–1078, September 2007.

A. Kiedzierska, H. Czepczynska, K. Smietana, and J. Otlewski. Expression, purification and crystallization of cysteine-rich human protein muskelin in *Escherichia coli*. *Protein Expr Purif*, 60(1):82–88, July 2008.

K.H. Kim, S.K. Hong, K.Y. Hwang, and E.E. Kim. Structure of mouse muskelin discoidin domain and biochemical characterization of its self-association. *Acta Crystallogr D Biol Crystallogr*, 70(Pt 11):2863–2874, November 2014.

M.H. Kim, D.R. Cooper, A. Oleksy, Y. Devedjiev, U. Derewenda, O. Reiner, J. Otlewski, and Z.S. Derewenda. The structure of the N-terminal domain of the product of the lissencephaly gene *Lis1* and its functional implications. *Structure*, 12(6):987–998, June 2004.

N. Kobayashi, J. Yang, A. Ueda, T. Suzuki, K. Tomaru, M. Takeno, K. Okuda, and Y. Ishigatsubo. RanBPM, Muskelin, p48EMLP, p44CTLH, and the armadillo-repeat proteins ARMC8α and ARMC8β are components of the CTLH complex. *Gene*, 396 (2):236–247, July 2007.

M. Komatsu, H. Kurokawa, S. Waguri, K. Taguchi, A. Kobayashi, Y. Ichimura, Y.-S. Sou, I. Ueno, A. Sakamoto, K.I. Tong, M. Kim, Y. Nishito, S.-I. Iemura, T. Natsume, T. Ueno, E. Kominami, H. Motohashi, K. Tanaka, and M. Yamamoto. The selective autophagy substrate p62 activates the stress responsive transcription factor Nrf2 through inactivation of Keap1. *Nat Cell Biol*, 12(3):213–223, February 2010.

P.V. Konarev, V.V. Volkov, A.V. Sokolova, M.H.J. Koch, and D.I. Svergun. *PRIMUS*: a Windows PC-based system for small-angle scattering data analysis. *J Appl Crystallogr*, 36:1277–1282, October 2003.

S.. Kosugi, M. Hasebe, M. Tomita, and H. Yanagawa. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc Natl Acad Sci USA*, 106(25):10171–10176, June 2009.

D. Krause. Biochemische Untersuchungen am Multidomänenprotein Muskelin. Bachelorarbeit im Studienfach Biologie, Julius-Maximilians-Universität Würzburg, 2011.

E. Krissinel and K. Henrick. Inference of macromolecular assemblies from crystalline state. *J Mol Biol*, 372(3):774–797, September 2007.

T. la Cour, L. Kiemer, A. Molgaard, R. Gupta, K. Skriver, and S. Brunak. Analysis and prediction of leucine-rich nuclear export signals. *Protein Eng Des Sel*, 17(6): 527–536, August 2004.

U.K. Laemmli. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227(5259):680–685, August 1970.

J. Lawler and M. Detmar. Tumor progression: the effects of thrombospondin-1 and -2. *Int J Biochem Cell Biol*, 36(6):1038–1045, June 2004.

J. Lawler and R.O. Hynes. The structure of human thrombospondin, an adhesive glycoprotein with multiple calcium-binding sites and homologies with several different proteins. *J Cell Biol*, 103(5):1635–1648, November 1986.

D.R. Ledee, C.Y. Gao, R. Seth, R.N. Fariss, B.K. Tripathi, and P.S. Zelenka. A specific interaction between muskelin and the cyclin-dependent kinase 5 activator p39 promotes peripheral localization of muskelin. *J Biol Chem*, 280(22):21376–21383, March 2005.

A. Leitner, T. Walzthoeni, A. Kahraman, F. Herzog, O. Rinner, M. Beck, and R. Aebersold. Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol Cell Proteomics*, 9(8):1634–1649, August 2010.

A.G.W. Leslie and H.R. Powell. Processing diffraction data with mosflm. In R.J. Read and J.L. Sussman, editors, *Evolving Methods for Macromolecular Crystallography*, volume 245, pages 41–51. Springer Netherlands, 2007.

Ivica Letunic and Peer Bork. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucl Acids Res*, 39(Web Server issue):W475–8, July 2011.

J. Lew, K. Beaudette, C.M. Litwin, and J.H Wang. Purification and characterization of a novel proline-directed protein kinase from bovine brain. *J Biol Chem*, 267(19): 13383–13390, July 1992.

M. Z. Li and S.J. Elledge. Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC. *Nat Methods*, 4(3):251–256, March 2007.

X. Li. Crystal structure of the Kelch domain of human Keap1. *J Biol Chem*, 279(52): 54750–54758, September 2004.

S. Liang, C. Zhang, S. Liu, and Y. Zhou. Protein binding site prediction using an empirical scoring function. *Nucl Acids Res*, 34(13):3698–3707, August 2006.

S.-C. Lo, X. Li, M.T. Henzl, L.J. Beamer, and M. Hannink. Structure of the Keap1:Nrf2 interface provides mechanistic insight into Nrf2 signaling. *EMBO J*, 25(15):3605–3617, August 2006.

S. Loebrich, R. Bähring, T. Katsuno, S. Tsukita, and M. Kneussel. Activated radixin is essential for GABA$_A$ receptor α5 subunit anchoring at the actin cytoskeleton. *EMBO J*, 25(5):987–999, March 2006.

B. Luscher, T. Fuchs, and C.L. Kilpatrick. GABA$_A$ Receptor trafficking-mediated plasticity of inhibitory synapses. *Neuron*, 70(3):385–409, May 2011.

S. Macedo-Ribeiro, W. Bode, R. Huber, M.A. Quinn-Allen, S.W. Kim, T.L. Ortel, G.P. Bourenkov, H.D. Bartunik, M.T. Stubbs, W.H. Kane, and P. Fuentes-Prior. Crystal structures of the membrane-binding C2 domain of human coagulation factor V. *Nature*, 402(6760):434–439, November 1999.

S.V. Mathieu, K.S. Aragão, A. Imberty, and A. Varrot. Discoidin I from Dictyostelium discoideum and Interactions with oligosaccharides: specificity, affinity, crystal structures, and comparison with discoidin II. *J Mol Biol*, 400(3):540–554, July 2010.

A.J. McCoy, R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni, and R.J. Read. *Phaser* crystallographic software. *J Appl Crystallogr*, 40(Pt 4):658–674, August 2007.

R. Menssen, J. Schweiggert, J. Schreiner, D. Kusevic, J. Reuther, B. Braun, and D.H. Wolf. Exploring the topology of the Gid complex, the E3 ubiquitin ligase involved in catabolite-induced degradation of gluconeogenic enzymes. *J Biol Chem*, 287(30): 25602–25614, July 2012.

A. Mikolajka, X. Yan, G.M. Popowicz, P. Smialowski, E.A. Nigg, and T.A. Holak. Structure of the N-terminal domain of the FOP (FGFR1OP) protein and implications for its dimerization and centrosomal localization. *J Mol Biol*, 359(4):863–875, June 2006.

P.S. Miller and A.R. Aricescu. Crystal structure of a human GABA$_A$ receptor. *Nature*, 512(7514):270–275, August 2014.

H. Moehler. GABA$_A$ receptors in central nervous system disease: anxiety, epilepsy, and insomnia. *J Recept Signal Transduct Res*, 26(5-6):731–740, 2006.

J. Mukherjee, K. Kretschmannova, G. Gouzer, H.-M. Maric, S. Ramsden, V. Tretter, K. Harvey, P.A. Davies, A. Triller, H. Schindelin, and S.J. Moss. The residence time of GABA$_A$Rs at inhibitory synapses is determined by direct binding of the receptor α1 subunit to gephyrin. *J Neurosci*, 31(41):14677–14687, October 2011.

