Structural and Biochemical Characterization of LINC Complexes

by

Victor E. Cruz Ruiz
B.S., University of Puerto Rico at Aguadilla
Aguadilla, PR 2011

SUBMITTED TO THE DEPARTMENT OF BIOLOGY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

AT THE

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

APRIL 2017 [June 2017]

Signature of Author..... Signature redacted .....

Victor E. Cruz Ruiz
Department of Biology
April 11, 2017

Certified by... Signature redacted .....

Thomas U. Schwartz
Professor of Biology
Thesis Supervisor

Accepted by..... Signature redacted .....

Amy E. Keating
Professor of Biology
Chair, Biology Graduate Committee

Structural and Biochemical Characterization of LINC Complexes

Victor E. Cruz Ruiz

Abstract

The nuclear envelope (NE) is comprised of a double membrane bilayer that physically separates the nucleoplasm from the cytoplasm. Information can be transmitted through the NE by molecular exchange through the nuclear pore complex (NPC) and by transduction of mechanical forces mediated by the linker of nucleoskeleton to cytoskeleton (LINC) complexes. LINC complexes are composed of two proteins. The nuclear half is formed by SUN proteins and the cytoplasmic half by KASH-peptide containing proteins. Each KASH protein interacts with different elements of the cytoskeleton and serve a distinct function. What dictates the pairing of the diverse repertoire of SUN- and KASH-proteins? Mechanistic details on the regulation of SUN-KASH interactions have so far remained largely elusive. To address this problem, we have solved high resolution X-ray crystal structures of SUN2 in complex with various KASH peptides. These structures revealed two distinct binding modes between SUN and KASH. Sequence analysis can be used to distinguish between these alternative binding modes. Additional biochemical characterization showed that SUN trimers can bind up to three different KASH peptides simultaneously, adding an unexpected layer of complexity to LINC complexes.

A hallmark of SUN proteins is the elongated coiled-coil domain that precedes the SUN domain. This coiled-coil domain likely spans the width of the perinuclear space (PNS) and may be involved in mediating higher order assemblies of LINC complexes. We have extensively characterized the oligomeric state of the coiled-coil domain in solution, and have mapped the regions that are critical for trimerization. We believe that the best strategy moving forward is to structurally characterize the coiled-coil of SUN proteins, to which I have contributed important initial results.

The cytoplasmic domains of KASH proteins, also known as Nesprins, physically anchor the outer nuclear membrane (ONM) with various cytoskeletal proteins. Nesprin-2, for example, directly binds to actin through its N terminus and indirectly through interactions with other actin binding proteins. Nesprin-2 actin complexes are required for nuclear nuclear polarization during fibroblast migration. I have started to characterize the interactions between Nesprin-2 and FHOD1, as well as Fascin-1, both well known actin binders. Additionally, I have initiated the X-ray crystallographic analysis to obtain a detailed structural understanding of these complexes and provide detailed insight into the pathologies associated with aberrant interactions.

Thesis Supervisor: Thomas U. Schwartz
Title: Professor of Biology

Table of Contents

Acknowledgments

These 5 years at MIT have been some of the most exciting of my life. During this time, I have grown as a scientist and as a person, even tying the knot with my long time girlfriend Francheska. I have learned that success as a scientist comes as much from knowledge and deep thinking, as from the "stick-to-itness" that is critical in pushing through a difficult obstacle when pursuing a research problem. It goes without saying, during my time at MIT I have met and worked with some amazing people.

To my former lab mates, especially Kasper, and Greg. Much of my growth as a scientist, especially during my first years in the laboratory blossomed from our conversations. To my current lab mates, thank you for your friendship, for the memorable lunch discussions, and for your scientific advice. I would also like to thank my graduate committee, Bob Sauer and Cathy Drennan. Your scientific input and intellectual encouragement has been invaluable during my graduate career.

To my advisor Thomas, I thank you for taking a chance on me and allowing me to join your lab. The way I think has been profoundly impacted by our discussions, scientific and otherwise. You have taught me what it means to be a scientist. The environment fostered in your laboratory is intellectually nurturing and friendly, and it is my sincere hope that this culture becomes the norm of what a research laboratory should be.

To my wife Francheska, you have made the difficult times endurable and the good times unforgettable. You keep me grounded, every day you remind me of the things that are truly important in life, te amo morena. To my family, my sister Deborah and my parents, your encouragement and support throughout this process has been beyond what can be expected, and I am forever grateful for it. To my closest friends, Josh, Christian, Frank, Frankie, Alberto, and Vivek. I thank you for your advice, for listening, and lending a hand during difficult times.

Introduction

Introduction

A hallmark of eukaryotic cells is the presence of membrane bound organelles that compartmentalize the interior of cells into organelles with discrete functions. The nucleus is the largest of these organelles. It houses the genetic material of a cell, and is the site of ribosome biogenesis, transcription, and replication of DNA. The content of the nucleus is physically separated from the cytoplasm by a double membrane bilayer called the nuclear envelope (NE). The NE is composed of an outer nuclear membrane (ONM) and an inner nuclear membrane (INM). The ONM is contiguous with the endoplasmic reticulum (ER). The INM and ONM are separated by a lumen known as the perinuclear space (PNS). At discrete sites on the NE the INM and ONM are fused, forming circular openings occupied by the nuclear pore complex (NPC). The NPC primarily provides a gateway for selective molecular exchange between the nucleus and the cytoplasm (Kabachinski, 2015; Knockenhauer, 2016). As such, the NPC serves as a gateway through which the nucleus and cytoplasm can communicate. The NPC also restricts protein traffic from the ONM to the INM, which contributes to the unique protein composition of the latter.

The transmission of mechanical forces into the nucleus provides another form of communication aside from molecular exchange. The molecular tethers that span the PNS and connect the nucleus to the cytoskeleton are the LINC complexes (Figure 1). LINC complexes are formed by the INM SUN proteins, and the ONM KASH proteins (Starr, 2010). Effective mechanotransduction relies on the mechanical rigidity of the

nucleus that is maintained by a network of nuclear proteins. The nuclear-facing leaflet of the INM is lined by a meshwork of proteins, predominantly composed of lamin A/C that provides rigidity to the nuclear membranes. Additionally, many soluble nuclear proteins associate with lamins and create foci where heterochromatin, and transmembrane INM proteins can be retained (Ungricht, 2015). Collectively, this protein network is termed the nucleoskeleton. Mutations within nucleoskeletal proteins can cause highly diverse diseases, including muscular dystrophies, premature aging, and neurological disorders. These diseases are collectively termed laminopathies (Zhang, 2007; Haque, 2010; Chen, 2012; Starr, 2012; Horn, 2013; Barateau, 2017).

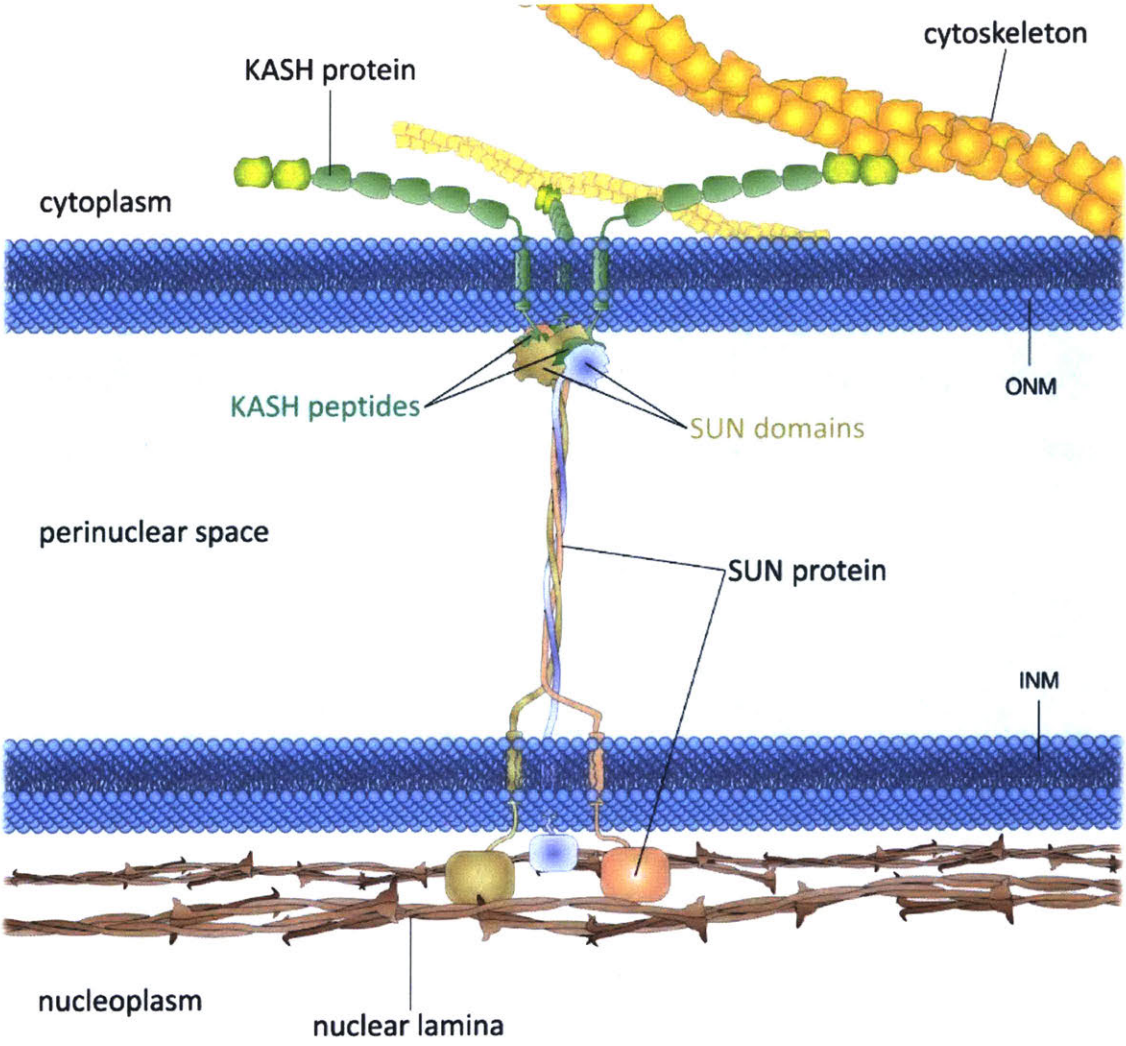


Figure 1. Overview of the nuclear envelope. Cartoon representation of SUN trimers bound to KASH proteins. At the nuclear side, SUN is connected to nuclear lamina (lamin A/C), while at the cytoplasmic side KASH proteins bind to cytoskeletal proteins. *Adapted from Chang, 2016.*

SUN proteins

SUN proteins are type II membrane proteins that are present in all eukaryotes. SUN proteins possess a highly conserved ~175 amino acid, C-terminal domain called the SUN domain. SUN proteins were first identified in *D. melanogaster* and *C. elegans*

(SAD1 and UNC84 respectively) in mutants with nuclear positioning and cell migration defects (Starr, 2010). Simpler single celled organisms such as *S. pombe* possess a single SUN protein, while more complex organisms, such as *H. sapiens* possess various SUN proteins. In humans five SUN proteins have been identified to date, SUN1 and SUN2 are expressed in all tissues and possess somewhat redundant roles while SUN3, SUN4 and SUN5 are more specialized and are only expressed in testis. While deletion mutants of either SUN1 or SUN2 are viable, the double knockout causes embryonic lethality in mice (Crisp, 2006). These mice can be rescued by expressing either SUN1 or SUN2 in neuronal cells.

The domain architecture of most SUN proteins is conserved. Generally, SUN proteins possess a variable N-terminal domain that projects into the nucleus followed by a transmembrane helix that spans the INM. The TM-helix is followed by an elongated, trimeric coiled-coil domain that presumably spans the width of the PNS. At the C terminus, the eponymous SUN domain is responsible for KASH binding (Sosa, 2013).

KASH proteins

KASH proteins, similar to SUN, are type II transmembrane proteins. In contrast to SUN, KASH proteins are tail anchored at the ONM instead of the INM. The most striking feature of KASH proteins is the highly conserved luminal domain that projects into the PNS. This short 30-40 amino acid element called the KASH peptide is necessary and sufficient for ONM positioning and for SUN binding. The KASH domain

comprises both the KASH peptide and the conserved transmembrane helix preceding it. Curiously, the transmembrane helix of the KASH domain is highly conserved in sequence, perhaps reflecting a functional role beyond mere ONM anchorage (Sosa, 2013).

To date, six KASH domain-containing proteins have been identified in humans (Chang, 2015). KASH1-4 are known as Nesprins, and mainly fold into modular repeats of three helix bundles termed spectrin repeats. Nesprin-1 and Nesprin-2 are very large proteins (800 kDa and 600 kDa, respectively) that possess an N-terminal actin binding domain (ABD) that is indispensable for nuclear coupling to actin. Nesprin-1 and Nesprin-2 have multiple isoforms, some of which lack the KASH peptide altogether and others result in truncated forms of these proteins missing most of their spectrin repeats. The functional significance of these isoforms is still a matter of speculation.

Nesprin-3α, on the other hand, binds to plectin through its own N-terminal domain. Plectin then tethers the nucleus to intermediate filaments or actin. Additionally, Nesprin-3α has been shown to regulate nuclear size and morphology. The N-terminal domain of Nesprin-3α directly interacts with the ABD of both Nesprin-1 and Nesprin-2, whether this interaction inhibits actin binding of Nesprin-1/2 and what function this serves is still unclear (Lu, 2012). Nesprin-4 and 5 both bind to microtubules through motor proteins kinesin-1 and the dynein-dynactin complex, respectively (Morimoto, 2012; Horn, 2013). Finally, the most recent addition to the KASH domain repertoire in

humans is Lymphoid Restricted Membrane Protein (LRMP or KASH6) that mediates nucleus-centrosome attachment and is involved in pronuclear congression during fertilization (Lindeman, 2012).

SUN-KASH interactions

The main function of SUN-KASH complexes is to mechanically tether the nucleus to the cytoskeleton, thus the SUN-KASH interaction was expected to be quite stable. Structural analysis of LINC complexes highlights how form meets function in this peculiar structure (Figure 2a). The crystal structure of the LINC complex reveals that the oligomeric state of SUN2 is trimeric, and that a single SUN2 trimer can bind up to three KASH peptides simultaneously. The trimer is established through the coiled-coil portion of SUN2 that possesses an undecan repeat and folds in a non-canonical right-handed trimeric coiled-coil (Sosa, 2012; Zhou, 2012). This is followed by the SUN domain; at its core the SUN domain is formed by a beta sandwich with multiple unique features. First, an internal disulfide bond in each SUN2 monomer stabilizes a loop that coordinates a potassium ion in SUN2 trimers.

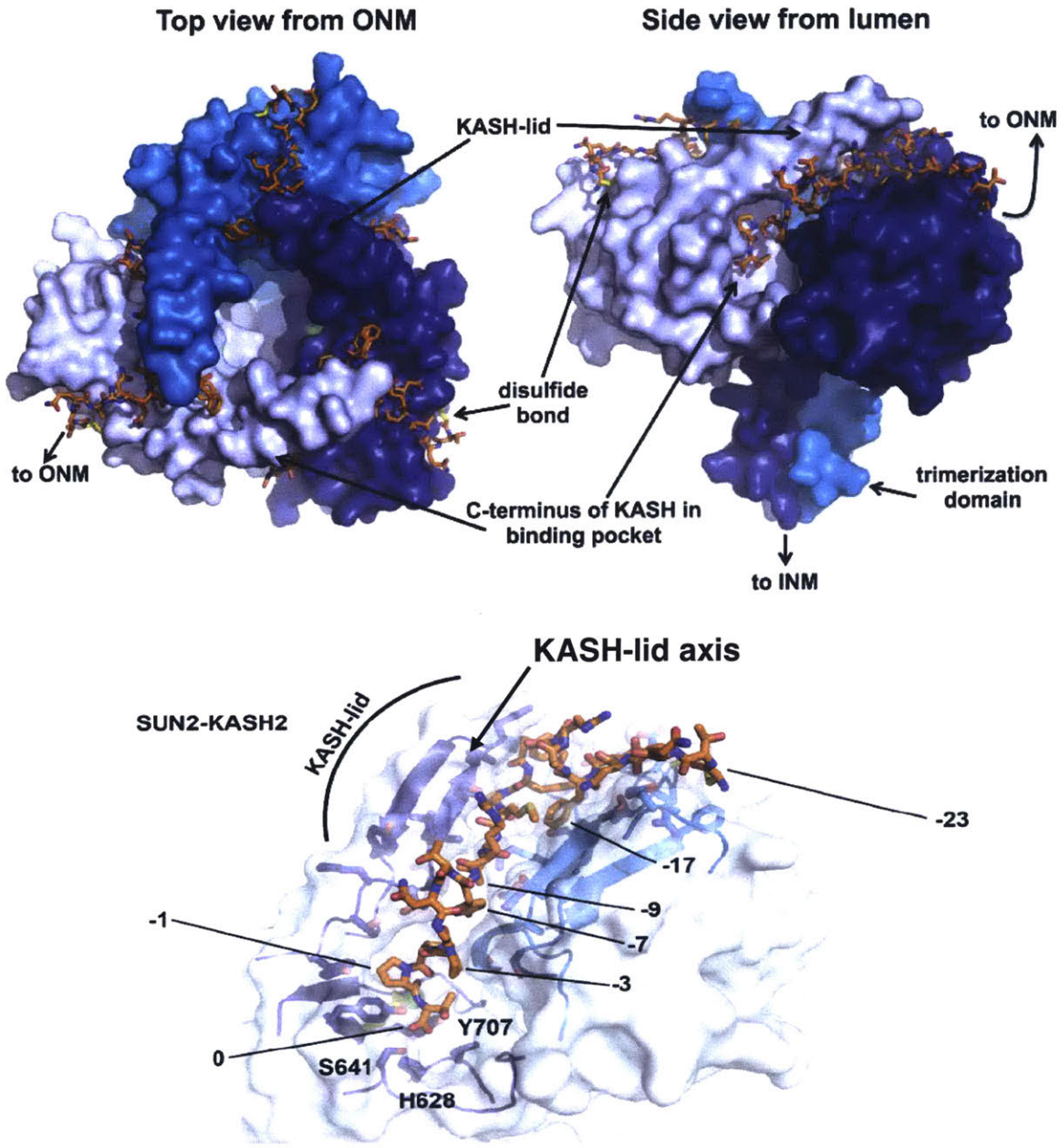


Figure 2. SUN2-KASH2 structural overview. Top, overview of SUN2 trimer bound to KASH2. SUN2 is shown as a surface representation with each subunit of the trimer colored a different shade of blue. KASH peptides are shown in stick representation. Bottom, detailed view of SUN2-KASH2 interactions, the KASH

peptide is sandwiched between two SUN2 monomers, conserved residues that mediate SUN2-KASH2 binding are numbered as described in Sosa, 2012. *Adapted from Sosa 2012.*

Proper coordination of this loop is necessary for LINC complex formation, since changing its register by deleting a residue abolishes KASH binding. The second major feature is the so-called KASH lid, which forms a beta hairpin when bound to KASH but is unstructured in the apo form of SUN2.

The SUN KASH complex structure shows that each KASH peptide is inserted between the interface of two neighboring SUN2 subunits (Figure 2b). This means that SUN2 trimerization is necessary for KASH binding, which has been confirmed with biochemical characterization of SUN KASH in solution (Sosa, 2013; Demircioglu, 2015). The very C terminus of a KASH peptide is inserted into a pocket on the surface of SUN2, revealing why the PPPX motif is so well conserved and why it is always C-terminal. Extending the C terminus of a KASH peptide by a single residue abolishes SUN KASH binding, highlighting the importance of the register of a KASH peptide upon complex formation. The C-terminal X residue at position 0 is usually preceded by three trans prolines at positions -1, -2, and -3 that navigate the surface of a SUN monomer and are partially solvent exposed, these are followed by a solvent exposed region of variable length. Next, the highly conserved residues at positions -7 and -9 pack into the hydrophobic core of a SUN2 monomer and are sandwiched between the core of one SUN2 monomer and the KASH lid of a neighboring SUN monomer. The KASH lid can twist on its axis to accommodate bulkier residues of different KASH peptides. At position

-11, KASH1 and KASH2 possess a highly conserved proline that causes the KASH peptide to sharply kink at about 90° towards the periphery of the SUN trimer, away from the central three fold symmetry axis. As the KASH peptide meanders along the surface of SUN2 a hydrophobic residue on either KASH1 or KASH2 is buried, otherwise this region shows low conservation and mostly weak interactions between SUN2 and KASH1 or KASH2. The last ordered residue seen in the crystal structure is the cysteine at position -23, that forms a disulfide bond with a SUN2 monomer, further stabilizing the interaction and covalently linking SUN2 to either KASH1 or KASH2 (Sosa, 2012).

The extensive contacts formed between SUN and KASH, in combination with the disulfide bond greatly enhance the stability of a LINC complex and presents structural evidence on how they withstand great mechanical forces. Interactions between SUN proteins and KASH peptides is promiscuous, (this work, Chapter 1, Figure 1). Because the function of each SUN-KASH pair is unique, it is unclear what mechanism cells use to distinguish between different LINC complexes and how the assembly and disassembly are regulated. We posit that SUN – KASH1/2 binding generally occurs through the same mechanism, since KASH1 and KASH2 are highly homologous. Structural details of how SUN proteins are able to recognize and bind to KASH peptides with lower sequence homology would provide insight into the regulation of LINC complexes.

The Coiled-coil Domain of SUN2

In order for LINC complexes to form, SUN proteins must span the width of the PNS. The width of the PNS is ~50nm, and it must be spanned by the ~300 amino acids of SUN2 that lie between the TM helix and the SUN domain. In order for these ~300 residues to span 50nm they must form an elongated domain (Sosa, 2012) . Secondary structure prediction indicates that this domain is indeed completely helical, and coiled-coil prediction indicates at least three large regions that are predicted to be coiled-coil, confirming that this domain is indeed elongated. However, structure prediction for trimeric coiled-coils is less accurate than it is for dimeric coils, thus it is a distinct possibility that the entire domain folds into a trimeric coiled-coil. Structures of trimeric coiled-coil of this length are rare in the Protein Data Bank (PDB). Structures of this length can have unexpected features such as a switch in handedness or changes in pitch throughout the length of the structure (Alvarez, 2010). These aberrations from the canonical structure can have specific physiological roles, which is why it is important to elucidate these structural details.

It has been proposed that the coiled-coil domain of SUN2 may function as a spacer that determines the width of the PNS (Sosa, 2012). A deletion of SUN2 in *HeLa* cells causes dilation of the PNS (Crisp, 2006). However, in *C. elegans* the spacing of the PNS was shown to be independent of the length of UNC84 (the *C. elegans* homolog of SUN2), albeit only in non-force bearing cells. In cells under mechanical stress, however, the deletion of the majority of the predicted coiled-coil domain of UNC84 did cause perinuclear dilation (Cain, 2014).

A recently published structure of murine SUN2 showed that the coiled-coil segment proximal to the SUN domain can pack against the globular domain and through interactions with the KASH lid may stabilize a monomeric form of SUN2 (Figure 3) (Nie, 2016). Based on the structure the authors postulate an intrinsic inhibitory mechanism for LINC complexes, although this awaits supporting *in vivo* data. Additionally, a short segment of the trimeric coiled-coil of murine SUN2 was also solved in this study, and that structure revealed a pseudo-stable oligomer. Surprisingly, polar residues were buried in the core of the coiled-coil, decreasing the stability of the oligomeric state of the coiled-coil. It is possible that LINC complexes are disassembled by regulating the oligomeric state of SUN. The only evidence for monomers of SUN domains is this structural work (Nie, et al. 2015). In addition, if the entire perinuclear domain is used SUN2 always forms trimers (this work, Chapter 2, Figure 3; Sosa, 2012; Nie, et al. 2015). The function of the monomeric auto inhibitory state, and if it exists *in vivo* remains unclear. It is clear, however, that the coiled-coil domain of SUN2 is functionally relevant in LINC complex physiology. It has also been suggested that the coiled-coil domain of SUN2 may mediate lateral interactions between parallel LINC complexes in order to form high order arrays of LINC complexes (Zhou, 2012; Wang, 2012).

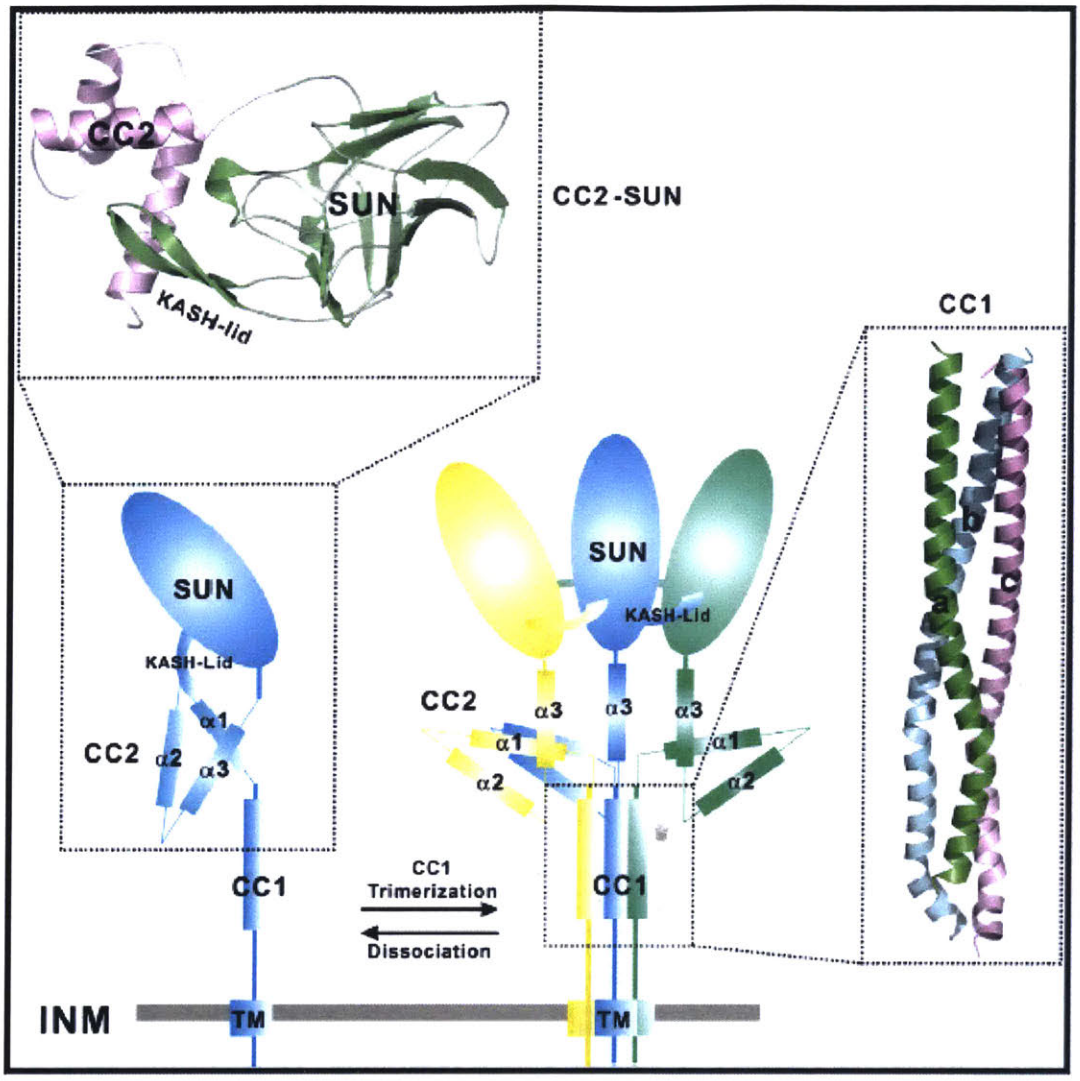


Figure 3. Structure of the coiled-coil domain of murine SUN2. Crystal structure of an autoinhibitory conformation of SUN2 coiled-coil, and of the trimeric coiled-coil of murine SUN2. Cartoon is not to scale, both structures encompass about half of the extended coiled-coil domain of SUN2. *Adapted from Nie, 2016.*

Human SUN proteins with shorter coiled-coil domain are exclusively found in testis, and are called SUN3, SUN4 and SUN5, respectively. During spermatid maturation, the nucleus undergoes dramatic conformational changes. The round

spermatid nuclei mature into an elongated shape, a process that requires SUN4-microtubule interactions (Calvi, 2015; Pasch, 2015). These elongated nuclei have areas of higher membrane curvature in which SUN4 as well as SUN3 are enriched. It is possible that the shorter coiled-coils form or stabilize regions of high membrane curvature in sperm nuclei. Alternatively, SUN3 and SUN4 may be enriched at these sites because their shorter length matches the perinuclear spacing at areas of high curvature.

Assembly of LINC Complexes

SUN and KASH are type II transmembrane proteins, and they are both inserted into the ER. SUN protein insertion probably occurs through the Sec61 channel. KASH proteins are tail anchored to the ER membrane, therefore they are most likely inserted via the GET complex. Upon insertion, the nuclear domain of SUN proteins (the N terminus) is exposed to the cytoplasm and must diffuse through the ER and ONM membranes past the NPC and into the nucleus (Ungritch, 2015). Fully assembled LINC complexes are only observed at the NE, which elicits the interesting question about the regulation of LINC formation. What inhibits LINC complexes to form at undesired location, i.e. throughout the tubular ER?

If LINC complexes are only assembled upon arrival of the SUN protein at the INM, there are a few possible ways to envision this process. One possibility is that the coiled-coil domain of SUN2 can keep the SUN domain in a monomeric state thereby

inhibiting KASH binding until SUN has reached the INM. Alternatively, SUN or KASH may have binding partners that inhibit complex formation until the proteins are both localized appropriately. Since in the structures, the prolines forming the "PPPX" motif are always in a trans conformation a third mechanism may also involve maintaining the highly conserved C-terminal prolines of KASH in an inhibitory cis conformation so long as the properly localized SUN protein has not been found. A proline isomerase may then serve as a trigger for binding.

LINC Complexes and Mechanosensing

The function of LINC complexes extends beyond just transmitting mechanical forces exerted on the nucleus by various cytoskeletal elements. Because cells actively respond to mechanical stimuli, and LINC complexes propagate mechanical stimuli into the nucleus, they can potentially transmit information from a cells surrounding into the nucleus. Various experiments have confirmed that the nucleus actively responds to mechanical stimuli in a number of ways. For example, optical tweezer studies show that isolated nuclei respond to mechanical stress by stiffening the nucleus at the site of force induction. This process is mediated by LINC complexes, specifically SUN2 and Nesprin-1 LINC complexes. Nuclear stiffening occurs through the phosphorylation of Emerin, an INM protein, that interacts with lamin A/C and increases the local density of lamins at the sight of force transmission (Guilluy et al., 2014). Cells under mechanical stress are also known to activate specific transcriptional programs to withstand external forces or respond to them (Alam et al., 2016). Whether these programs are activated through the

dissociation of protein complexes from the nucleoskeleton, or, alternatively, through local reorganization of lamins is still unclear. While there is evidence for both mechanisms, it is unclear whether or not they are coordinated, and to which extent these processes may be tissue-specific.

Laminopathies and LINC Complexes

Mutations that affect the mechanical integrity of the nucleus cause to a number of diseases, collectively termed laminopathies. Both SUN and KASH proteins are implicated in laminopathies, emphasizing their importance in cellular physiology and human health (Haque, 2010; Kandert, 2007). This is reflected at a cellular level by the misshapen nuclei typically associated with laminopathies. At a larger scale, laminopathies primarily affect neuromuscular tissues, and give rise to various muscular dystrophies such as Emery Dreifuss Muscular Dystrophy (EDMD) and Spinocerebral Ataxia. Interestingly, SUN proteins seem to accelerate the progression of laminopathy symptoms, potentially through an overactive DNA damage response, or alternatively through application of force to structurally compromised nuclei (Chen, 2012; Starr, 2012).