C.H. Na, D.R. Jones, Y. Yang, X. Wang, Y. Xu, and J. Peng. Synaptic protein ubiquitination in rat brain revealed by antibody-based ubiquitome analysis. *J Proteome Res*, 11(9):4722–4732, September 2012.

D.E. Naylor, H. Liu, and C.G. Wasterlain. Trafficking of GABA$_A$ receptors, loss of inhibition, and a mechanism for pharmacoresistance in status epilepticus. *J Neurosci*, 25(34):7724–7733, August 2005.

H. Neuvirth, R. Raz, and G. Schreiber. ProMate: a structure based prediction program to identify the location of protein-protein binding sites. *J Mol Biol*, 338(1):181–199, April 2004.

S.L. Newstead, J.N. Watson, A.J. Bennet, and G. Taylor. Galactose recognition by the carbohydrate-binding module of a bacterial sialidase. *Acta Crystallogr D Biol Crystallogr*, 61(Pt 11):1483–1491, November 2005.

A.N. Nguyen Ba, A. Pogoutse, N. Provart, and A.M. Moses. NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. *BMC Bioinformatics*, 10(1):202, 2009.

H. Nishitani, E. Hirose, Y. Uchimura, M. Nakamura, M. Umeda, K. Nishii, N. Mori, and T. Nishimoto. Full-sized *RanBPM* cDNA encodes a protein possessing a long stretch

of proline and glutamine within the N-terminal region, comprising a large protein complex. *Gene*, 272(1-2):25–33, July 2001.

J. Nymann-Andersen, G.W. Sawyer, and R.W. Olsen. Interaction between GABA$_A$ receptor subunit intracellular loops: implications for higher order complex formation. *J Neurochem*, 83(5):1164–1171, December 2002.

J. Oberoi, L. Fairall, P.J. Watson, J.-C. Yang, Z. Czimmerer, T. Kampmann, B.T. Goult, J.A. Greenwood, J.T. Gooch, B.C. Kallenberger, L. Nagy, D. Neuhaus, and J.W.R. Schwabe. Structural basis for the assembly of the SMRT/NCoR core transcriptional repression machinery. *Nat Struct Mol Biol*, 18(2):177–184, January 2011.

R.W. Olsen and W. Sieghart. International Union of Pharmacology. LXX. Subtypes of γ-aminobutyric acid$_A$ receptors: classification on the basis of subunit composition, pharmacology, and function. Update. *Pharmacol Rev*, 60(3):243–260, September 2008.

B. Padmanabhan, K.I. Tong, T. Ohta, Y. Nakamura, M. Scharlock, M. Ohtsuji, M.-I. Kang, A. Kobayashi, S. Yokoyama, and M. Yamamoto. Structural basis for defects of Keap1 activity provoked by its point mutations in lung cancer. *Mol Cell*, 21(5): 689–700, March 2006.

B. Padmanabhan, Y. Nakamura, and S. Yokoyama. Structural analysis of the complex of Keap1 with a prothymosin α peptide. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, 64(4):233–238, April 2008.

J.A. Peters, M.A. Cooper, J.E. Carland, M.R. Livesey, T.G. Hales, and J.J. Lambert. Novel structural determinants of single channel conductance and ion selectivity in 5-hydroxytryptamine type 3 and nicotinic acetylcholine receptors. *J Physiol*, 588 (4):587–596, February 2010.

M.V. Petoukhov, D. Franke, A.V. Shkumatov, G. Tria, A.G. Kikhney, M. Gajda, C. Gorba, H.D.T. Mertens, P.V. Konarev, and D.I. Svergun. New developments in the *ATSAS* program package for small-angle scattering data analysis. *J Appl Crystallogr*, 45(Pt 2):342–350, April 2012.

I.K.H. Poon and D.A. Jans. Regulation of nuclear transport: central role in development and transformation? *Traffic*, 6(3):173–186, January 2005.

E. Potterton, P. Briggs, M. Turkenburg, and E. Dodson. A graphical user interface to the *CCP4* program suite. *Acta Crystallogr D Biol Crystallogr*, 59(Pt 7):1131–1137, July 2003.

S. Prag, G.D.M. Collett, and J.C. Adams. Molecular analysis of muskelin identifies a conserved discoidin-like domain that contributes to protein self-association. *Biochem J*, 381(Pt 2):547–559, July 2004.

S. Prag, A. De Arcangelis, E. Georges-Labouesse, and J.C. Adams. Regulation of post-translational modifications of muskelin by protein kinase C. *Int J Biochem Cell Biol*, 39(2):366–378, 2007.

K.P. Pratt, B.W. Shen, K. Takeshima, E.W. Davie, K. Fujikawa, and B.L. Stoddard. Structure of the C2 domain of human factor VIII at 1.5 Å resolution. *Nature*, 402 (6760):439–442, November 1999.

S. Qin and H.-X. Zhou. meta-PPISP: a meta web server for protein-protein interaction site prediction. *Bioinformatics*, 23(24):3386–3387, December 2007.

J.B. Rannes, A. Ioannou, S.C. Willies, G. Grogan, C. Behrens, S.L. Flitsch, and N.J. Turner. Glycoprotein labeling using engineered variants of galactose oxidase obtained by directed evolution. *J Am Chem Soc*, 133(22):8436–8439, June 2011.

J. Regelmann, T. Schüle, F.S. Josupeit, J. Horak, M. Rose, K.-D. Entian, M. Thumm, and D.H. Wolf. Catabolite degradation of fructose-1,6-bisphosphatase in the yeast *Saccharomyces cerevisiae*: a genome-wide screen identifies eight novel GID genes and indicates the existence of two degradation pathways. *Mol Biol Cell*, 14(4): 1652–1663, April 2003.

W.C. Risher and C. Eroglu. Thrombospondins as key regulators of synaptogenesis in the central nervous system. *Matrix Biol*, 31(3):170–177, April 2012.

A. Roy, A. Kucukural, and Y. Zhang. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc*, 5(4):725–738, March 2010.

R. Roy, S. Hohng, and T. Ha. A practical guide to single-molecule FRET. *Nat Methods*, 5(6):507–516, June 2008.

M. Saito, H. Toyoda, H. Sato, H. Ishii, and Y. Kang. Rapid use-dependent down-regulation of γ-aminobutyric acid type A receptors in rat mesencephalic trigeminal neurons. *J Neurosci Research*, 87(14):3120–3133, November 2009.

R.S. Saliba, G. Michels, T.C. Jacob, M.N. Pangalos, and S.J. Moss. Activity-Dependent ubiquitination of $GABA_A$ receptors regulates their accumulation at synaptic sites. *J Neurosci*, 27(48):13341–13351, November 2007.

O. Santt, T. Pfirrmann, B. Braun, J. Juretschke, P. Kimmig, H. Scheel, K. Hofmann, M. Thumm, and D.H. Wolf. The yeast GID complex, a novel ubiquitin ligase (E3) involved in the regulation of carbohydrate metabolism. *Mol Biol Cell*, 19(8):3323–3333, August 2008.

S. Sauer. Biochemische und strukturelle Untersuchungen der CTLH-Untereinheiten RanBPM und Muskelin. Bachelorarbeit im Studienfach Biochemie, Julius-Maximilians-Universität Würzburg, 2014.

N.C. Saxena and R.L. Macdonald. Properties of putative cerebellar γ-aminobutyric acid$_A$ receptor isoforms. *Mol Pharmacol*, 49(3):567–579, March 1996.

P. Schuck. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J*, 78(3):1606–1619, March 2000.

T. Schüle, M. Rose, K.D. Entian, M. Thumm, and D.H. Wolf. Ubc8p functions in catabolite degradation of fructose-1, 6-bisphosphatase in yeast. *EMBO J*, 19(10): 2161–2167, May 2000.

J. Schultz, F. Milpetz, P. Bork, and C.P. Ponting. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci USA*, 95(11): 5857–5864, May 1998.