Various diseases are specifically related to defects of KASH proteins. A number of pathogenic variants of SYNE1 (the gene encoding Nesprin-1) have been shown to cause Cerebral Ataxia (Gros-Luis, et al., 2006; Attali, 2009). Interestingly, these mutants all cause early termination of Nesprin-1 synthesis, producing KASH-free proteins. A

point mutant of Nesprin-2 has been associated with EDMD, a muscular dystrophy that predominantly affects skeletal muscles. EDMD is most often triggered by mutations in the gene encoding the INM protein emerin (Zhang, et al., 2007; Mejat and Misteli, 2010; Barateau, 2017). The EDMD mutant of Nesprin-2 affects the calponin homology domain of Nesprin-2 and thus actin binding.

A KASH-less form of Nesprin-4 has been shown to cause degenerative loss of hearing. In this case KASH4 is coupled to SUN1, and loss of either protein affects nuclear morphology of inner ear hair cells that leads to loss of hearing (Horn, et al., 2013). In both cases nuclei fail to maintain their basal localization, stressing the diverse pathologies that can arise due to defects, or in this case the absence of a particular LINC complexes.

Higher Order Assemblies of LINC Complexes

Nuclear migration and positioning is a mechanically daunting task that requires the nucleus to transmit and resist large forces generated by the cytoskeleton. As such, a single LINC complex would be unable to withstand these forces and the formation of LINC complex clusters is most likely necessitated (Sosa, 2013). There are various ways that LINC complexes might form high order arrays. First, the coiled-coil domain of SUN proteins lie parallel to one another when LINC complexes are assembled since the perinuclear domain must be extended in order to span the width of the PNS. Lateral contacts between parallel helices, exchanges of helices between adjacent coiled-coils of

LINC complexes, or the presence of a crosslinking protein could provide a mechanism of array formation that is regulated at the PNS (Zhou, 2012). Alternatively, the structure of SUN2 in complex with either KASH1 or KASH2 provides a possible clue. Since KASH1 and KASH2 form a disulfide bond with SUN2, the transmembrane helices of KASH1/2 of a single LINC complex would sit far apart. The distance does not suggest any interaction between KASH domains of a single LINC complex. However, the transmembrane helices of neighboring LINC complexes would be juxtaposed and could contact transmembrane helices of neighboring LINC complexes. The high sequence conservation of the transmembrane helices of KASH proteins suggest that they serve as more than just membrane anchors. If they can form oligomeric structures themselves, they would likely generate 2D clusters of LINC complexes (Sosa, 2013). A third mechanism would rely on a cytoplasmic protein bundling together LINC complexes, perhaps a cytoskeletal connection could mediate this.

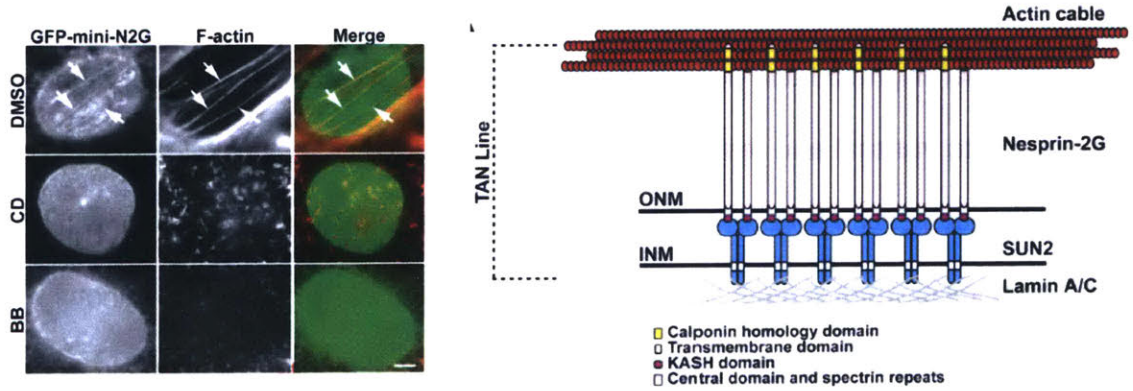


Figure 4. TAN lines are arrays of LINC complexes formed during fibroblast migration. Left, Nesprin-2 and actin colocalizes to form arrays that mediate nuclear positioning and cell migration. This process depends on the presence of actin fibers and myosin. Right, proposed model of TAN lines, SUN2

and Nesprin-2 move the nucleus by tethering the organelle to moving actin cables. *Adapted from Luxton, 2011.*

Indeed, arrays of LINC complexes known as transmembrane actin associated lines (TAN lines) are observed during fibroblast migration (Figure 4). TAN lines are characterized by thick bundles of actin filaments that migrate rearward with respect to the direction the cell is moving. In a SUN2-Nesprin-2 dependent manner the nucleus is tethered to these migrating actin bundles and positioned in the rear of the cell, away from the leading edge (Luxton, 2010). If Nesprin-2 is removed, the thick actin bundles still form but the nucleus fails to properly position itself which results in a failure of cells to migrate. This shows that the actin bundles form and migrate independently from LINC complexes. Migration can be partially rescued by expressing a mini-Nesprin-2 protein that possesses the KASH domain and the actin binding site of Nesprin-2 but lacks most of its spectrin repeats. However, this is only a partial rescue, suggesting that some other unknown element enhances TAN line stability. Interestingly, TAN lines are unique to SUN2 and Nesprin-2, despite the somewhat redundant role of SUN1/SUN2 and of Nesprin-1/Nesprin-2 (Luxton, 2010; Luxton, 2014).

Nesprin-2 and FHOD1

Recently, a formin homology protein, known as FHOD1 has been identified as a component of TAN lines (Kutscheidt, 2014). Formins are typically associated with stress fiber formation and interact with actin through their FH1 and FH2 domains promoting actin filament growth (actin polymerization) (Schonichen, 2013). They are grouped into

various subclasses based on FH2 sequence divergence. The domain organization of FHOD1 is highly modular, with an N-terminal GTPase binding domain (GBD) that shows low sequence conservation with other GBDs of formins. Next is the diaphanous inhibitory domain (DID) that participates in autoregulation of formins. FHOD1 possesses a predicted coiled-coil motif that has been shown to be required for self association, thus like other formins FHOD1 is most likely a dimer (Madrid, 2004). This is followed by the FH1 and FH2, and finally the diaphanous autoregulatory domain (DAD) which folds back and interacts with the DID of the same protein thus inhibiting the actin binding activity of FH2 (Campellone and Welch, 2010). Additionally, FHOD1 binds directly to the Rho kinase ROCK1 which also phosphorylates the DAD domain of FHOD1 preventing auto inhibition. Unregulated FHOD1, similar to FHOD1 in its phosphorylated state, promotes stress fiber formation and gives rise to elongated cell morphology (Hannemann, 2008).

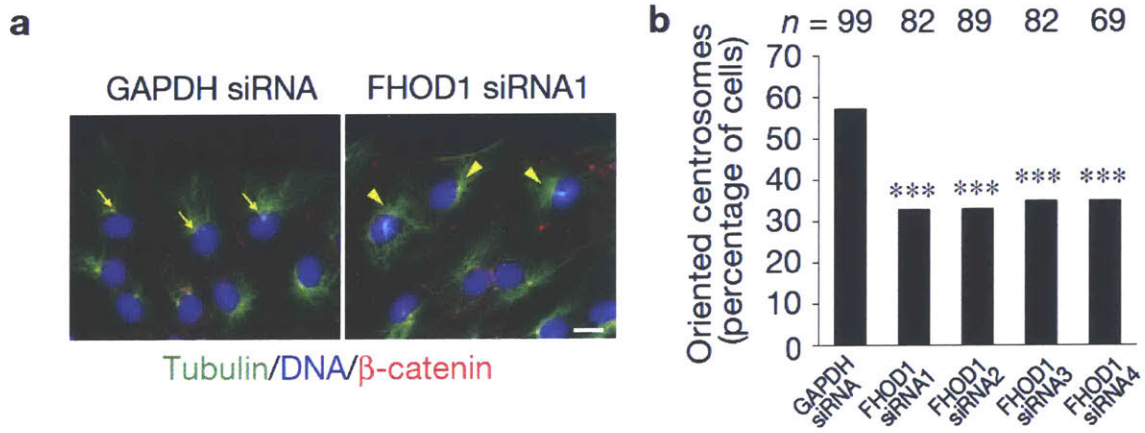


Figure 5. FHOD1 contributes to nuclear positioning and TAN line formation. a) Fibroblasts position their nuclei away from the leading edge as they migrate to the wound site. Yellow arrows show nuclear rearward positioning; arrowheads show nuclei that fail to properly localize in the absence of FHOD1. b) Quantification of a) using four different siRNAs. *Adapted from Kutscheidt, 2016.*

FHOD1, like other formin homology proteins interacts with actin, but FHOD1 bundles preformed actin filaments instead of promoting the elongation of existing ones. Unique among formins, FHOD1 has been shown to bind to Nesprin-2 via spectrin repeats 11 – 12. This interaction is mediated by the structurally divergent GBD and the conserved DID domain of FHOD1. siRNA knockdown of FHOD1 impairs TAN line formation but does not prevent it, and diminishes nuclear rearward localization but does not abolish it (Figure 5a and 5b). This supports a model in Nesprin-2 filaments need multiple points of contact to actin fibers in order to resist the large forces required to move the nucleus (Kutscheidt, 2014).

Nesprin-2 and Fascin-1

A second indirect interaction between Nesprin-2 and actin was recently discovered, and characterized. This is mediated once again by an actin bundling protein, known as Fascin-1. Fascin-1 can directly bind to spectrin repeats 51-53 of Nesprin-2, and together with the FHOD-1 and Nesprin-2 interaction forms an additional contact point between Nesprin-2 and actin filaments. The interaction between Nesprin-2 and Fascin-1 only involves one of the four β-trefoil domains of Fascin-1, namely β-trefoil 3. This interaction does not inhibit the actin bundling role of Fascin-1 (Jayo, 2016).

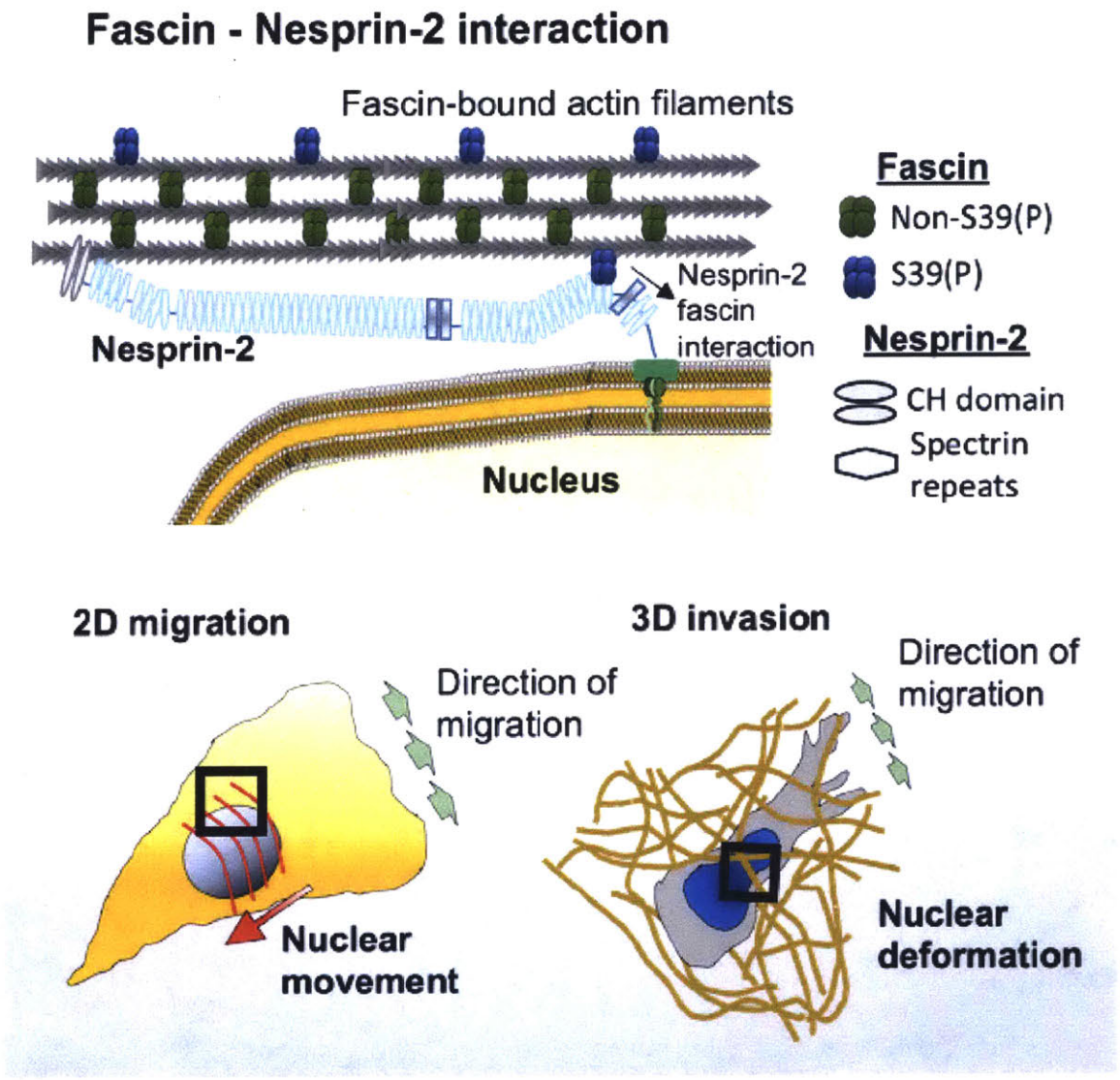


Figure 6. Cartoon model of Fascin-1 and Nesprin-2 interaction. Fascin-1 is an actin bundling protein, that can be regulated by phosphorylation. Fascin-1 can also bind to Nesprin-2, thereby connecting the nucleus to actin filaments. Fascin-1 and Nesprin-2 interactions are important for nuclear positioning, nuclear deformation, and cell migration through confined environments. *Adapted from Jayo, 2016.*

Loss of Fascin-1 reduces nuclear deformability, and diminishes the ability of cells to traverse through tight spaces (Figure 6). Because of the rigidity of the nucleus it must

deform in order for the cell to squeeze through openings that are smaller than the diameter of the organelle. Interestingly, overexpression of the β-trefoil 3 domain of Fascin-1 outcompetes cognate Fascin-1 and N2G interactions uncoupling N2G SR51-53 from actin (Jayo, 2016). The overexpression of β-trefoil is enough to inhibit cell migration through pores smaller than 10 μm. This process is of particular interest in the field of cancer metastasis in which the cell must traverse the dense meshwork of the extracellular matrix, and often must traverse the tight junctions between cells, these features makes Fascin-1 a potential drug target (Chen, 2010; Zhang, 2015). Structural details of the Fascin-1 and Nesprin-2 interaction would provide a great leap forward in the efforts to develop a suitable drug to inhibit Fascin-1 mediated nuclear deformation.

Nesprin-3α

Nesprin-3α is the shortest of the nesprin protein family, possessing a molecular weight of only ~100 kDa. Nesprin-3α is composed of 8 predicted SRs, the N-terminal SR binds to plectin, which in turn can bind to actin through its own CH1/2 domains. The N-terminal SR1 of Nesprin-3α also interacts with the ABD of Nesprin-1 and Nesprin-2, potentially regulating actin binding of these two nesprins (Lu, 2012). In addition, interactions between Nesprin-3α and Nesprin-1/2 have been shown to control nuclear size, perhaps by forming a nesprin/actin network that can mechanically constrict the nucleus (Lu, 2012). Similar to Fascin-1, this control of nuclear size may be important for cell migration under physiological conditions, as well as during metastasis (Jayo, 2016).