F.R. Schumacher, F.J. Sorrell, D.R. Alessi, A.N. Bullock, and T. Kurz. Structural and biochemical characterization of the KLHL3-WNK kinase interaction important in blood pressure regulation. *Biochem J*, 460(2):237–246, May 2014.

L.A. Schwarz and G.N. Patrick. Ubiquitin-dependent endocytosis, trafficking and turnover of neuronal membrane proteins. *Mol Cell Neurosci*, 49(3):387–393, March 2012.

E. Severi, A. Muller, J.R. Potts, A. Leech, D. Williamson, K.S. Wilson, and G.H. Thomas. Sialic acid mutarotation is catalyzed by the *Escherichia coli* β-propeller protein YjhT. *J Biol Chem*, 283(8):4841–4849, February 2008.

V.A.L. Seymour, J.P. Curmi, S.M. Howitt, M.G. Casarotto, D.R. Laver, and M.L. Tierney. Selective modulation of different GABA$_A$ receptor isoforms by diazepam and etomidate in hippocampal neurons. *Int J Biochem Cell Biol*, 44(9):1491–1500, September 2012.

W. Sieghart. Structure and pharmacology of γ-aminobutyric acid$_A$ receptor subtypes. *Pharmacol Rev*, 47(2):181–234, June 1995.

W. Sieghart and G. Sperk. Subunit composition, distribution and function of GABA$_A$ receptor subtypes. *Curr Top Med Chem*, 2(8):795–816, August 2002.

F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soeding, J.D. Thompson, and D.G. Higgins. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*, 7, October 2011.

E. Sigel and M.E. Steinmann. Structure, function, and modulation of GABA$_A$ receptors. *J Biol Chem*, 287(48):40224–40231, November 2012.

J. Simon, H. Wakimoto, N. Fujita, M. Lalande, and E.A. Barnard. Analysis of the set of GABA$_A$ receptor genes in the human genome. *J Biol Chem*, 279(40):41422–41435, September 2004.

B.A. Sosa, A. Rothballer, U. Kutay, and T.U. Schwartz. LINC complexes form by binding of three KASH peptides to domain interfaces of trimeric SUN proteins. *Cell*, 149(5): 1035–1047, May 2012.

K. Suhre and Y.H. Sanejouand. *ElNémo*: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucl Acids Res*, 32(Web Server issue):W610–W614, 2004.

D. Svergun, C. Barberato, and M.H.J. Koch. *CRYSOL* - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Cryst*, pages 1–6, December 1995.

D.I. Svergun. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr*, 25:495–503, August 1992.

D.I. Svergun, M.V. Petoukhov, and M.H.J. Koch. Determination of domain structure of proteins from X-ray solution scattering. *Biophys J*, 80(6):2946–2953, June 2001.

N. Tagnaouti, S. Loebrich, F. Heisler, Y. Pechmann, S. Fehr, A. De Arcangelis, E. Georges-Labouesse, J.C. Adams, and M. Kneussel. Neuronal expression of muskelin in the rodent central nervous system. *BMC Neurosci*, 8(1):28, 2007.

K.I. Tong, B. Padmanabhan, A. Kobayashi, C. Shang, Y. Hirotsu, Si. Yokoyama, and M. Yamamoto. Different electrostatic potentials define ETGE and DLG motifs as hinge and latch in oxidative stress response. *Mol Cell Biol*, 27(21):7511–7521, November 2007.

V. Tretter, B. Kerschner, I. Milenkovic, S.L. Ramsden, J. Ramerstorfer, L. Saiepour, H.-M. Maric, S.J. Moss, H. Schindelin, R.J. Harvey, W. Sieghart, and K. Harvey. Molecular basis of the γ-aminobutyric acid A receptor α3 subunit interaction with the clustering protein gephyrin. *J Biol Chem*, 286(43):37702–37711, October 2011.

V. Tretter, J. Mukherjee, H.-M. Maric, H. Schindelin, W. Sieghart, and S.J. Moss. Gephyrin, the enigmatic organizer at GABAergic synapses. *Front Cell Neurosci*, 6: 23, May 2012.

B.K. Tripathi, D.R. Lowy, and P.S. Zelenka. The Cdk5 activator p39 specifically links muskelin to myosin II and regulates stress fiber formation and actin organization in lens. *Exp Cell Res*, August 2014.

C.-H. Tung, J.-W. Huang, and J.-M. Yang. Kappa-alpha plot derived structural alphabet and BLOSUM-like substitution matrix for rapid search of protein structure database. *Genome Biol*, 8(3):R31, March 2007.

W.B. Turnbull and A.H. Daranas. On the value of c: can low affinity systems be studied by isothermal titration calorimetry? *J Am Chem Soc*, 125(48):14859–14866, December 2003.

M. Uhlén, L. Fagerberg, B.M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, C. Kampf, E. Sjöstedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C.A.-K. Szigyarto, J. Odeberg, D. Djureinovic, J.O. Takanen, S. Hober, T. Alm, P.-H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J.M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, and F. Pontén. Proteomics. Tissue-based map of the human proteome. *Science*, 347(6220):1260419, January 2015.

M. Umeda, H. Nishitani, and T. Nishimoto. A novel nuclear protein, Twa1, and Muskelin comprise a complex with RanBPM. *Gene*, 303:47–54, January 2003.

N. Unwin. Refined structure of the nicotinic acetylcholine receptor at 4 Å resolution. *J Mol Biol*, 346(4):967–989, March 2005.

A.A. Vagin, R.A. Steiner, A.A. Lebedev, L. Potterton, S. McNicholas, F. Long, and G.N. Murshudov. *REFMAC5* dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr*, 60(Pt 12 Pt 1):2184–2195, December 2004.

W.S.J Valdar. Scoring residue conservation. *Proteins*, 48(2):227–241, August 2002.

M. Valiyaveettil, A.A. Bentley, P. Gursahaney, R. Hussien, R. Chakravarti, N. Kureishy, S. Prag, and J.C. Adams. Novel role of the muskelin-RanBP9 complex as a nucleocytoplasmic mediator of cell morphology regulation. *J Cell Biol*, 182(4):727–739, August 2008.

S.J.L. van Wijk, S.J. de Vries, P. Kemmeren, A. Huang, R. Boelens, A.M.J.J. Bonvin, and H.T.M. Timmers. A comprehensive framework of E2-RING E3 interactions of the human ubiquitin-proteasome system. *Mol Syst Biol*, 5:295, August 2009.

C.W. Vander Kooi, M.A. Jusino, B. Perman, D.B. Neau, H.D. Bellamy, and D.J. Leahy. Structural basis for ligand and heparin binding to neuropilin B domains. *Proc Natl Acad Sci USA*, 104(15):6152–6157, April 2007.

T.A. Verdoorn. Formation of heteromeric γ-aminobutyric acid type A receptors containing two different α subunits. *Mol Pharmacol*, 45(3):475–480, March 1994.

L.D. Walensky and G.H. Bird. Hydrocarbon-stapled peptides: principles, practice, and progress. *J Med Chem*, 57(15):6275–6288, August 2014.

N.G. Walter, C.-Y. Huang, A.J. Manzo, and M.A. Sobhy. Do-it-yourself guide: how to use the modern single-molecule toolkit. *Nat Methods*, 5(6):475–489, June 2008.

T.S. Walter, C. Meier, R. Assenberg, K.-F. Au, J. Ren, A. Verma, J.E. Nettleship, R.J. Owens, D.I. Stuart, and J.M. Grimes. Lysine methylation as a routine rescue strategy for protein crystallization. *Structure*, 14(11):1617–1622, November 2006.

L. Wamprecht. Analyse der head-to-tail Interaktion von Muskelin. Bachelorarbeit im Studienfach Biochemie, Julius-Maximilians-Universität Würzburg, 2012.

J.W. Wang, J.R. Chen, Y.X. Gu, C.D. Zheng, and H.F. Fan. Direct-method SAD phasing with partial-structure iteration: towards automation. *Acta Crystallogr D Biol Crystallogr*, 60(Pt 11):1991–1996, November 2004.

S.D. Weeks, M. Drinker, and P.J. Loll. Ligation independent cloning vectors for expression of SUMO fusions. *Protein Expr Purif*, 53(1):40–50, May 2007.

T.J. Wheeler, J. Clements, and R.D. Finn. Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. *BMC Bioinformatics*, 15:7, January 2014.

M.D. Winn, C.C. Ballard, K.D. Cowtan, E.J. Dodson, P. Emsley, P.R. Evans, R.M. Keegan, E.B. Krissinel, A.G.W. Leslie, A. McCoy, S.J. McNicholas, G.N. Murshudov, N.S. Pannu, E.A. Potterton, H.R. Powell, R.J. Read, A. Vagin, and K.S. Wilson. Overview

of the *CCP4* suite and current developments. *Acta Crystallogr D Biol Crystallogr*, 67 (Pt 4):235–242, April 2011.

T. Wiseman, S. Williston, J.F. Brandts, and L.N. Lin. Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem*, 179 (1):131–137, May 1989.

D. Xu, A. Farmer, G. Collett, N.V. Grishin, and Y.M. Chook. Sequence and structural analyses of nuclear export signals in the NESdb database. *Mol Biol Cell*, 23(18): 3677–3693, September 2012.

F. Xue and L. Cooley. *kelch* encodes a component of intercellular bridges in Drosophila egg chambers. *Cell*, 72(5):681–693, March 1993.

G. Yachdav, E. Kloppmann, L. Kajan, M. Hecht, T. Goldberg, T. Hamp, P. Hönigschmid, A. Schafferhans, M. Roos, M. Bernhofer, L. Richter, H. Ashkenazy, M. Punta, A. Schlessinger, Y. Bromberg, R. Schneider, G. Vriend, C. Sander, N. Ben-Tal, and B. Rost. PredictProtein - an open resource for online prediction of protein structural and functional features. *Nucl Acids Res*, 42(Web Server issue):W337–43, July 2014.

J.-M. Yang and C.-H. Tung. Protein structure database search and evolutionary classification. *Nucl Acids Res*, 34(13):3646–3659, August 2006.

Y. Zhang. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, 9 (1):40, March 2008.

APPENDIX

6.1 Abbreviations

6xHis	Hexahistidine
A$_{280}$	Absorbance at 280 nm
AUC	Analytical ultracentrifugation
BESSY	Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung
BS^3	Bis[sulfosuccinimidyl] suberate
CD	Circular dichroism
CNS	Central nervous system
CV	Column volume
dATP	2'-Deoxyadenosine 5'-triphosphate
dCTP	2'-Deoxycytidine 5'-triphosphate
dGTP	2'-Deoxyguanosine 5'-triphosphate
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
dTTP	2'-Deoxythymidine 5'-triphosphate
dtUD1	Doubly tagged catalytic domain of the *S. cerevisiae* ubiquitin-like-specific protease 1
E-value	Expectation value
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum
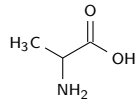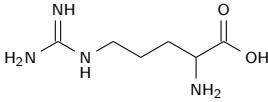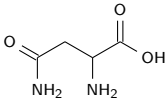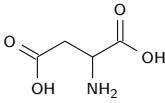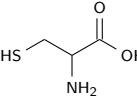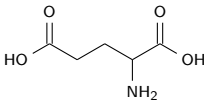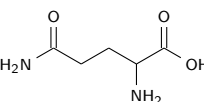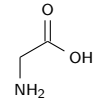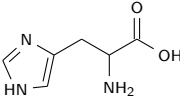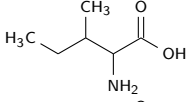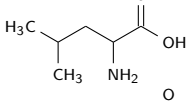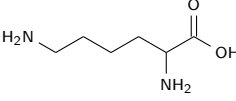
ESRF	European synchroton radiation facility
FRET	Förster resonance energy transfer
GABA	γ-aminobutyric acid
HEK	Human embryonic kidney
HMM	Hidden Markov model
ID	Identifier
IPTG	Isopropyl-β-D-thiogalactopyranoside
ITC	Isothermal titration calorimetry
K_D	Dissociation constant
LB	Lysogeny broth
MALS	Multi-angle light scattering
NAGE	Native agarose gel electrophoresis
NES	Nuclear export signal
NLS	Nuclear localization signal
NMA	Normal mode analysis
NMR	Nuclear magnetic resonance
OD_{600}	Optical density at 600 nm
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PMSF	Phenylmethylsulfonyl fluoride
r.m.s.d.	Root mean square deviation
SAD	Single-wavelength anomalous diffraction
SAXS	Small-angle X-ray scattering
SDS	Sodium dodecyl sulfate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
SLIC	Sequence and ligation independent cloning
SV-AUC	Sedimentation velocity analytical ultracentrifugation
TCEP	Tris[2-carboxyethyl]phosphine
TEV	Tobacco etch virus
TLS	Translation, libration, screw-motion

Tris	Tris[hydroxymethyl]-aminomethane
UV	Ultraviolet

Protein names
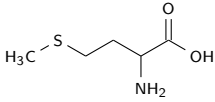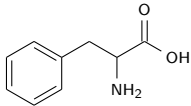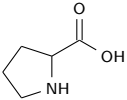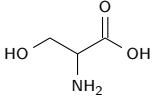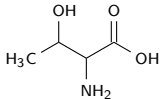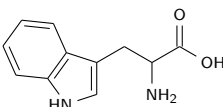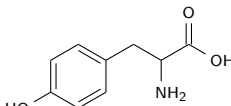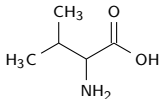
ARMC8	Armadillo repeat-containing protein 8
Cdk5	Cyclin-dependent-like kinase 5
DIC1A	Dynein intermediate chain isoform 1A
FOP	Fi bro blast growth factor receptor 1 oncogene partner
GFP	Green fluorescent protein
Gid	Glucose induced degradation
GST	Glutathione S transferase
HO-1	Heme-oxygenase-1
Keap1	Kelch-like ECH-associated protein 1
KLHL3	Kelch-like protein 3
Krp1	Kelch-related protein 1
Lasp1	LIM and SH3 domain protein 1
MAEA	Macrophage erythroblast attacher
Nrf2	Nuclear factor (erythroid-derived 2)-like 2
PKC	Protein kinase C
RanBP9	Ran-binding protein 9
Rmnd5a	Protein RMD5 homolog A
SUMO	Small ubiquitin like modifier
TBL1	Transducin β-like protein 1
Twa1	Two-hybrid associated protein to RanBP9
WNK4	Protein kinase with no lysine (K) 4

6.2 Amino acids

Amino acid	3-letter code	1-letter code	Grouped as	Skeletal formula
Alanine	Ala	A	non-polar aliphatic	
Arginine	Arg	R	positive	
Asparagine	Asn	N	polar	
Aspartic acid	Asp	D	negative	
Cysteine	Cys	C	*unique*	
Glutamic acid	Glu	E	negative	
Glutamine	Gln	Q	polar	
Glycine	Gly	G	*unique*	
Histidine	His	H	positive	
Isoleucine	Ile	I	non-polar aliphatic	
Leucine	Leu	L	non-polar aliphatic	
Lysine	Lys	K	positive	

Amino acid	3-letter code	1-letter code	Grouped as	Skeletal formula
Methionine	Met	M	non-polar aliphatic	
Phenylalanine	Phe	F	non-polar aromatic	
Proline	Pro	P	*unique*	
Serine	Ser	S	polar	
Threonine	Thr	T	polar	
Tryptophan	Trp	W	non-polar aromatic	
Tyrosine	Tyr	Y	non-polar aromatic	
Valine	Val	V	non-polar aliphatic	