Recent work on non-adherent fibroblasts has shown an interesting form of cell migration through 3D matrices, termed A1 amoeboid (Liu, 2015). This consists of cell migrating through small openings through which the nucleus must deform to fit, the actomyosin network pulls the nucleus forward creating positive pressure in front of the nucleus that pushes the membrane on the leading edge outward towards the direction of migration (Petrie, 2014). This piston like method is typically inhibited in malignant tumors that can rely on secreted metalloproteinases to enlarge the size of pores in 3D collagen matrices to migrate through confined spaces, although in the absence of secreted metalloproteinases malignant cells have been shown to use A1 amoeboid mechanism (Petrie, 2017). It is interesting that the intermediate filaments Vimentin, lamin A, and Nesprin-3α are necessary for this piston-like cell migration to occur. This probably means that LINC complexes are also involved in this mechanism, since components on both sides of the NE that interact with LINC complexes are players in this process.

References

Alam, S. G., Zhang, Q., Prasad, N., Li, Y., Chamala, S., Kuchibhotla, R., et al. (2016). The mammalian LINC complex regulates genome transcriptional responses to substrate rigidity. *Nature Publishing Group*, 1–11. http://doi.org/10.1038/srep38063

Alvarez, B. H., Gruber, M., Ursinus, A., Dunin-Horkawicz, S., Lupas, A. N., & Zeth, K. (2010). A transition from strong right-handed to canonical left-handed supercoiling in a conserved coiled-coil segment of trimeric autotransporter adhesins. *Journal of Structural Biology*, *170*(2), 236–245. http://doi.org/10.1016/j.jsb.2010.02.009

Attali, R., Warwar, N., Israel, A., Gurt, I., McNally, E., Puckelwartz, M., et al. (2009). Mutation of SYNE-1, encoding an essential component of the nuclear lamina, is responsible for autosomal recessive arthrogryposis. *Human Molecular Genetics*, *18*(18), 3462–3469. http://doi.org/10.1093/hmg/ddp290

Barateau, A., Vadrot, N., Vicart, P., Ferreiro, A., Mayer, M., Héron, D., et al. (2017). A Novel Lamin A Mutant Responsible for Congenital Muscular Dystrophy Causes Distinct Abnormalities of the Cell Nucleus. *PLoS ONE*, *12*(1), e0169189–18. http://doi.org/10.1371/journal.pone.0169189

Cain, N. E., Tapley, E. C., McDonald, K. L., Cain, B. M., & Starr, D. A. (2014). The SUN protein UNC-84 is required only in force-bearing cells to maintain nuclear envelope

architecture. *The Journal of Cell Biology*, *206*(2), 163–172. http://doi.org/10.1083/jcb.201405081

Calvi, A., Wong, A. S. W., Wright, G., Wong, E. S. M., Loo, T. H., Stewart, C. L., & Burke, B. (2015). SUN4 is essential for nuclear remodeling during mammalian spermiogenesis. *Developmental Biology*, *407*(2), 321–330. http://doi.org/10.1016/j.ydbio.2015.09.010

Campellone, K. G., & Welch, M. D. (2010). A nucleator arms race: cellular control of actin assembly. *Nature Reviews Microbiology*, *11*(4), 237–251. http://doi.org/10.1038/nrm2867

Chen, C.-Y., Chi, Y.-H., Mutalif, R. A., Starost, M. F., Myers, T. G., Anderson, S. A., et al. (2012). Accumulation of the Inner Nuclear Envelope Protein Sun1 Is Pathogenic in Progeric and Dystrophic Laminopathies. *Cell*, *149*(3), 565–577. http://doi.org/10.1016/j.cell.2012.01.059

Chen, L., Yang, S., Jakoncic, J., Zhang, J. J., & Huang, X.-Y. (2010). Migrastatin analogues target fascin to block tumour metastasis. *Nature*, *464*(7291), 1062–1066. http://doi.org/10.1038/nature08978

Crisp, M., Liu, Q., Roux, K., Rattner, J. B., Shanahan, C., Burke, B., et al. (2006). Coupling of the nucleus and cytoplasm. *The Journal of Cell Biology*, *172*(1), 41–53. http://doi.org/10.1083/jcb.200509124

Gros-Louis, F., Dupré, N., Dion, P., Fox, M. A., Laurent, S., Verreault, S., et al. (2006). Mutations in SYNE1 lead to a newly discovered form of autosomal recessive cerebellar ataxia. *Nature Genetics*, *39*(1), 80–85. http://doi.org/10.1038/ng1927

Guilluy, C., Osborne, L. D., Van Landeghem, L., Sharek, L., Superfine, R., Garcia-Mata, R., & Burridge, K. (2014). Isolated nuclei adapt to force and reveal a mechanotransduction pathway in the nucleus. *Nature Cell Biology*, *16*(4), 376–381. http://doi.org/10.1038/ncb2927

Hannemann, S., Madrid, R., Stastna, J., Kitzing, T., Gasteier, J., Schönichen, A., et al. (2008). The Diaphanous-related Formin FHOD1 associates with ROCK1 and promotes Src-dependent plasma membrane blebbing. *Journal of Biological Chemistry*, *283*(41), 27891–27903. http://doi.org/10.1074/jbc.M801800200

Haque, F., Mazzeo, D., Patel, J. T., Smallwood, D. T., Ellis, J. A., Shanahan, C. M., & Shackleton, S. (2010). Mammalian SUN Protein Interaction Networks at the Inner Nuclear Membrane and Their Role in Laminopathy Disease Processes. *Journal of*

Biological Chemistry, *285*(5), 3487–3498. http://doi.org/10.1074/jbc.M109.071910

Horn, H. F., Brownstein, Z., Lenz, D. R., Shivatzki, S., Dror, A. A., Dagan-Rosenfeld, O., et al. (2013). The LINC complex is essential for hearing. *Journal of Clinical Investigation*, 1–11. http://doi.org/10.1172/JCI66911

Jayo, A., Malboubi, M., Antoku, S., Chang, W., Ortiz-Zapater, E., Groen, C., et al. (2016). Fascin Regulates Nuclear Movement and Deformation in Migrating Cells. *Developmental Cell*, *38*(4), 371–383. http://doi.org/10.1016/j.devcel.2016.07.021

Kabachinski, G., & Schwartz, T. U. (2015). The nuclear pore complex - structure and function at a glance. *Journal of Cell Science*, *128*(3), 423–429. http://doi.org/10.1242/jcs.083246

Kandert, S., Luke, Y., Kleinhenz, T., Neumann, S., Lu, W., Jaeger, V. M., et al. (2007). Nesprin-2 giant safeguards nuclear envelope architecture in LMNA S143F progeria cells. *Human Molecular Genetics*, *16*(23), 2944–2959. http://doi.org/10.1093/hmg/ddm255

Knockenhauer, K. E., & Schwartz, T. U. (2016). The Nuclear Pore Complex as a Flexible and Dynamic Gate. *Cell*, *164*(6), 1162–1171. http://doi.org/10.1016/j.cell.2016.01.034

Kutscheidt, S., Zhu, R., Antoku, S., Luxton, G. W. G., Stagljar, I., Fackler, O. T., & Gundersen, G. G. (2014). FHOD1 interaction with nesprin-2G mediates TAN line formation and nuclear movement. *Nature Cell Biology*, *16*(7), 708–715. http://doi.org/10.1038/ncb2981

Liu, Y. J., Le Berre, M., Lautenschlaeger, F., Maiuri, P., Callan-Jones, A., Heuzé, M., et al. (2015). Confinement and Low Adhesion Induce Fast Amoeboid Migration of Slow Mesenchymal Cells. *Cell*, *160*(4), 659–672. http://doi.org/10.1016/j.cell.2015.01.007

Lu, W., Schneider, M., Neumann, S., Jaeger, V.-M., Taranum, S., Munck, M., et al. (2012). Nesprin interchain associations control nuclear size. *Cellular and Molecular Life Sciences*, *69*(20), 3493–3509. http://doi.org/10.1007/s00018-012-1034-1

Luxton, G. W. G., Gomes, E. R., Folker, E. S., Vintinner, E., & Gundersen, G. G. (2010). Linear Arrays of Nuclear Envelope Proteins Harness Retrograde Actin Flow for Nuclear Movement. *Science*, *329*(5994), 956–959. http://doi.org/10.1126/science.1189072

Luxton, G. W. G., Gomes, E. R., Folker, E. S., Worman, H., & Gundersen, G. G. (2014). TAN lines. *Nucleus*, *2*(3), 173–181. http://doi.org/10.4161/nucl.2.3.16243

Madrid, R., Gasteier, J. E., Bouchet, J., Schröder, S., Geyer, M., Benichou, S., & Fackler, O. T. (2004). Oligomerization of the diaphanous-related formin FHOD1 requires a coiled-coil motif critical for its cytoskeletal and transcriptional activities. *FEBS Letters*, *579*(2), 441–448. http://doi.org/10.1016/j.febslet.2004.12.009

Mejat, A. & Misteli, T., (2010). LINC complexes in health and disease, *Nucleus* 1:1 40–52.

Nie, S., Ke, H., Gao, F., Ren, J., Wang, M., Huo, L., et al. (2016). Coiled-Coil Domains of SUN Proteins as Intrinsic Dynamic Regulators. *Structure/Folding and Design*, *24*(1), 80–91. http://doi.org/10.1016/j.str.2015.10.024

Pasch, E., Link, J., Beck, C., Scheuerle, S., & Alsheimer, M. (2015). The LINC complex component Sun4 plays a crucial role in sperm head formation and fertility. *Biology Open*, *4*(12), 1792–1802. http://doi.org/10.1242/bio.015768

Petrie, R. J., Koo, H., & Yamada, K. M. (2014). Generation of compartmentalized pressure by a nuclear piston governs cell motility in a 3D matrix. *Science*, *345*(6200), 1062–1065. http://doi.org/10.1126/science.1256965

Petrie, R. J., Harlin, H. M., Korsak, L. I. T., & Yamada, K. M. (2017). Activating the nuclear piston mechanism of 3D migration in tumor cells. *The Journal of Cell*

Biology, *216*(1), 93–100. http://doi.org/10.1083/jcb.201605097

Schonichen, A., Mannherz, H. G., Behrmann, E., Mazur, A. J., Kuhn, S., Silvan, U., et al. (2013). FHOD1 is a combined actin filament capping and bundling factor that selectively associates with actin arcs and stress fibers. *Journal of Cell Science*, *126*(8), 1891–1901. http://doi.org/10.1242/jcs.126706

Sosa, B. A., Rothballer, A., Kutay, U., Schwartz, T. U. (2012). LINC Complexes Form by Binding of Three KASH Peptides to Domain Interfaces of Trimeric SUN Proteins. Cell, 149, 1035-1047. doi:10.1016/j.cell.2012.03.046

Sosa, B. A., Kutay, U., & Schwartz, T. U. (2013). Structural insights into LINC complexes. *Current Opinion in Structural Biology*, *23*(2), 285–291. http://doi.org/10.1016/j.sbi.2013.03.005

Starr, D. A., & Fridolfsson, H. N. (2010). Interactions Between Nuclei and the Cytoskeleton Are Mediated by SUN-KASH Nuclear-Envelope Bridges. *Annual Review of Cell and Developmental Biology*, *26*(1), 421–444. http://doi.org/10.1146/annurev-cellbio-100109-104037

Starr, D. A. (2012). Laminopathies: Too Much SUN Is a Bad Thing. *Current Biology*, *22*(17), R678–R680. http://doi.org/10.1016/j.cub.2012.06.070

Ungricht, R., Klann, M., Horvath, P., & Kutay, U. (2015). Diffusion and retention are major determinants of protein targeting to the inner nuclear membrane. *The Journal of Cell Biology*, *209*(5), 687–704. http://doi.org/10.1083/jcb.201409127

Wang, W., Shi, Z., Jiao, S., Chen, C., Wang, H., Liu, G., et al. (2012). Structural insights into SUN-KASH complexes across the nuclear envelope. *Cell Research*, *22*(10), 1440–1452. http://doi.org/10.1038/cr.2012.126

Zhang, Q., Bethmann, C., Worth, N. F., Davies, J. D., Wasner, C., Feuer, A., et al. (2007). Nesprin-1 and -2 are involved in the pathogenesis of Emery Dreifuss muscular dystrophy and are critical for nuclear envelope integrity. *Human Molecular Genetics*, *16*(23), 2816–2833. http://doi.org/10.1093/hmg/ddm238

Zhang, T., Ling, X.-L., Hou, X.-M., & Zhao, D. (2015). Clinicopathological significance of fascin-1 expression in patients with non-small cell lung cancer. *OncoTargets and Therapy*, 1589–7. http://doi.org/10.2147/OTT.S84308

Zhou, Z., Du, X., Cai, Z., Song, X., Zhang, H., Mizuno, T., et al. (2012). Structure of Sad1-UNC84 Homology (SUN) Domain Defines Features of Molecular Bridge in

Nuclear Envelope. *Journal of Biological Chemistry*, *287*(8), 5317–5326.

http://doi.org/10.1074/jbc.M111.304543

Chapter 1: The Structures of SUN KASH Complexes Reveal Two Alternative Binding Modes

Summary

Linker of nucleoskeleton to cytoskeleton (LINC) complexes are molecular tethers that span the nuclear membranes and physically connect the nucleus to the cytoskeleton. They are composed of outer nuclear membrane spanning KASH-containing proteins, and SUN-proteins that traverse the inner nuclear membrane. LINC complexes mediate the transmission of mechanical forces into the nucleus and facilitate processes such as nuclear anchorage, nuclear migration, and homologous chromosome pairing during meiosis. Here we present high resolution X-ray crystal structures of SUN2 in complex with KASH3, KASH4, and KASH5, respectively. Compared to SUN2:KASH1/2, we observe an alternative binding mode that maintains core interactions between SUN and KASH, but inhibits the formation of a disulfide bond with SUN2. We establish that SUN2 can bind to different KASH peptides simultaneously, which may have distinct functional consequences. The lack of a covalent interaction between SUN2 and KASH3/4/5 hints at a regulatory mechanism mediated by force transmission.

Introduction

The contents of the nucleus in eukaryotic cells are physically separated from the cytoplasm by a double membrane bilayer, the nuclear envelope (NE). The outer nuclear membrane (ONM) faces the cytoplasm and is contiguous with the endoplasmic reticulum. The inner nuclear membrane (INM) is separated from the ONM by the perinuclear space, which has a relatively uniform thickness of ~40-50 nm (Franke, 1981; Cohen, 2002) INM and ONM are fused at distinct circular openings throughout the NE, which are occupied by nuclear pore complexes (NPCs). In metazoa, the nuclear face of the INM is lined with a filamentous protein network, the lamina (Turgay, 2017). Lamins, the proteins that constitute the lamina, provide mechanical stiffness to the nucleus and interact with peripheral heterochromatin (Bridger, 2007). The structural integrity of the NE is key in maintaining cellular homeostasis, and in a cell's ability to respond to mechanical stimuli (Starr, 2012; Barateau, 2017).

Communication between the nucleus and the cytoplasm occurs mainly through two mechanisms. First, NPCs facilitates molecular exchange between the nucleus and cytoplasm (Kabachinski, 2015; Knockenhauer 2016). In addition, information can be transmitted into the nucleus through mechanical signaling mediated by molecular tethers known as the Linkers of Nucleoskeleton to Cytoskeleton (LINC) complexes (Starr, 2010).

The core of LINC complexes is formed in the PNS, between two proteins. SUN (Sad1 and UNC84) proteins form the nuclear half of LINC complexes, while KASH (Klarsicht, ANC-1 and Syne Homology) proteins are tail anchored at the ONM and constitute the cytoplasmic end of LINC complexes (Razafsky & Hodzic, 2009; Starr & Fischer, 2005).

SUN proteins are conserved in eukaryotes (Malone, 1999; Starr, 2010). They possess an N-terminal domain that projects into the nucleus, followed by a transmembrane helix that spans the INM. The nuclear domain of SUN proteins interacts with lamins and heterochromatin anchoring SUN at the INM. The perinuclear domain of SUN proteins consists of an extended coiled-coil that leads to the conserved C-terminal SUN domain which binds to KASH proteins (Sosa 2012; Chang, 2015). Presently, five SUN proteins have been identified in humans. SUN1 and 2 are present in all tissues and possess partially redundant functions, while SUN3, 4, and 5 are expressed in testis and possess a shorter coiled-coil domain than their SUN1/2 counterparts (Crisp, 2006; Wang, 2006; Sosa, 2013; Lindeman 2012; Pasch, 2014).