151

6.3 Screen compositions

ThermoFluor screen

Well	Composition	Well	Composition
A1	0.1 M citric acid pH 4.5	E1	0.1 M sodium cacodylate pH 6.0
A2	0.1 M BisTris pH 7.0	E2	0.1 M BisTris propane pH 7.0
A3	0.1 M imidazole pH 6.5	E3	0.1 M MOPS pH 7.0
A4	0.1 M HEPES pH 8.0	E4	0.1 M bicine pH 9.0
A5	0.1 M Tris pH 8.5	E5	0.1 M glycyl-glycine pH 8.5
B1	0.1 M acetat pH 4.6	F1	0.1 M sodium cacodylate pH 6.5
B2	0.1 M ADA pH 6.5	F2	0.1 M PIPES pH 6.5
B3	0.1 M imidazole pH 8.0	F3	0.1 M MOPS pH 7.5
B4	0.1 M HEPES pH 8.5	F4	0.1 M Tris pH 7.0
B5	0.1 M Tris pH 9.0	F5	0.1 M CHES pH 9.0
C1	0.1 M MES pH 5.5	G1	0.1 M BisTris pH 5.5
C2	0.1 M ADA pH 7.0	G2	0.1 M PIPES pH 7.0
C3	0.1 M sodium potassium phosphate pH 6.8	G3	0.1 M HEPES pH 7.0
C4	0.1 M Bicine pH 8.0	G4	0.1 M Tris pH 7.5
C5	0.1 M TAPS pH 8.0	G5	0.1 M Ches pH 9.5
D1	0.1 M MES pH 6.5	H1	0.1 M BisTris pH 6.5
D2	0.1 M BisTris propane pH 6.0	H2	0.1 M PIPES pH 7.5
D3	0.1 M sodium potassium phosphate pH 7.5	H3	0.1 M HEPES pH 7.5
D4	0.1 M Bicine pH 8.5	H4	0.1 M Tris pH 8.0
D5	0.1 M TAPS pH 9.0	H5	0.1 M Caps pH 9.8

Salt screen

Well	Composition		
A1	0.1 M potassium chloride,	0.1 M Tris pH 8.0,	12 % PEG 2000 MME
A2	0.1 M calcium chloride,	0.1 M Tris pH 8.0,	12 % PEG 2000 MME
A3	0.1 M potassium chloride,	0.1 M Tris pH 8.0,	15 % PEG 2000 MME
A4	0.1 M calcium chloride,	0.1 M Tris pH 8.0,	15 % PEG 2000 MME
A5	0.1 M potassium chloride,	0.1 M Tris pH 8.0,	18 % PEG 2000 MME
A6	0.1 M calcium chloride,	0.1 M Tris pH 8.0,	18 % PEG 2000 MME
A7	0.1 M potassium chloride,	0.1 M Tris ph 8.5,	12 % PEG 2000 MME
A8	0.1 M calcium chloride,	0.1 M Tris ph 8.5,	12 % PEG 2000 MME
A9	0.1 M potassium chloride,	0.1 M Tris ph 8.5,	15 % PEG 2000 MME
A10	0.1 M calcium chloride,	0.1 M Tris ph 8.5,	15 % PEG 2000 MME
A11	0.1 M potassium chloride,	0.1 M Tris ph 8.5,	18 % PEG 2000 MME
A12	0.1 M calcium chloride,	0.1 M Tris ph 8.5,	18 % PEG 2000 MME
B1	0.1 M potassium bromide,	0.1 M Tris pH 8.0,	12 % PEG 2000 MME
B2	0.1 M sodium chloride,	0.1 M Tris pH 8.0,	12 % PEG 2000 MME

Well	Composition		
B3	0.1 M potassium bromide,	0.1 M Tris pH 8.0,	15 % PEG 2000 MME
B4	0.1 M sodium chloride,	0.1 M Tris pH 8.0,	15 % PEG 2000 MME
B5	0.1 M potassium bromide,	0.1 M Tris pH 8.0,	18 % PEG 2000 MME
B6	0.1 M sodium chloride,	0.1 M Tris pH 8.0,	18 % PEG 2000 MME
B7	0.1 M potassium bromide,	0.1 M Tris ph 8.5,	12 % PEG 2000 MME
B8	0.1 M sodium chloride,	0.1 M Tris ph 8.5,	12 % PEG 2000 MME
B9	0.1 M potassium bromide,	0.1 M Tris ph 8.5,	15 % PEG 2000 MME
B10	0.1 M sodium chloride,	0.1 M Tris ph 8.5,	15 % PEG 2000 MME
B11	0.1 M potassium bromide,	0.1 M Tris ph 8.5,	18 % PEG 2000 MME
B12	0.1 M sodium chloride,	0.1 M Tris ph 8.5,	18 % PEG 2000 MME
C1	0.1 M potassium fluoride,	0.1 M Tris pH 8.0,	12 % PEG 2000 MME
C2	0.1 M ammonium chloride,	0.1 M Tris pH 8.0,	12 % PEG 2000 MME
C3	0.1 M potassium fluoride,	0.1 M Tris pH 8.0,	15 % PEG 2000 MME
C4	0.1 M ammonium chloride,	0.1 M Tris pH 8.0,	15 % PEG 2000 MME
C5	0.1 M potassium fluoride,	0.1 M Tris pH 8.0,	18 % PEG 2000 MME
C6	0.1 M ammonium chloride,	0.1 M Tris pH 8.0,	18 % PEG 2000 MME
C7	0.1 M potassium fluoride,	0.1 M Tris ph 8.5,	12 % PEG 2000 MME
C8	0.1 M ammonium chloride,	0.1 M Tris ph 8.5,	12 % PEG 2000 MME
C9	0.1 M potassium fluoride,	0.1 M Tris ph 8.5,	15 % PEG 2000 MME
C10	0.1 M ammonium chloride,	0.1 M Tris ph 8.5,	15 % PEG 2000 MME
C11	0.1 M potassium fluoride,	0.1 M Tris ph 8.5,	18 % PEG 2000 MME
C12	0.1 M ammonium chloride,	0.1 M Tris ph 8.5,	18 % PEG 2000 MME
D1	0.1 M potassium citrate,	0.1 M Tris pH 8.0,	12 % PEG 2000 MME
D2	50 mM cadmium chloride,	0.1 M Tris pH 8.0,	12 % PEG 2000 MME
D3	0.1 M potassium citrate,	0.1 M Tris pH 8.0,	15 % PEG 2000 MME
D4	50 mM cadmium chloride,	0.1 M Tris pH 8.0,	15 % PEG 2000 MME
D5	0.1 M potassium citrate,	0.1 M Tris pH 8.0,	18 % PEG 2000 MME
D6	50 mM cadmium chloride,	0.1 M Tris pH 8.0,	18 % PEG 2000 MME
D7	0.1 M potassium citrate,	0.1 M Tris ph 8.5,	12 % PEG 2000 MME
D8	50 mM cadmium chloride,	0.1 M Tris ph 8.5,	12 % PEG 2000 MME
D9	0.1 M potassium citrate,	0.1 M Tris ph 8.5,	15 % PEG 2000 MME
D10	50 mM cadmium chloride,	0.1 M Tris ph 8.5,	15 % PEG 2000 MME
D11	0.1 M potassium citrate,	0.1 M Tris ph 8.5,	18 % PEG 2000 MME
D12	50 mM cadmium chloride,	0.1 M Tris ph 8.5,	18 % PEG 2000 MME
E1	0.1 M potassium formate,	0.1 M Tris pH 8.0,	12 % PEG 2000 MME
E2	0.1 M magnesium chloride,	0.1 M Tris pH 8.0,	12 % PEG 2000 MME
E3	0.1 M potassium formate,	0.1 M Tris pH 8.0,	15 % PEG 2000 MME
E4	0.1 M magnesium chloride,	0.1 M Tris pH 8.0,	15 % PEG 2000 MME
E5	0.1 M potassium formate,	0.1 M Tris pH 8.0,	18 % PEG 2000 MME
E6	0.1 M magnesium chloride,	0.1 M Tris pH 8.0,	18 % PEG 2000 MME
E7	0.1 M potassium formate,	0.1 M Tris ph 8.5,	12 % PEG 2000 MME
E8	0.1 M magnesium chloride,	0.1 M Tris ph 8.5,	12 % PEG 2000 MME
E9	0.1 M potassium formate,	0.1 M Tris ph 8.5,	15 % PEG 2000 MME
E10	0.1 M magnesium chloride,	0.1 M Tris ph 8.5,	15 % PEG 2000 MME