The tail anchored KASH proteins project their ~30 C-terminal residues into the PNS. This perinuclear element alone is known as the KASH 'peptide', while in combination with the preceding transmembrane helix it is referred to as the KASH 'domain'. Their C-terminal "PPPX" motif is highly conserved among KASH

peptides and is required for binding to SUN proteins. Extending this motif by a single residue abolishes binding to SUN (Sosa, 2012). Currently, six KASH domain-containing proteins have been identified in humans. KASH1-4 are known as nesprins (nuclear envelope spectrin repeats), KASH5 is also known as CCDC155, and KASH6 is known as lymphoid restricted membrane protein (LRMP). Nesprin-1 and 2 have various isoforms, the largest of which can bind directly to actin through an N-terminal calponin homology (CH) domain (Zhang, 2002). Nesprin-1 interacts with Matrin3 which in turn associates with mRNA processing bodies, and regulates miRNA mediated gene silencing. Suggesting that Nesprin-1 function may extend beyond just mechanical tethering (Rajgor, 2016). Nesprin-2, on the other hand, can bind to actin indirectly through two additional sites. The formin homology domain protein FHOD1 interacts with Nesprin-2 through spectrin repeats 11-12, while itself interacting with actin (Kutscheidt, 2014). The second indirect actin-binding site is mediated by Fascin-1 that interacts with Nesprin-2's INM proximal spectrin repeats 51-53 (Jayo, 2016). These multi-point interactions between Nesprin-2 and actin mediate nuclear positioning during fibroblast migration, where bundles of actin filaments are tethered to the nucleus through interactions with Nesprin-2 (Kutscheidt, 2014).

Nesprin-3, in contrast to Nesprin-1 and -2, binds to intermediate filaments, rather than actin. The interaction is mediated by the giant protein plectin, which

binds the N-terminal domain of Nesprin-3 (Morgan, 2011). Nesprin-4 and 5 both bind to microtubules through motor proteins kinesin-1 and dynein-dynactin complex, respectively (Morimoto, 2012; Horn, 2013b; Chang, 2015). The most recent addition to the spectrum of KASH-containing proteins in humans is Lymphoid Restricted Membrane Protein (LRMP or KASH6) that is involved in nucleus-centrosome attachment and is pivotal in pronucleus congression during fertilization (Lindeman, 2012). Underlying their importance in cellular physiology, defects in LINC complexes have been associated to a number of diseases. KASH-less forms of Nesprin-1 and Nesprin-2 are associated with neurological and muscular defects such as Emery Dreifuss muscular dystrophy (EDMD) and spinocerebral ataxia (Attali, 2009; Zhang, 2007), while mutants of Nesprin-4 cause progressive high-frequency hearing loss (Horn, 2013a). Understanding how LINC complexes are regulated is pivotal to our understanding of these diseases.

The structures of the human hetero-hexameric SUN2-KASH1 and SUN2-KASH2 complexes provided insight into how they facilitate transmission of large mechanical forces (Sosa, 2012; Wang, 2012). The C-terminal SUN domain of SUN2 folds into a compact β sandwich, which trimerizes into a trefoil, supported by a preceding three-stranded coiled-coil element. KASH1 or 2 bind at the interface between adjacent SUN domains. A loop, disordered in the apo-structure, folds into a beta hairpin and clamps down the KASH peptide to

stabilize the interaction. This element is called the 'KASH lid' (Sosa, 2012). Additionally, each SUN2 trimer binds up to 3 KASH peptides at once, increasing the interaction strength and distributing expected mechanical force across three discrete sites within the complex. Finally, a conserved disulfide bond between SUN2 and KASH1/2 likely enhances the stability of the interaction even further.

In this study, we were interested in analyzing whether the wide spectrum of potential SUN-KASH pairings would follow the pattern established by the SUN2-KASH1/2 complexes. We sought to address this question using structural and biochemical tools. We solved high-resolution crystal structures of SUN2 in complex with KASH3, KASH4, and KASH5, to complement the published SUN2-KASH1 and SUN2-KASH2 complexes. The new structures reveal distinct binding interactions of SUN2 with different KASH peptides. Taken together, this data feeds into the notion that humans have evolved an elaborate LINC-complex network, with the possibility for regulation on multiple levels.

Results

SUN2 binds to all 6 human KASH peptides

To begin our study, we first examined direct binding of human SUN2 to all six currently known KASH proteins. We co-expressed SUN2 with the predicted perinuclear KASH peptides of Nesprin-1, -2, -3, -4, CCDC155, or LRMP in *E. coli*. From here on, we will refer to these peptides as KASH1-6 for simplicity. KASH peptides were N-terminally fused to 6 histidines, followed by a Gb1 solubility tag (Hammarström, 2006), and a human rhinovirus 3C cleavage site. SUN2 was expressed with a trimerizing GCN4 tag (Ciani, 2010), but no affinity tag. SUN2-KASH complexes were first nickel affinity purified, thereby eliminating apo-SUN2. After proteolytically cleaving Gb1 from KASH, SUN2-KASH complexes were separated on size exclusion chromatography, with free Gb1 eluting as a separate peak. This way, we were able to verify stable interaction of SUN2 with all 6 KASH peptides.

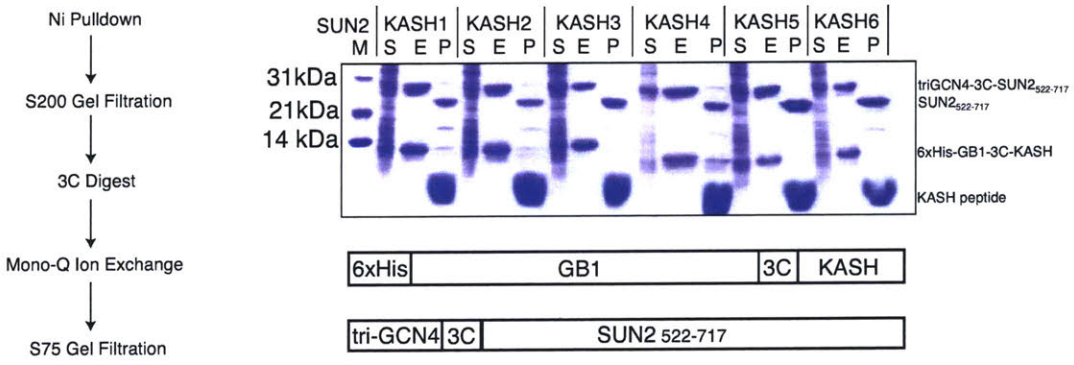
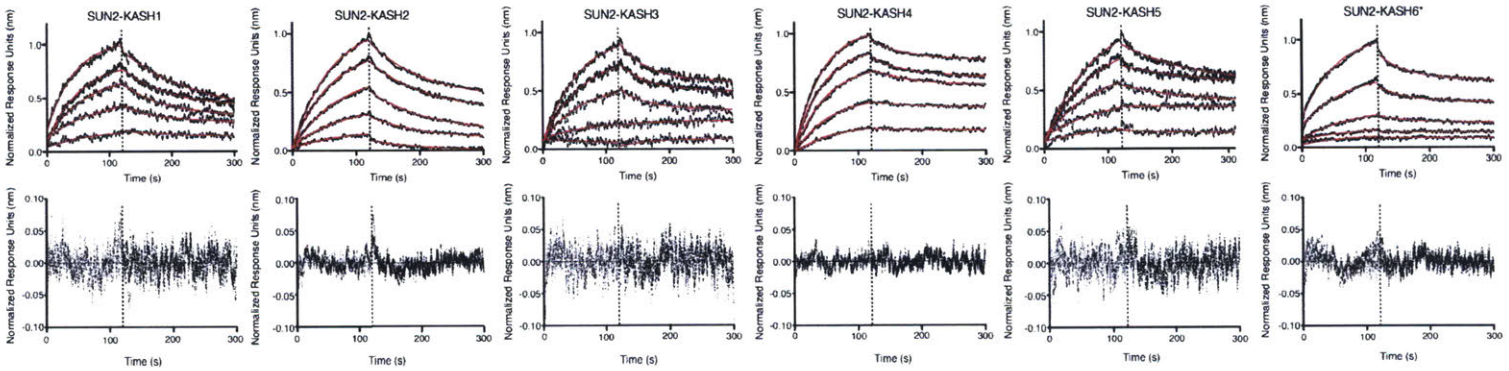


Figure 1. Overview of SUN2-KASH constructs and purification. Left, flowchart of complex purification. Gel shows representative purification of SUN2 in complex with all six KASH peptides.

Bottom, shows a schematic diagram of the constructs purified. M is protein marker, S is the soluble fraction of the lysate, E is the nickel elution, and P is the purified complex after one size exclusion step and cleavage of the solubility tags.

The conserved core of KASH peptides bind SUN2 with similar affinity

Next, we examined the binding affinity for SUN2 and all 6 KASH peptides using biolayer-interferometry (BLI). Similar to what has been previously reported, we directionally immobilized GFP fused KASH to streptavidin-coated optical tips using an N-terminal, biotinylated AviTag (Wang, 2012; Ölstund, 2009). Trimeric apo-SUN2 were used as analytes. All binding data were fit to single exponential association and dissociation kinetics. Our experiments show that KASH1, 2, 3, 4, and 5 bind SUN2 with K_D's in the range of 0.7 μM – 4 μM, while KASH6 is a weaker binder, having a K_D in the order of 15 μM (Figure 2). These results are consistent with our analysis of SUN2-KASH6 complexes that show incomplete occupancy and disassembly of the complex during size exclusion chromatography (data not shown). Primarily for this reason, SUN2-KASH6 is excluded from our structural analysis of LINC complexes.



	K_{on} $M^{-1}s^{-1}$	K_{off} s^{-1}	K_D M	K_D error M
SUN2-KASH1	1.41E+04	7.78E-03	5.51E-07	3.68E-08
SUN2-KASH2	5.51E+03	1.13E-02	2.05E-06	6.31E-08
SUN2-KASH3	4.91E+03	1.25E-02	2.55E-06	7.25E-07
SUN2-KASH4	1.70E+07	8.40E-02	4.93E-07	1.98E-08
SUN2-KASH5	3.71E+03	9.89E-03	2.67E-06	2.07E-07
SUN2-KASH6	7.26E+02	1.07E-02	1.48E-05	7.92E-07

Figure 2. Kinetic analysis of SUN2-KASH interactions. Top panels show normalized binding and dissociation curves. Red lines are the curves generated from fitting the data. Bottom panels are the residuals from the fitting analysis. Kinetic parameters and dissociation constant from fitting are shown in the table. Concentration range used was 1000nM, 750nM, 500nM, 250nM, and 100nM. *Indicates that the concentrations used in this experiment were 10000nM, 5000nM, 2500nM, 1000nM 750nM.

Crystallographic analysis of SUN2 KASH complexes

In order to test whether similar binding kinetics reflect similar binding modes, we set out to structurally characterize the SUN2-KASH complexes by X-ray crystallography. Using a purification strategy discussed previously (Demircioglu, 2015), our initial attempts yielded poorly diffracting crystals of SUN2-KASH3 and SUN2-KASH5. Based on PISA analysis of the crystal packing interactions of these poorly diffracting crystals, we designed SUN2 mutants that could potentially enhance crystal packing, and hence produce better diffracting crystals. The point mutant that yielded the best crystals of SUN2 in complex with KASH3 and KASH5 mutated SUN2 at Q534D and residues at 683–685 of SUN2 were replaced by T683G, M684R, and A685G. For SUN2-KASH4 crystals, wildtype SUN2 was used.

Data collection and refinement statistics

Protein	SUN2$_{500-717}$	SUN2$_{522-717}$ KASH1	SUN2$_{522-717}$ KASH2	SUN2$_{522-717}$ KASH3	SUN2$_{522-717}$ KASH4	SUN2$_{522-717}$ KASH5
Organism	*H. sapiens*	*H. sapiens*	*H. sapiens*	*H. sapiens*	*H. sapiens*	*H. sapiens*
Data collection						
Space group	R 3 2	R 3 2	R 3 2	P 63 2 2	P 63 2 2	C 2
a, b, c (Å)	79.2, 79.2, 199.5	79.6, 79.6, 256.4	79.3, 79.3, 260.0	79.0, 79.0, 127.4	80.8, 80.8, 173.9	135.6, 78.4, 94.6
α, β, γ (°)	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 118.7, 90.0
Wavelength (Å)	0.9792	0.9792	0.9792	0.9791	0.9792	0.9791
Completeness (%)	100.0 (98.0)	99.6 (97.7)	99.9 (99.8)	100.0 (99.8)	100.0 (99.9)	97.0 (72.4)
Redundancy	3.7 (3.0)	3.5 (3.4)	5.9 (6.1)	35.5 (28.5)	34.8 (30.0)	3.0 (1.5)
Rp.i.m. (%)	2.8 (100.0)	4.7 (43.4)	2.5 (30.7)	0.9 (50.0)	2.6 (82.0)	4.6 (36.5)
I/σ	31.1 (0.7)	24.9 (1.5)	21.0 (2.7)	52.4 (1.7)	47.3 (1.4)	16.3 (1.3)
CC1/2 (%)	99.6 (56.7)	99.8 (62.0)	99.9 (77.9)	100.0 (94.8)	99.9 (70.6)	99.8 (70.5)
Refinement						
Resolution range (Å)	56.5 – 1.9	45.0 -2.3	36.1 -2.7	43.7 – 1.8	86.9 – 1.5	45.9 – 2.5
Rwork	17.4	17.7	18.2	20.0	15.3	22.7
Rfree	22.4	22.7	24.1	23.1	18.5	27.4
Rmsd						
Bond length (Å)	0.014	0.018	0.013	0.015	0.014	0.006
Bond angles(°)	1.176	1.916	1.732	1.176	1.454	0.958
Ramachandran Plot						
Favored (%)	95.6	94.5	93.1	95.6	95.9	92
Allowed (%)	3.4	5.5	6.4	3.4	3.2	6.8
Outliers (%)	1.0	0.0	0.5	0.9	1.0	1.3

Table 1. Data collection and refinement statistics. Values shown in parenthesis correspond to the highest resolution shell. For completion, statistics of SUN2 in complex with KASH1 and KASH2 are also shown (Sosa, 2012).

Using the combination of a new tagging strategy, the purification scheme described here, and point mutants for some cases, we obtained well diffracting crystals for apo-SUN2, SUN2-KASH3, SUN2-KASH4, and SUN2-KASH5 (Table 1) (Demircioglu, 2015). Expectedly, all structures pack into similar crystal lattices as those observed in previous studies, therefore they were easily solved by molecular replacement (Sosa, 2012). The search model used was SUN2 in complex with KASH1 (pdb code 4DXR) in which the KASH lid and the KASH peptide were removed for the search. Both the KASH lid and the KASH peptide were then manually built into the resulting electron density. The general features of SUN-KASH engagement are maintained in all complexes (Figure 3). SUN2 forms a trimer, even in the apo-form, and all KASH peptides bind at the interface between adjacent SUN2 monomers despite sequence variations between KASH peptides. However, we also observe a distinct difference between SUN2 bound to KASH1 or 2 as opposed to KASH3, 4, or 5 (Figure 3). While KASH1 and 2 kink toward the periphery of the SUN domain, KASH3, 4, and 5 extend towards the neighboring KASH lid from the adjacent SUN2 monomer. The solvent exposed surface of the KASH lid shows strong conservation that remained unexplained by previous structures of SUN-KASH complexes. The binding mode we observe between SUN2 and KASH3, 4, and 5 however show that the KASH peptide extends toward these residues, providing a basis for the observed conservation of the solvent exposed face of the KASH lid.

When KASH1, 2, 3, 4, and 5 are aligned, residues between positions 0 to -10 only slightly diverge, with an RMSD of about 0.8 Å over 11 residues. The main differences occur at position -11. In a subset of KASH peptides this residue is conserved as a proline, while it is non-conserved in the other subset (Figure 4). Superimposing the structures of KASH2 and KASH4 it is clear that Pro -11 determines the directionality of the KASH peptide upon exiting from under the KASH lid (Figure 4). Interestingly, the N-terminal portions of KASH peptides lacking Pro -11 are the least conserved. This is explained by the fact that these KASH peptides do not share the same binding surface with SUN2 once they exit from under the KASH lid. Both KASH1 and 2 take a sharp 90° turn at Pro -11 that leads the KASH peptide away from the 3-fold symmetry axis and instead over the immunoglobulin (Ig) fold of a neighboring SUN monomer and towards Cys563 where a disulfide bridge is formed. Because KASH3, 4, and 5 lack Pro -11, they adopt the alternate binding conformation described here (Figure 3).

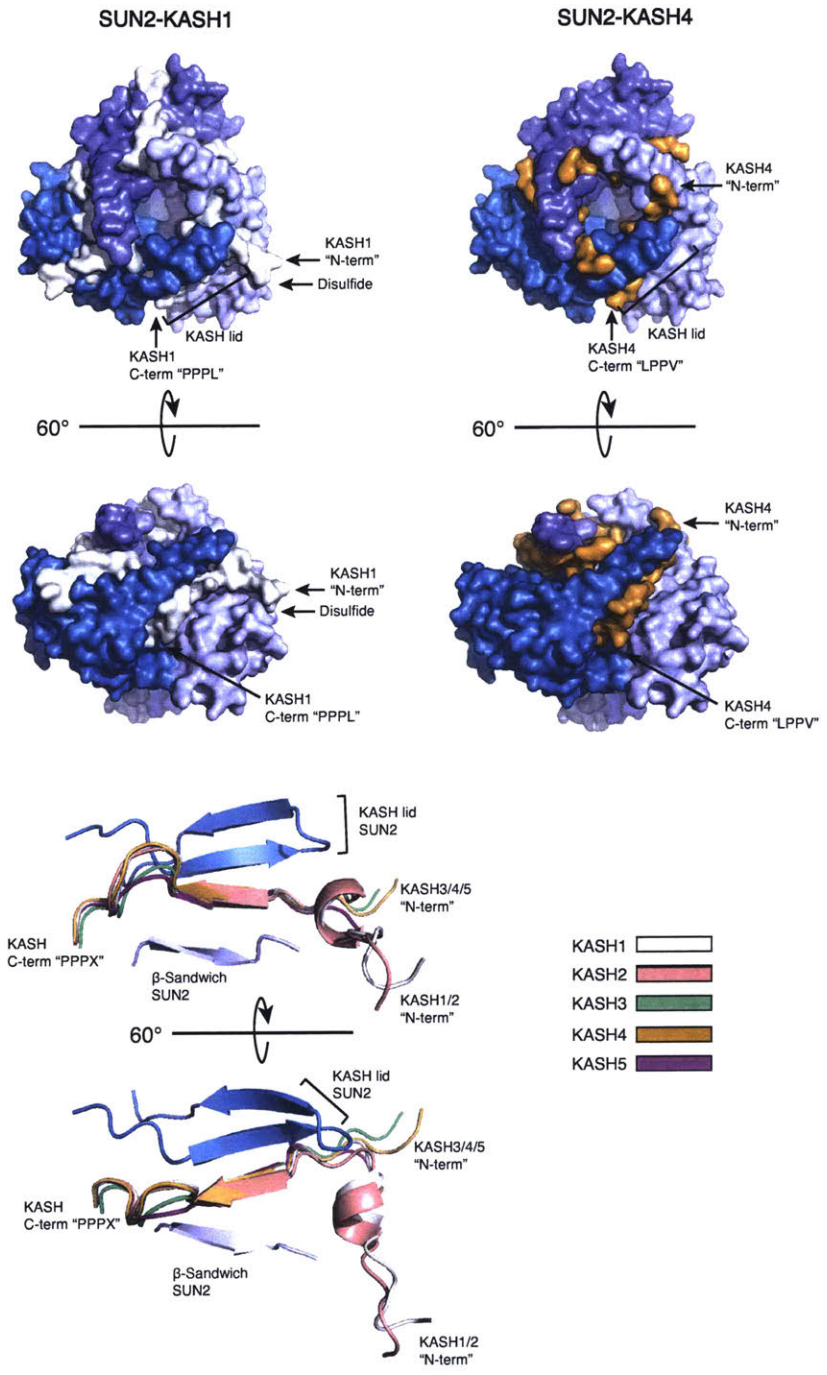


Figure 3. SUN2 can bind to KASH peptides in two alternate conformations. Top, overview of SUN2 in complex with KASH1 and with KASH4. Bottom, cartoon representation of the two alternative binding modes, focusing on differences between KASH peptides.

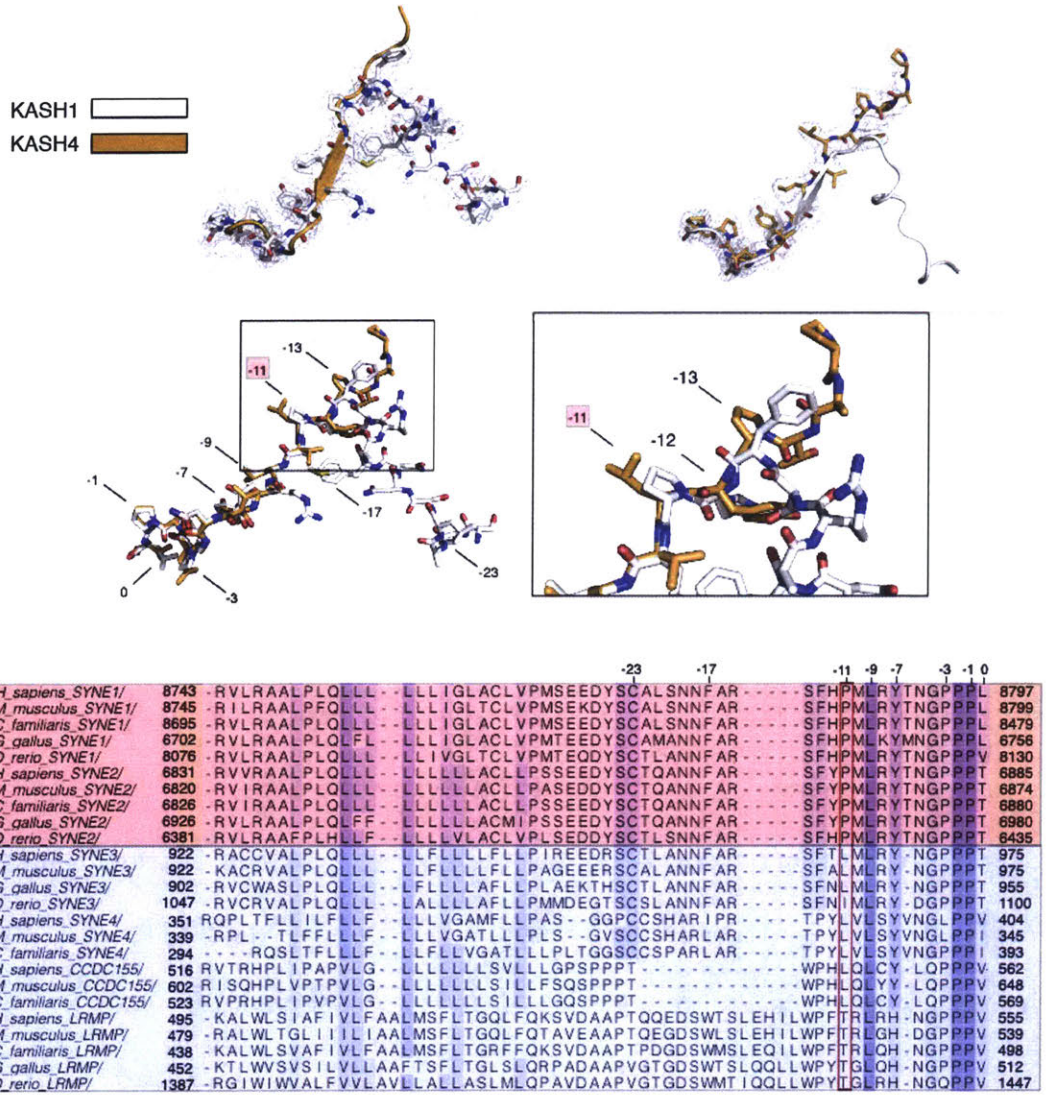


Figure 4. Proline at position -11 determines KASH binding mode. Top, alignment of KASH1 and KASH4 enveloped in electron density clearly highlight the two distinct binding modes. Middle, residues that are conserved between KASH peptides interact nearly identically with SUN2, proline -11 is highlighted as mediating binding modes. Bottom, multiple sequence alignment between KASH1-6 of vertebrates, proline -11 is consistently present in KASH1 and 2 but absent from KASH3-6.

The KASH lid itself only has minor changes in conformation, such as a slight rotation or a small displacement on its axis to accommodate bulkier, or more compact side chains present in each KASH peptide. Notably, KASH sequences between position -4 through -6 of different lengths are easily accommodated and are partially solvent exposed. KASH3/5 possess two residues here whereas KASH1/2/4 have 3. To accommodate the lacking residues at positions -5 and -4, KASH3 and 5 form a shorter loop that still maintains the correct register of the remainder of the KASH peptide. Residues preceding this loop are well aligned between different KASH peptides as well as residues after this loop. The tyrosine at positions -7 clearly functions as an anchor that is critical in determining the register of the KASH peptide for residues -8 to -11 and facilitates the hydrogen-bonding pattern between the KASH peptide and the KASH lid. The solvent exposed loop between positions -4 to -6 together with the anchoring by Tyr -7 grants flexibility in the sequence and length of residues that can be accommodated between the Tyr -7 and the conserved PPPX motif.

The cation loop of SUN2

We can clearly resolve the electron density of the cation loop in all the structures reported here, and it is particularly well ordered in the high resolution structures of the SUN2-KASH3 and SUN2-KASH4 complexes. The cation loop of SUN2 in complex with KASH3 is identical to what has been observed in the apo form of SUN2, as well as SUN2 bound to KASH1/2 (Sosa, 2012). In these

structures a potassium ion is coordinated at the center of the cation loop by the carbonyl groups of Val 590, Gln 593, Asp 595, Asn 600, Tyr 707, and by a well ordered water molecule. Surprisingly, the ion loop of SUN2 in complex with KASH4 lacks a cation, even though the loop is still well defined in our electron density (Figure 5). Instead of a potassium ion, this structure contains two water molecules that coordinate the cation loop interactions. The first water molecule hydrogen bonds with the main chain carbonyl of Val 590, Gln 593, with the main chain amide of Asp 595, and with its neighboring water molecule. The second water coordinates the carbonyl of Asn 600 and of Tyr 707. The main chain carbonyl of Asp 595 now hydrogen bonds with the amine of Asn 600's side chain and its backbone carbonyl now points outside of the cation loop. In the presence of a potassium ion, the side chain of Asp 595 pairs with the amine of the Asn 600 side chain, however, in the absence of a potassium ion the side chain of Asp 595 points away from the cation loop and is no longer as well ordered in our density maps.

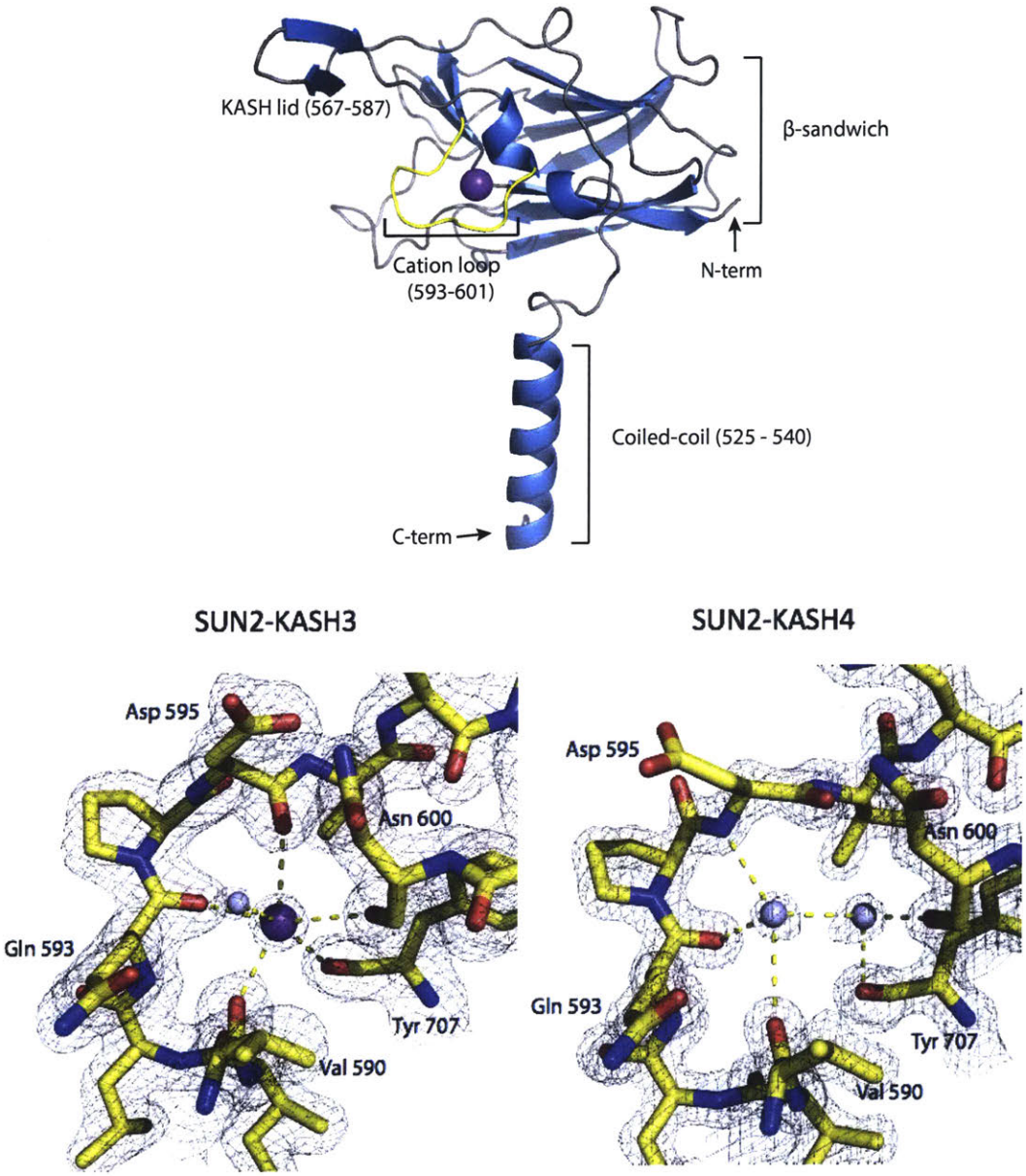


Figure 5. The cation loop of SUN2 can be coordinated in two different ways. Top, overview of SUN2 monomer structure highlighting the cation loop of SUN2. Bottom left, the cation loop of the SUN2-KASH3 complex is representative of what has been described to date. A potassium ion is coordinated by multiple backbone carbonyl interactions. Bottom right, in the structure of SUN2-KASH4 the ion loop is coordinated by two water molecules. Electron density in both cases is contoured at an RMSD of 2.0.

Mixed occupancy LINC complexes

After confirming that the core interactions between SUN2 and KASH1-5 are similar, we decided to test if a single SUN2 trimer could bind to two different KASH peptides simultaneously and if there is a preference for either binding mode (Figure 6). The same purification scheme shown in Figure 1 was used here, but an additional , His-tag-free GFP-fused KASH peptide was co-expressed alongside. Since we affinity-purify using one of the KASH peptides only, any additional KASH peptide that is co-purified must therefore be bound to the same SUN2 trimer. For example, we tested whether SUN2 could bind to KASH1 and 2 simultaneously. Here, we pulled on 6XHis-Gb1-KASH2 which co-purified both SUN2 and KASH1. This also held true for complex formation with KASH3 and KASH4, confirming that SUN2 can bind multiple KASH peptides at once that share the same binding mode. We are also able to purify SUN2 bound to both KASH2 and KASH3 simultaneously, showing that SUN2 can bind two KASH peptides that adopt different binding modes at once.