Well	Composition		
E11	0.1 M potassium formate,	0.1 M Tris ph 8.5,	18 % PEG 2000 MME
E12	0.1 M magnesium chloride,	0.1 M Tris ph 8.5,	18 % PEG 2000 MME
F1	0.1 M potassium nitrate,	0.1 M Tris pH 8.0,	12 % PEG 2000 MME
F2	0.1 M lithium chloride,	0.1 M Tris pH 8.0,	12 % PEG 2000 MME
F3	0.1 M potassium nitrate,	0.1 M Tris pH 8.0,	15 % PEG 2000 MME
F4	0.1 M lithium chloride,	0.1 M Tris pH 8.0,	15 % PEG 2000 MME
F5	0.1 M potassium nitrate,	0.1 M Tris pH 8.0,	18 % PEG 2000 MME
F6	0.1 M lithium chloride,	0.1 M Tris pH 8.0,	18 % PEG 2000 MME
F7	0.1 M potassium nitrate,	0.1 M Tris ph 8.5,	12 % PEG 2000 MME
F8	0.1 M lithium chloride,	0.1 M Tris ph 8.5,	12 % PEG 2000 MME
F9	0.1 M potassium nitrate,	0.1 M Tris ph 8.5,	15 % PEG 2000 MME
F10	0.1 M lithium chloride,	0.1 M Tris ph 8.5,	15 % PEG 2000 MME
F11	0.1 M potassium nitrate,	0.1 M Tris ph 8.5,	18 % PEG 2000 MME
F12	0.1 M lithium chloride,	0.1 M Tris ph 8.5,	18 % PEG 2000 MME
G1	0.1 M potassium iodide,	0.1 M Tris pH 8.0,	12 % PEG 2000 MME
G2	0.1 M sodium malonate,	0.1 M Tris pH 8.0,	12 % PEG 2000 MME
G3	0.1 M potassium iodide,	0.1 M Tris pH 8.0,	15 % PEG 2000 MME
G4	0.1 M sodium malonate,	0.1 M Tris pH 8.0,	15 % PEG 2000 MME
G5	0.1 M potassium iodide,	0.1 M Tris pH 8.0,	18 % PEG 2000 MME
G6	0.1 M sodium malonate,	0.1 M Tris pH 8.0,	18 % PEG 2000 MME
G7	0.1 M potassium iodide,	0.1 M Tris ph 8.5,	12 % PEG 2000 MME
G8	0.1 M sodium malonate,	0.1 M Tris ph 8.5,	12 % PEG 2000 MME
G9	0.1 M potassium iodide,	0.1 M Tris ph 8.5,	15 % PEG 2000 MME
G10	0.1 M sodium malonate,	0.1 M Tris ph 8.5,	15 % PEG 2000 MME
G11	0.1 M potassium iodide,	0.1 M Tris ph 8.5,	18 % PEG 2000 MME
G12	0.1 M sodium malonate,	0.1 M Tris ph 8.5,	18 % PEG 2000 MME
H1	0.1 M cesium chloride,	0.1 M Tris pH 8.0,	12 % PEG 2000 MME
H2	0.1 M potassium sulfate,	0.1 M Tris pH 8.0,	12 % PEG 2000 MME
H3	0.1 M cesium chloride,	0.1 M Tris pH 8.0,	15 % PEG 2000 MME
H4	0.1 M potassium sulfate,	0.1 M Tris pH 8.0,	15 % PEG 2000 MME
H5	0.1 M cesium chloride,	0.1 M Tris pH 8.0,	18 % PEG 2000 MME
H6	0.1 M potassium sulfate,	0.1 M Tris pH 8.0,	18 % PEG 2000 MME
H7	0.1 M cesium chloride,	0.1 M Tris ph 8.5,	12 % PEG 2000 MME
H8	0.1 M potassium sulfate,	0.1 M Tris ph 8.5,	12 % PEG 2000 MME
H9	0.1 M cesium chloride,	0.1 M Tris ph 8.5,	15 % PEG 2000 MME
H10	0.1 M potassium sulfate,	0.1 M Tris ph 8.5,	15 % PEG 2000 MME
H11	0.1 M cesium chloride,	0.1 M Tris ph 8.5,	18 % PEG 2000 MME
H12	0.1 M potassium sulfate,	0.1 M Tris ph 8.5,	18 % PEG 2000 MME

6.4 Construct list

ID	Protein (amino acids)	Mutations	Species	Vector	Modifications N-terminal	C-terminal
M1	Mkln (M1-D156)		*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M2	Mkln (M1-E205)		*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M3	Mkln (M1-G272)		*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M4	Mkln (D157-L735)		*R. norvegicus*	pETM-SUMO	6xHis-SUMO	6xHis
M5	Mkln (E205-L735)		*R. norvegicus*	pETM-SUMO	6xHis-SUMO	6xHis
M6	Mkln (G273-L735)		*R. norvegicus*	pETM-SUMO	6xHis-SUMO	6xHis
M7	Mkln (M1-Y247)		*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M8	Mkln (Y247-L375)		*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M9	Mkln (M1-L735)		*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M10	Mkln (D157-G272)		*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M11	Mkln (M1-D156)		*R. norvegicus*	pETM-11	6xHis-TEV	
M12	Mkln (M1-E205)		*R. norvegicus*	pETM-11	6xHis-TEV	
M13	Mkln (M1-G272)		*R. norvegicus*	pETM-11	6xHis-TEV	
M14	Mkln (D157-L735)		*R. norvegicus*	pETM-11	6xHis-TEV	6xHis
M15	Mkln (E205-L735)		*R. norvegicus*	pETM-11	6xHis-TEV	6xHis
M16	Mkln (G273-L735)		*R. norvegicus*	pETM-11	6xHis-TEV	6xHis
M19	Mkln (M1-L735)		*R. norvegicus*	pETM-11	6xHis-TEV	6xHis
M21	Mkln (M1-D156)		*R. norvegicus*	pETM-30	6xHis-GST-TEV	
M22	Mkln (M1-E205)		*R. norvegicus*	pETM-30	6xHis-GST-TEV	
M23	Mkln (M1-G272)		*R. norvegicus*	pETM-30	6xHis-GST-TEV	
M24	Mkln (D157-L735)		*R. norvegicus*	pETM-30	6xHis-GST-TEV	6xHis
M25	Mkln (E205-L735)		*R. norvegicus*	pETM-30	6xHis-GST-TEV	6xHis
M26	Mkln (G273-L735)		*R. norvegicus*	pETM-30	6xHis-GST-TEV	6xHis
M29	Mkln (M1-L735)		*R. norvegicus*	pETM-30	6xHis-GST-TEV	
M30	Mkln (M1-E205)	F184E L196Q	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M31	Mkln (M1-E205)	F184E L196K	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M32	Mkln (M1-E205)	F184Q L196E	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M33	Mkln (M1-E205)	F184K L196E	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M34	Mkln (M1-G272)	F184E L196Q	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M35	Mkln (M1-G272)	F184E L196K	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M36	Mkln (M1-G272)	F184Q L196E	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M37	Mkln (M1-G272)	F184K L196E	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M38	Mkln (E205-L735)	F184E L196Q	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	6xHis
M39	Mkln (E205-L735)	F184E L196K	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	6xHis
M40	Mkln (E205-L735)	F184Q L196E	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	6xHis
M41	Mkln (E205-L735)	F184K L196E	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	6xHis
M42	Mkln (M1-L735)	F184E L196Q	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M43	Mkln (M1-L735)	F184E L196K	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M44	Mkln (M1-L735)	F184Q L196E	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M45	Mkln (M1-L735)	F184K L196E	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M46	Mkln (D157-L735)	S324A T515A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	6xHis
M47	Mkln (D157-L735)	S324E T515E	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	6xHis
M48	Mkln (E205-L735)	S324A T515A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	6xHis
M49	Mkln (E205-L735)	S324E T515E	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	6xHis
M50	Mkln (M1-L735)	S324A T515A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M51	Mkln (M1-L735)	S324E T515E	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M52	Mkln (D157-L735)		*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M53	Mkln (E205-L735)		*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M54	Mkln (D157-R625)		*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M55	Mkln (E205-R625)		*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M56	Mkln (G273-R625)		*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M57	Mkln (E12-E205)		*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M58	Mkln (M1-H187)		*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M59	Mkln (E12-H187)		*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M60	Mkln (M1-D156)	E50A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	