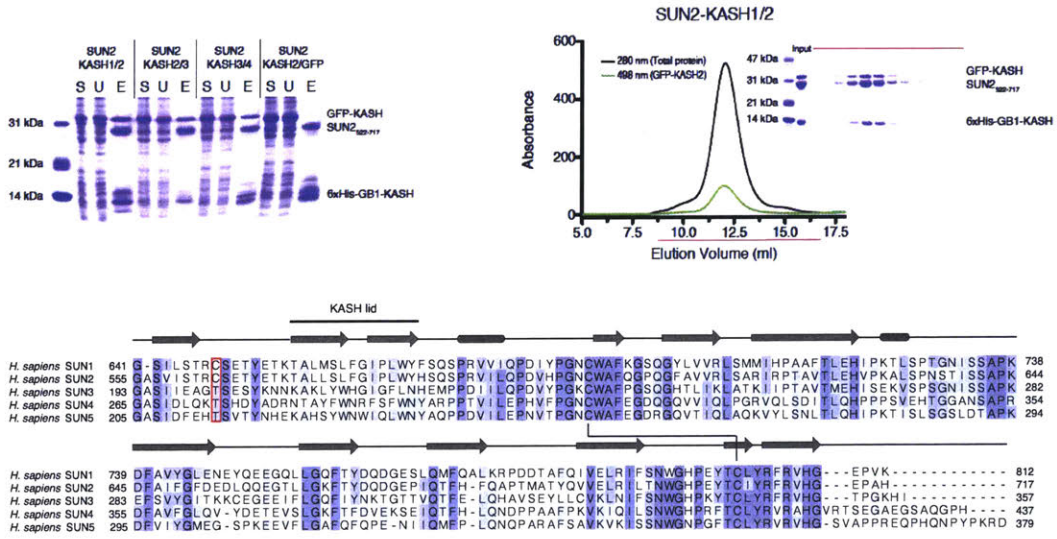


Figure 6. SUN2 can bind to multiple different KASH peptides simultaneously. Top left, purification of SUN2 in complex with two different KASH peptides. S shows the soluble fraction of the lysate, U shows the fraction that is unbound to nickel, E is the nickel elution. Top right, size exclusion chromatograph and the corresponding gel show that the complex is stable, black trace is total protein detected at 280nm and green trace shows GFP-KASH absorbance at 498nm. Bottom, multiple sequence alignment of *H. sapiens* SUN1-SUN5. Cys 563 is only conserved in SUN1 and SUN2.

Discussion

Here we present biochemical evidence that SUN2 can directly interact with all known human KASH peptides, forming at least 6 distinct LINC complexes. In addition, we determined the structures of SUN2 bound to KASH3, KASH4, and KASH5. We report binding affinities between SUN2 and KASH1-6, and we show evidence for heterogeneous LINC complexes in which a single SUN2 trimer can bind to multiple different KASH peptides simultaneously.

Previous work showed that SUN and KASH interactions occur with promiscuity, which raises the question of how these SUN KASH bridges can be specifically assembled and regulated. Lack of structural details of SUN1, SUN3, SUN4, and SUN5 leaves some gaps in our understanding of how SUN-KASH pairing occurs in these LINC complexes, although tissue specific expression of SUN3-5 is perhaps a way of specifically assembling LINC complexes involving these SUN proteins (Calvi, 2015; Pasch, 2015). SUN1 and SUN2, however, are expressed in all tissues and in various cases specific pairs of LINC complexes are known to be functionally relevant. For example, SUN1-KASH4 is important for high frequency hearing, while SUN1-KASH5 mediates chromosome pairing during meiosis, and SUN2-KASH2 mediates TAN line formation and nuclear polarization (Morimoto, 2012; Horn, 2013a; Horn, 2013b; Luxton, 2010). Because core interactions between SUN2 and all KASH peptides are conserved, they are unlikely candidates for regulating specificity. We propose that the two binding

modes we describe here are a mechanism of regulation between SUN2 and KASH. In our structures, KASH3-5 do not form disulfide bonds with SUN2, and thus their interaction is non-covalent and weaker than the interaction between SUN2 and KASH1/2 which form disulfide bonds (Sosa, 2012). Since KASH3-5 only non-covalently interact with SUN2 it is a distinct possibility that these LINC complexes cannot withstand the forces needed to anchor or move the nucleus (Jahed, 2015). Perhaps these "aberrant" LINC complexes can be specifically disassembled by the application of force from the cytoskeleton during mechanically taxing processes, and provide apo-SUN2 trimers to interact with either KASH1 or KASH2 that can effectively transmit mechanical force. However, SUN2 KASH3/4/5 complexes may still serve a role in sensing mechanical input and translating this signal into the nucleus, as opposed to moving the large organelle.

Based on early structures of SUN2 bound to KASH1 or KASH2, we speculated that each KASH binding site was most likely independent from the neighboring two. Here we take this a step further and provide the first piece of evidence for a heterogeneous LINC complex that pairs a single SUN2 trimer with 2 different KASH peptides. Because most KASH peptides are expressed in all tissues and the core interactions between SUN2 and KASH1-5 are essentially identical, the formation of LINC complexes with mixed occupancy is likely to occur in a cell. Once force is applied to these mixed LINC complexes it is likely

that non-covalently bound KASH peptides can be pulled apart from SUN2(Jahed. 2015).

In the context of SUN2, KASH peptides lacking a proline at position -11 are unable to meander along the surface of SUN2 towards the conserved Cys563 on SUN2. KASH3 and KASH4 both lack proline -11 but also posses a cysteine at position -23. Since SUN1 is able to couple to both of these peptides, and each of these LINC complexes have specific physiological roles, it is a distinct possibility that SUN1 can form a disulfide bond with KASH3 and KASH4. The covalent coupling between SUN1 and KASH3 and 4 could be regulated through the same mechanism we propose for SUN2 with KASH1 and 2. KASH5, in turn, lacks both proline -11 and cysteine -23, but colocalizes predominantly with SUN1. While KASH6 represents a unique case, since expression is enriched in testis, and KASH6 is also the weakest binder to SUN2 in all of our experiments (Lindeman, 2012). We hypothesize KASH6 is not a cognate partner of SUN2 nor SUN1, but instead one of the testis specific SUN3, SUN4, or SUN5. The testis specific SUN proteins lack a cysteine at position 563 which inhibits them from covalently binding to KASH peptides (Figure 6). Another function of non-cognate SUN and KASH coupling could be to maintain a pool of all KASH peptides at the NE. From this pool specific pairs can be disassembled by the application of force from the cytoplasm, which allows uncoupled KASH peptides to pair with SUN trimers in order to enrich for SUN-KASH pairs capable of forming disulfide bonds.

We have previously shown that the cation loop of SUN2 is needed for KASH binding. Here we report a structure of SUN2 in which the ion loop lacks an ion, but is instead coordinated by two water molecules. We only observe this in the context of SUN2-KASH4 complexes. While the cation loop is essential for KASH binding, it is unclear if different coordination provides yet another layer of regulation in LINC complex formation and maintenance.

Materials and Methods

Plasmid construction, protein expression and purification

DNA sequences containing human SUN2 were cloned into a modified bicistronic ampicillin resistant pETDuet-1 vector (EMD Millipore), superfolder GFP-KASH fusions (sfGFP) were cloned and expressed into a modified ampicillin resistant vector, pET-30b(+) (EMD Biosciences). For protein binding assays, a biotin acceptor peptide AAAGLNDIFEAQKIEWH was N-terminally fused to $SUN2_{500\text{-}717}$, preceded by a 6xHistidine tag. Each KASH peptide (KASH1-6) was C-terminally fused to 6xHistidine-sfGFP using inverse PCR. For crystallography, 6xHistidine tagged $SUN2_{522\text{-}717}$ was cloned into the first multiple cloning site (MCS) and MBP-fused $KASH4_{379\text{-}404}$ was cloned into the second MCS. For SUN2-$KASH3_{947\text{-}975}$ and SUN2-$KASH5_{542\text{-}562}$ the KASH peptides were C-terminally fused to 6xHistidine-Gb1 tags and $SUN2_{522\text{-}717}$ remained untagged (Cheng and Patel, 2001). All SUN2 and KASH constructs were N-terminally fused with a human rhinovirus 3C protease cleavage site.

Transformed LOBSTR(DE3)-RIL (Andersen, 2013) bacterial expression cells were grown at 37 °C to an OD_{600} of 0.6, then shifted to 18 °C and induced with 0.2 M isopropyl β-D-1-thiogalactopyranoside for 16 hours. Cells were harvested by centrifugation at 6000 g, resuspended in lysis buffer (50 mM potassium phosphate, pH 8.0, 400 mM NaCl, 40 mM imidazole) and lysed using an LM20 Microfluidizer Processor (Microfluidics). The lysate was cleared by

centrifugation at 10000 g for 25 minutes. The soluble fraction was incubated with Nickel Sepharose 6 Fast Flow beads (GE Healthcare) for 30 minutes at 4 °C in batch. After the beads were washed with lysis buffer, the protein was eluted (10 mM Tris/HCl pH 8.0, 150 mM NaCl, 250 mM imidazole). The protein was further purified by size exclusion chromatography using an S75 or S200 16/60 column (GE Healthcare) equilibrated in running buffer (10 mM Tris/HCl, pH 8.0, 150 mM NaCl, and 0.2 mM EDTA).

Protein affinity- and solubility-tags were removed with 3C protease and the protein complexes were separated from fusion tags by a second size exclusion step using an S75 16/60 column, under the same conditions.

In vitro binding experiments

Assays were performed at 30°C in a buffer containing 20 mM HEPES pH 8.0, 100 mM KCl, 1% BSA, and 0.05% TWEEN-20. BLI experiments were performed using an Octet RED96 (ForteBio) instrument. For all experiments Avi-tagged SUN2 was immobilized on streptavidin sensors, at a thickness of 1.5 nm. A baseline was collected in assay buffer, and for each binding step a control sensor with unbound sfGFP-KASH was used to subtract non-specific binding at the highest concentration of SUN2 corresponding to each experiment.

Kinetic data analysis

To obtain k_{on} and k_{off}, kinetic data was fit to single exponential association or decay functions using nonlinear-least-squares algorithms implemented in Prism (GraphPad Software). K_D was obtained by calculating the ratio of k_{on} and k_{off}.

Crystallization

Purified Apo-$SUN2_{500-717}$ was concentrated to 7 mg/ml and crystallized in 14.5% (w/v) polyethylene glycol (PEG) 3350, 100 mM potassium formate pH 8.0, and 200 mM sodium bromide. Rhombohedral crystals appeared after 5 days at 18°C. SUN2-KASH3 complex was concentrated to 5 mg/ml, and crystallized in 16-18% PEG3350, 100 mM ammonium citrate pH 7.0, 100 mM BisTris/HCl pH 6.5, and 10 mM nickel (II) chloride. Large, plate-shaped crystals grew within 2 days at 18°C. SUN2-KASH4 crystallized at 7 mg/ml in 17% PEG3350, 200 mM magnesium chloride, and 100 mM BisTris/HCl pH 6.5. Crystals appeared after 1 day and finished growing within 3 days, at 18°C. SUN2-KASH5 was crystallized at 10 mg/ml, in 14% PEG3000, and 100 mM BisTris/HCl pH 6.5. Crystals appeared and finished growing within 1-12 hours at 4°C. Crystals were cryoprotected in their reservoir solution supplemented with 15% glycerol and flash-frozen in liquid nitrogen. Diffraction data were collected at 24ID-C at Argonne National Laboratories.

X-ray data collection and structure determination

All data processing was done using programs provided by SBgrid (Morin, 2013). Data reduction was carried out with HKL2000 (Otwinowski and Minor, 1997), molecular replacement with PHASER from the phenix suite, using apo-SUN2 structure (PDB code 4DXT) that lacked the KASH lid as a search model. The structures were manually built using Coot and refined phenix.refine (Emsley, 2010). Data collection and refinement statistics are summarized in Table 1. Structure figures were created using PyMOL (Schrödinger LLC).

References

Andersen, K. R., Leksa, N. C., & Schwartz, T. U. (2013). Optimized E. coliexpression strain LOBSTR eliminates common contaminants from His-tag purification. Proteins: Structure, Function, and Bioinformatics, 81(11), 1857–1861. http://doi.org/10.1002/prot.24364

Attali, R., Warwar, N., Israel, A., Gurt, I., McNally, E., Puckelwartz, M., et al. (2009). Mutation of SYNE-1, encoding an essential component of the nuclear lamina, is responsible for autosomal recessive arthrogryposis. *Human Molecular Genetics*, *18*(18), 3462–3469. http://doi.org/10.1093/hmg/ddp290

Barateau, A., Vadrot, N., Vicart, P., Ferreiro, A., Mayer, M., Héron, D., et al. (2017). A Novel Lamin A Mutant Responsible for Congenital Muscular Dystrophy Causes Distinct Abnormalities of the Cell Nucleus. *PLoS ONE*, *12*(1), e0169189–18. http://doi.org/10.1371/journal.pone.0169189

Bridger, J. M., Foeger, N., Kill, I. R., & Herrmann, H. (2007). The nuclear lamina. *FEBS Journal*, *274*(6), 1354–1361. http://doi.org/10.1111/j.1742-4658.2007.05694.x

Calvi, A., Wong, A. S. W., Wright, G., Wong, E. S. M., Loo, T. H., Stewart, C. L., & Burke, B. (2015). SUN4 is essential for nuclear remodeling during mammalian spermiogenesis. *Developmental Biology*, *407*(2), 321–330.

http://doi.org/10.1016/j.ydbio.2015.09.010

Chang, W., Worman, H. J., & Gundersen, G. G. (2015). Accessorizing and anchoring the LINC complex for multifunctionality. *The Journal of Cell Biology*, *208*(1), 11–22. http://doi.org/10.1083/jcb.201409047

Ciani, B., Bjelic, S., Honnappa, S., Jawhari, H., Jaussi, R., Payapilly, A., et al. (2010). Molecular basis of coiled-coil oligomerization-state specificity. *PNAS*, 107(46), 19850-19855.

Cohen, M., Tzur, Y. B., Neufeld, E., Feinstein, M., Delannoy, M. R., Wilson, K. L., and Gruenbaum, Y. (2002). Transmission electron microscope studies of the nuclear envelope in Caenorhabditis elegans embryos. *Journal of Structural Biology*, 140, 232–240.

Crisp, M., Liu, Q., Roux, K., Rattner, J. B., Shanahan, C., Burke, B., Stahl, P. D., and Hodzic, D. (2006). Coupling of the nucleus and cytoplasm. *The Journal of Cell Biology*, *172*(1), 41–53. http://doi.org/10.1083/jcb.200509124

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Cryst (2010). D66, 486-501 [Doi:10.1107/S0907444910007493]*, 1–16.

http://doi.org/10.1107/S0907444910007493

Franke, W. W., Scheer, U., Khrone, G., and Jarasch, E. D. (1981). The Nuclear Envelope and the Architecture of the Nuclear Periphery, *The Journal of Cell Biology, 91(3),* 39–50.