ID	Protein (amino acids)	Mutations	Species	Vector	Modifications N-terminal	C-terminal
M61	Mkln (M1-D156)	Y53A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M62	Mkln (M1-D156)	P54E	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M63	Mkln (M1-D156)	Y57E	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M64	Mkln (M1-D156)	F143A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M65	Mkln (M1-D156)	N144R	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M66	Mkln (M1-D156)	N144W	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M67	Mkln (D157-L735)	G280-D289 3A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M68	Mkln (D157-L735)	R315-H328 10A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M69	Mkln (D157-L735)	R345-S359 3A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M70	Mkln (D157-L735)	E376-H381 10A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M71	Mkln (D157-L735)	G405-G427 3A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M72	Mkln (D157-L735)	R442-G458 10A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M73	Mkln (D157-L735)	G475-D484 3A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M74	Mkln (D157-L735)	G501-Q516 10A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M75	Mkln (D157-L735)	G532-N546 3A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M76	Mkln (D157-L735)	K563-F585 10A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M77	Mkln (D157-L735)	G602-D617 3A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M78	Mkln (D157-L735)	G280-D289 3A R315-H328 10A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M79	Mkln (D157-L735)	R345-S359 3A E376-H381 10A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M80	Mkln (D157-L735)	G405-G427 3A R442-G458 10A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M81	Mkln (D157-L735)	G475-D484 3A G501-Q516 10A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M82	Mkln (D157-L735)	G532-N546 3A K563-F585 10A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M83	Mkln (E12-D156)		*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M84	Mkln (E12-G272)		*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M85	Mkln (M1-E205)	C180S	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M86	Mkln (M1-E205)	C180Q	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M87	Mkln (M1-E205)	C180W	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M88	Mkln (M1-E205)	R172A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M89	Mkln (M1-E205)	E173A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M90	Mkln (M1-E205)	R185A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M91	Mkln (M1-E205)	N188A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M92	Mkln (M1-L735)	F143A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M93	Mkln (M1-L735)	F143A F184E L196Q	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M94	Mkln (M1-L735)	N144W	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M95	Mkln (M1-L735)	N144W F184E L196Q	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M96	Mkln (M1-L735)	N144R	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M97	Mkln (M1-L735)	N144R F184E L196Q	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M98	Mkln (M1-E205)	F184A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M99	Mkln (M1-L735)	K182A H183A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M100	Mkln (M1-V700)		*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M101	Mkln (M1-E205)	K182A H183A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M102	Mkln (D157-L735)	K182A H183A	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M103	Mkln (D157-V700)		*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M104	Mkln (E205-V700)		*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M105	Mkln (V700-L735)		*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M106	Mkln (V700-L735)		*R. norvegicus*	pETM-30	6xHis-GST-TEV	
M107	Mkln (V700-L735)		*R. norvegicus*	pETM-41	6xHis-MBP-TEV	
M108	Mkln (M1-R625)		*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M109	Mkln (R625-L735)		*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M110	Mkln (Y247-R625)		*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M111	Mkln (E12-R625)	N144R	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M112	Mkln (E12-L735)	N144R	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M113	Mkln (M1-E205)	C13S C82S C128S C164S	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M114	Mkln (M1-L735)	C180S	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M115	Mkln (M1 - E205)	K199E	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M116	Mkln (M1 - L735)	K199E	*R. norvegicus*	pETM-SUMO	6xHis-SUMO	
M117	Mkln (M1-L735)		*R. norvegicus*	pBK-CMV		[1]
M118	Mkln (M1-L735)		*R. norvegicus*	pEGFP-C1*	mRFP	[1]
M119	Mkln (M1-L735)		*R. norvegicus*	pEGFP-C1*	mCherry	[1]
M120	Mkln (M1-L735)	N144R	*R. norvegicus*	pBK-CMV		

| ID | Protein (amino acids) | Mutations | Species | Vector | Modifications | | |
					N-terminal	C-terminal	
M121	Mkln (M1-L735)	N144R	*R. norvegicus*	pEGFP-C1*	mRFP		
M122	Mkln (M1-L735)	N144R	*R. norvegicus*	pEGFP-C1*	mCherry		
M123	Mkln (M1-L735)	F184E L196Q	*R. norvegicus*	pBK-CMV			
M124	Mkln (M1-L735)	F184E L196Q	*R. norvegicus*	pEGFP-C1*	mRFP		
M125	Mkln (M1-L735)	F184E L196Q	*R. norvegicus*	pEGFP-C1*	mCherry		
M126	Mkln (M1-L735)	N144R F184E L196Q	*R. norvegicus*	pBK-CMV			
M127	Mkln (M1-L735)	N144R F184E L196Q	*R. norvegicus*	pEGFP-C1*	mRFP		
M128	Mkln (M1-L735)	N144R F184E L196Q	*R. norvegicus*	pEGFP-C1*	mCherry		
M129	Mkln (M1-L735)	K182A H183A	*R. norvegicus*	pEGFP-C1*	mCherry		
M130	Mkln (M1-V700)		*R. norvegicus*	pEGFP-C1*	mCherry		
G1	GABA$_A$R α1 (N334-R420)		*R. norvegicus*	pETM-11	6xHis-TEV		[2]
G2	GABA$_A$R α2 (N335-R419)	V371G	*H. sapiens***	pETM-11	6xHis-TEV		[2]
G3	GABA$_A$R α3 (N360-K458)		*R. norvegicus*	pETM-11	6xHis-TEV		[2]
G4	GABA$_A$R α5 (N342-K429)		*R. norvegicus*	pETM-11	6xHis-TEV	6xHis	[2]
G5	GABA$_A$R α1 (N334-E407)		*R. norvegicus*	pETM-11	6xHis-TEV		
G6	GABA$_A$R α2 (N335-E406)	V371G	*H. sapiens***	pETM-11	6xHis-TEV		
G7	GABA$_A$R α3 (N360-A445)		*R. norvegicus*	pETM-11	6xHis-TEV		
G8	GABA$_A$R α5 (N342-E416)		*R. norvegicus*	pETM-11	6xHis-TEV		

[1] Plasmids were kindly proved by Frank Heisler, AG Kneussel, Center for Molecular Neurobiology (ZMNH), University Medical Center Hamburg-Eppendorf.

[2] Plasmids were kindly proved by Hans Maric, AG Schindelin, Rudolf Virchow Center for Biomedical Research (RVZ), University of Würzburg.

*pEGFP-C1 was modified by replacing the coding sequence of the fluorescent protein as described before (Heisler et al., 2011).

**Note that the amino acid sequence of the GABA$_A$R α2 intracellular loop of the rat and human homologues are 100 % identical.

6.5 Supplementary information on SAXS analysis
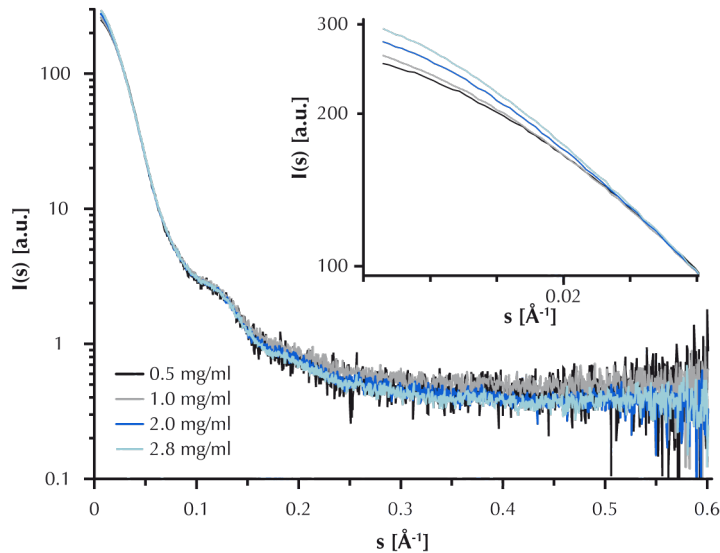
Figure 6.1 – Experimental scattering curves of Mkln$_{1-735}$ plotted on a logarithmic scale as a function of the momentum transfer (s = 4πsin(θ)/λ, with 2θ = scattering angle; λ = X-ray wavelength), low-angle range shown as insert. (M9)
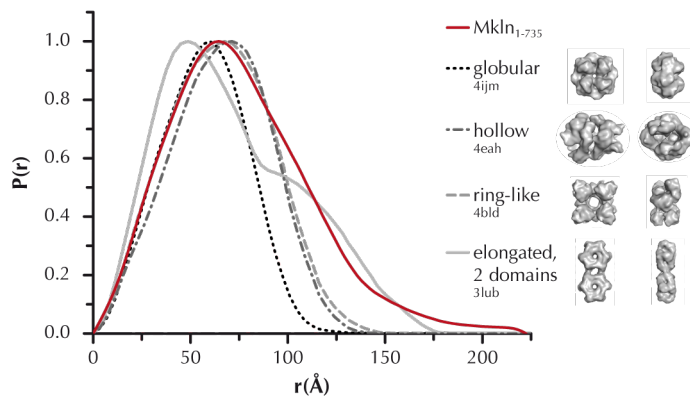
Figure 6.2 – Pair distance distribution function of Mkln$_{1-735}$ measured at a protein concentration of 0.5 mg/ml with pair distance distribution functions of several proteins (PDB codes given in the legend) of comparable size to the muskelin tetramer (320-360 kDa), derived with *GNOM* from theoretical scattering profiles calculated by *CRYSOL*, shown for comparison. (M9)
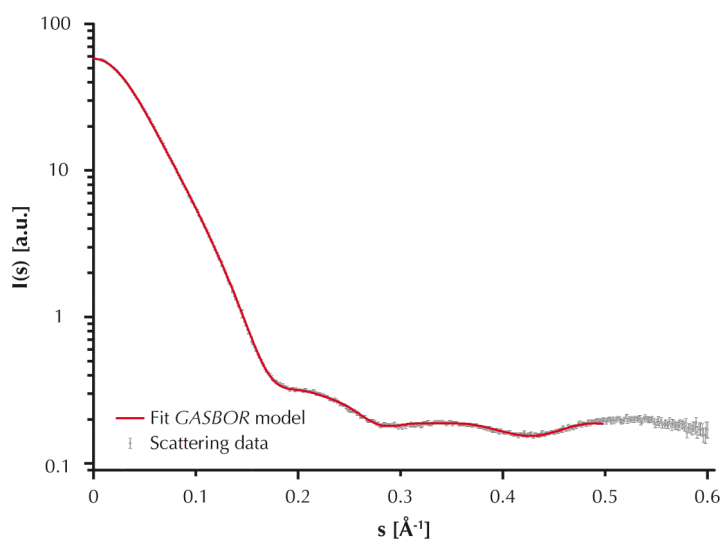
Figure 6.3 – Experimental scattering data with the fit of the best model of Mkln$_{205-735}$ plotted on a logarithmic scale as a function of the momentum transfer (s = 4πsin(θ)/λ, with 2θ = scattering angle; λ = X-ray wavelength). (M5)

LIST OF FIGURES

LIST OF TABLES

LIST OF PUBLICATIONS

Peer reviewed publications

C.F. Delto, F.F. Heisler, J. Kuper, B. Sander, M. Kneussel and H. Schindelin. The LisH motif of muskelin is crucial for oligomerization and governs intracellular localization. *Structure*, 23 (2): 364–373, February 2015

Congress contributions

C.F. Delto, F.F. Heisler, J. Kuper, M. Kneussel and H. Schindelin (March 2014). Crystal structure of Muskelin's discoidin and LisH-domain provides the basis to understand the mechanism and impact of its oligomerization. Presentation at the *22nd Annual Conference of the German Crystallographic Society*, Berlin (Germany)

C.F. Delto (October 2013): A complex story - how the oligomeric state governs the function of the protein muskelin. Presentation at *SCI Scientific Crosstalk - 8th International Symposium*, Würzburg (Germany)

C.F. Delto, F.F. Heisler, M. Kneussel and H. Schindelin (June 2013). Muskelin function is controlled by its oligomerization. Poster presented at *Inhibition in the CNS - Gordon Research Conference*, Les Diablerets (Switzerland)

CURRICULUM VITAE

Acknowledgements

I'd like to bring this work to completion by expressing my sincere gratitude towards the many people that accompanied and supported me on my way.

To be named first, my supervisor *Hermann Schindelin*, for his mentorship, his support, his patience, his trust in me, his keen questions that trained my scientific thinking and for helping me to grasp more of the true value of crystal structures beyond their aesthetics.

My second supervisor, *Matthias Kneussel,* to whom I owe great thanks in several ways; for sharing the muskelin project, for giving his advice and support since and for great ideas and scientific input in the discussions about it.

Thomas Raabe, who agreed to complete my supervisory committee and has contributed a valuable new perspective to the discussions in the meetings.

Frank Heisler, from whose generously shared data I profited so much, for bringing my biochemical data alive and adding the most intriguing aspect to it, for open discussions and for simply being the best collaboration partner one could wish for.

Kristina Haslinger, who was a great diploma student, a companion-in-cell-culture on the glycine receptor project, and the most melodious soprano in the lab.

My bachelor students *Daniel Krause, Laura Wamprecht, Jana Kolkmann* and *Susanne Sauer,* who dedicated their theses to the muskelin project - our joint work triggered many of its major progressions, and taught me some lessons about teaching.

Daniela Schneeberger, who, since day one, took care and helped to keep me on track; she held more than once mentally and literally the boxing pads for me.

Antje Schäfer, who was always willing to share the practical tricks of the trades and to listen to scientific and less scientific problems with her unmatched empathy.

Inga Birkholz, who became more than a colleague, but a true friend, to me.

My dear friend *Nora*, who helped me immensely with her compassionate, honest, and encouraging words.

My family, which was very supportive, sympathetic and forbearing with me and my commitment to work. The curiosity and enthusiasm that my mother-in-law *Almut* showed for my work was a balm for the scientist's soul.

My husband *Ralf*, who believed in me, when I couldn't myself, and stood by my side against all odds.

Affidavit

I hereby confirm that my thesis entitled

Structural and Biochemical Characterization of the GABA$_A$ Receptor Interacting Protein Muskelin

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 confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor similar form.

....

Place, Date   Signature