Hammarström, M., Woestenenk, E. A., Hellgren, N., Härd, T., & Berglund, H. (2006). Effect of N-terminal solubility enhancing fusion proteins on yield of purified target protein. *Journal of Structural and Functional Genomics*, *7*(1), 1–14. http://doi.org/10.1007/s10969-005-9003-7

Horn, H. F., Brownstein, Z., Lenz, D. R., Shivatzki, S., Dror, A. A., Dagan-Rosenfeld, O., et al. (2013a). The LINC complex is essential for hearing. *Journal of Clinical Investigation*, *123*(2), 740–750. http://doi.org/10.1172/JCI66911

Horn, H. F., Kim, D. I., Wright, G. D., Wong, E. S. M., Stewart, C. L., Burke, B., & Roux, K. J. (2013b). A mammalian KASH domain protein coupling meiotic chromosomes to the cytoskeleton. *The Journal of Cell Biology*, *202*(7), 1023–1039. http://doi.org/10.1083/jcb.201304004

Jahed, Z., Shams, H., & Mofrad, M. R. K. (2015). A Disulfide Bond Is Required for the Transmission of Forces through SUN-KASH Complexes. *Biophysj*, *109*(3), 501–509. http://doi.org/10.1016/j.bpj.2015.06.057

Jayo, A., Malboubi, M., Antoku, S., Chang, W., Ortiz-Zapater, E., Groen, C., et al. (2016). Fascin Regulates Nuclear Movement and Deformation in Migrating Cells. *Developmental Cell*, *38*(4), 371–383. http://doi.org/10.1016/j.devcel.2016.07.021

Kabachinski, G., & Schwartz, T. U. (2015). The nuclear pore complex - structure and function at a glance. *Journal of Cell Science*, *128*(3), 423–429. http://doi.org/10.1242/jcs.083246

Knockenhauer, K. E., & Schwartz, T. U. (2016). The Nuclear Pore Complex as a Flexible and Dynamic Gate. *Cell*, *164*(6), 1162–1171. http://doi.org/10.1016/j.cell.2016.01.034

Kutscheidt, S., Zhu, R., Antoku, S., Luxton, G. W. G., Stagljar, I., Fackler, O. T., & Gundersen, G. G. (2014). FHOD1 interaction with nesprin-2G mediates TAN line formation and nuclear movement. *Nature Cell Biology*, *16*(7), 708–715. http://doi.org/10.1038/ncb2981

Lindeman, R. E., & Pelegri, F. (2012). Localized Products of futile cycle/ lrmp Promote Centrosome-Nucleus Attachment in the Zebrafish Zygote. *Current biology*, *22*(10), 843–851. http://doi.org/10.1016/j.cub.2012.03.058

Malone, C. J., Fixsen, W. D., Horvitz, H. R., Han, M. (1999). UNC-84 localizes to the nuclear envelope and is required for nuclear migration and anchoring during C. elegans development. *Development*, 126, 3171-3181.

Morgan, J. T., Pfeiffer, E. R., Thirkill, T. L., Kumar, P., Peng, G., Fridofsson, H. N., Doublas, G. C., Starr, D. A., Barakat, A. I. (2011). Nesprin-3 regulates endothelial cell morphology, perinuclear cytoskeletal architecture, and flow-induced polarization. *Molecular Biology of the Cell.* 4324–4334. http://doi.org/10.1091/mbc.E11-04-0287)

Morimoto, A., Shibuya, H., Zhu, X., Kim, J., Ishiguro, K.-I., Han, M., & Watanabe, Y. (2012). A conserved KASH domain protein associates with telomeres, SUN1, and dynactin during mammalian meiosis. *The Journal of Cell Biology*, *198*(2), 165–172. http://doi.org/10.1083/jcb.201204085

Morin, A., Eisenbraun, B., Key, J., Sanschagrin, P. C., Timony, M. A., Ottaviano, M., & Sliz, P. (2013). Collaboration gets the most out of software. *eLife*, *2*, 213–6. http://doi.org/10.7554/eLife.01456

Otwinoski, Z., & Minor W. (1997). Processing of X-ray Diffraction Data in Oscillation mode. Methods in Enzymology, 276, 307–326.

Pasch, E., Link, J., Beck, C., Scheuerle, S., & Alsheimer, M. (2015). The LINC complex component Sun4 plays a crucial role in sperm head formation and fertility. *Biology Open*, *4*(12), 1792–1802. http://doi.org/10.1242/bio.015768

Rajgor, D., Hanley, J. G., & Shanahan, C. M. (2016). Identification of novel nesprin-1 binding partners and cytoplasmic matrin-3 in processing bodies. *Molecular Biology of the Cell.* 3894-3902. http://doi.org/10.1091/mbc.E16-06-0346)

Razafsky, D., & Hodzic, D. (2009). Bringing KASH under the SUN: the many faces of nucleo-cytoskeletal connections. *The Journal of Cell Biology*, *186*(4), 461–472. http://doi.org/10.1083/jcb.200906068

Starr, D. A., & Fischer, J. A. (2005). KASH 'n Karry: The KASH domain family of cargo-specific cytoskeletal adaptor proteins. *BioEssays*, *27*(11), 1136–1146. http://doi.org/10.1002/bies.20312

Starr, D. A. (2012). Laminopathies: Too Much SUN Is a Bad Thing. *Current Biology*, *22*(17), R678–R680. http://doi.org/10.1016/j.cub.2012.06.070

Starr, D. A., & Fridolfsson, H. N. (2010). Interactions Between Nuclei and the Cytoskeleton Are Mediated by SUN-KASH Nuclear-Envelope Bridges. *Annual Review of Cell and Developmental Biology*, *26*(1), 421–444. http://doi.org/10.1146/annurev-cellbio-100109-104037

Turgay, Y., Eibauer, M., Goldman, A. E., Shimi, T., Khayat, M., Ben-Harush, K., Dubrovsky-Gapp, A., Sapra, T., Godlman R. D., & Medalia, O. (2017). The molecular architecture of lamins in somatic cells. *Nature Publishing Group*, *543*(7644), 261–264. http://doi.org/10.1038/nature21382

Ungricht, R., Klann, M., Horvath, P., & Kutay, U. (2015). Diffusion and retention are major determinants of protein targeting to the inner nuclear membrane. *The Journal of Cell Biology*, *209*(5), 687–704. http://doi.org/10.1083/jcb.201409127

Wang, Q., Du, X., Cai, Z., and Greene, M. I. (2002). Characterization of the Structures Involved in Localization of the SUN Proteins to the Nuclear Envelope and the Centrosome. *DNA and Cell Biology*, 25, 554-562.

Wang, W., Shi, Z., Jiao, S., Chen, C., Wang, H., Liu, G., et al. (2012). Structural insights into SUN-KASH complexes across the nuclear envelope. *Cell Research*, *22*(10), 1440–1452. http://doi.org/10.1038/cr.2012.126

Zhang, Q., Bethmann, C., Worth, N. F., Davies, J. D., Wasner, C., Feuer, A., et al. (2007). Nesprin-1 and -2 are involved in the pathogenesis of Emery Dreifuss muscular dystrophy and are critical for nuclear envelope integrity. *Human Molecular Genetics*, *16*(23), 2816–2833. http://doi.org/10.1093/hmg/ddm238

Zhang, Q., Ragnauth, C., Greener, M. J., Shanahan, C. M., & Roberts, R. G. (2002). The Nesprins Are Giant Actin-Binding Proteins, Orthologous to *Drosophila melanogaster* Muscle Protein MSP-300. *Genomics*, *80*(5), 473–481. http://doi.org/10.1006/geno.2002.6859

Zhou, Z., Du, X., Cai, Z., Song, X., Zhang, H., Mizuno, T., et al. (2012). Structure of Sad1-UNC84 Homology (SUN) Domain Defines Features of Molecular Bridge in Nuclear Envelope. *Journal of Biological Chemistry*, *287*(8), 5317–5326. http://doi.org/10.1074/jbc.M111.304543

Chapter 2: Coiled-coil Domain of SUN2 and its Role in the Oligomerization of LINC Complexes

Summary

Mechanical communication between the nucleus and cytoskeleton depends on protein complexes that span the lumen of the perinuclear space (PNS) physically connecting the outer and inner nuclear membranes (ONM and INM, respectively). The INM SUN proteins together with ONM KASH proteins fulfill this requirement. SUN proteins possess a transmembrane helix that anchors them at the INM, followed by a large trimeric coiled-coil domain that spans the ~50 nm gap between the INM and the ONM. The C terminus of SUN proteins contains the conserved SUN domain that binds to and anchors the ONM KASH proteins. The coiled-coil domain of SUN2 is required for trimerization of SUN proteins, and trimer formation is a pre-requisite for KASH binding. Thus, the coiled-coil domain of SUN serves at least two distinct functions. First to regulate the oligomeric state of SUN, enabling KASH binding, and second to span the gap of the PNS thus placing the SUN domain proximal to the KASH containing ONM.

Here we present our preliminary analysis of the coiled-coil domain of SUN2. The coiled-coil domain requires the INM proximal element for trimerization of apo-SUN2. In addition, KASH stabilizes the SUN2 trimer through its interaction with the SUN domain. Without the SUN domain, coiled-coils expressed lacking the trimerization domain form soluble aggregates, possibly by formin off register oligomers.

We have designed various coiled-coil constructs for structural analysis. We have been able to crystallize a large portion of the coiled-coil domain of SUN2. Despite good diffraction data quality, solving the phase problem remains a work in progress.

Introduction

The linker of nucleoskeleton to cytoskeleton (LINC) complex physically tethers the nucleus to the cytoskeleton. In order to achieve this function LINC complexes must span the perinuclear space (PNS) and traverse both the outer nuclear membrane (ONM) and the inner nuclear membrane (INM). These criteria are met by the two components of LINC complexes, the ONM protein KASH, and the INM protein SUN (Starr 2010). KASH proteins are C-terminally anchored to the ONM and are retained by their interactions with the SUN domain of SUN proteins. KASH proteins only project their final ~30 C-terminal residues into the PNS, known as the KASH domain. SUN proteins also project their C-terminal domain into the PNS, however their perinuclear domain spans the ~50 nm separating the ONM from the INM (Sosa, 2013; Chang, 2015) (Figure 1).

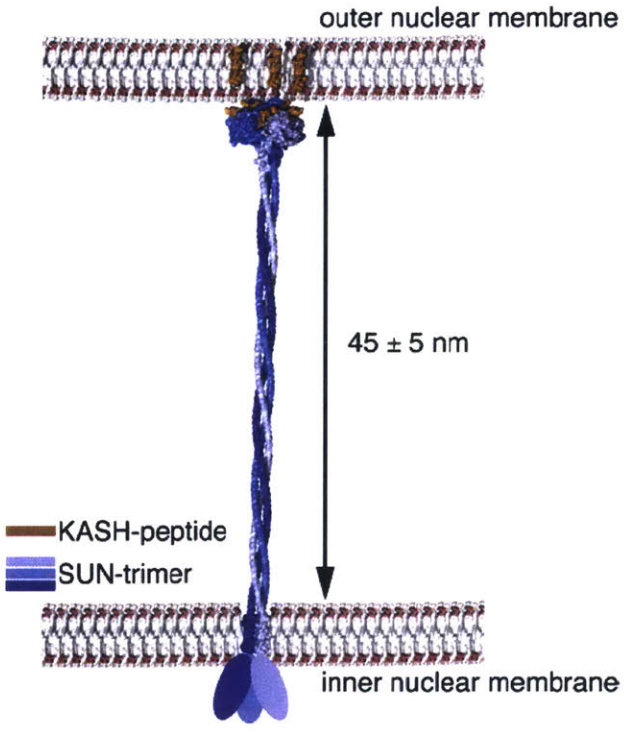


Figure 1. Model of SUN2 trimer. SUN2 is shown to scale with the PNS, the coiled-coil is modeled as an elongated trimer with a consistent pitch. *Adapted from Sosa, 2012.*

The very C terminus of SUN proteins contain the eponymous SUN domain that is preceded by an elongated coiled-coil domain. The coiled-coil domain of SUN proteins may act as a ruler that determines the spacing between the ONM and the INM (Sosa, 2013). This hypothesis is supported by the presence of SUN proteins with shorter coiled-coil domains in areas of higher membrane curvature, predominantly in spermatocytes (Sosa, 2012; Lindeman, 2012) (Figure 2). Shorter SUN proteins could stabilize this unfavorable membrane architecture through their interactions with KASH proteins. There is, however, some data that challenges this model. In *C. elegans*, shortening the length of the coiled-coil domain of UNC-84 (SUN2 homolog) does not shorten the spacing between the ONM and the INM despite correct localization of SUN and KASH (Cain, 2014). However, cells under mechanical stress were not able to retain appropriate spacing between the nuclear membranes, suggesting that the coiled-coil domain of UNC-84 may somehow contribute to LINC complex stability and not just bridge the gap between nuclear membranes.

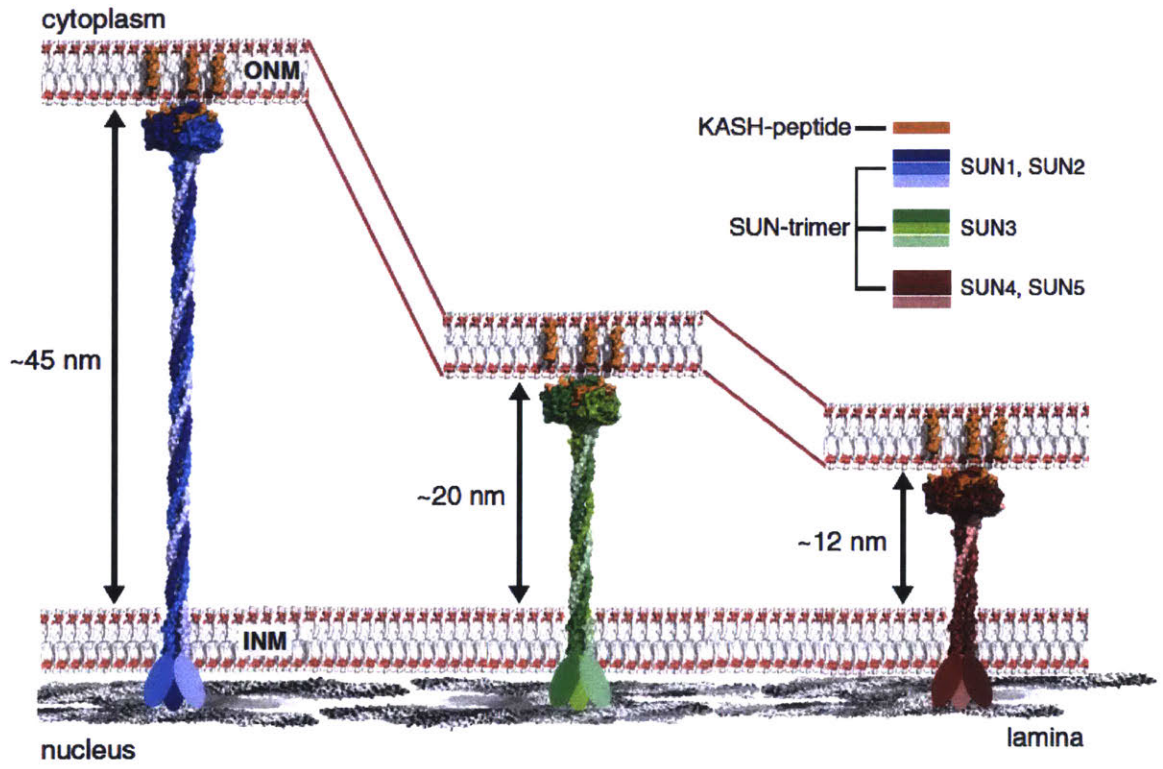


Figure 2. Perinuclear spacing may be regulated by SUN trimers of various lengths. If the PNS domain of SUN3 and SUN4/5 are modeled as trimeric coiled-coils, they would substantially reduce the spacing between the INM and ONM. *Adapted from Sosa, 2013.*

LINC complexes form molecular bridges that facilitate nuclear migration. This process requires the transmission of large forces from the cytoskeleton to the nucleus (Luxton, 2010). It is likely that arrays of LINC complexes are required to withstand such large forces. It has been proposed that the coiled-coil domain of SUN2 can mediate the clustering of LINC complexes, either through lateral interactions between neighboring coiled-coils, or through a process where a single helix from one coiled-coil can unwind and domain swap with a neighboring LINC complex (Zhou, 2012).

Early structural characterization of SUN2 revealed that SUN2 folds as a trimer, and that trimerization of SUN2 is a pre-requisite for KASH binding (Sosa, 2012; Zhou, 2012; Wang, 2012). Additionally, the structures revealed that N-terminal to the SUN domain is a trimeric right handed coiled-coil element with undecan repeat that is important for trimerization. It is likely that the entire perinuclear domain preceding the SUN domain folds as a trimeric coiled-coil. Recent structural work on various fragments of the coiled-coil domain of murine SUN2 has revealed interesting features (Nie, 2016). First, the handedness of the coiled-coil switches along the length of the coiled-coil. Second, various polar side chains are directed towards the core of the trimer, which is uncommon since this is typically dominated by hydrophobic interactions. Third, the coiled-coil proximal to the SUN domain was shown to interact with the SUN domain itself, stabilizing a monomer of SUN that is incapable of binding to KASH. It is still unclear if monomers of SUN2 are present in physiological conditions, and their role in regulating LINC complex assembly remains speculative.

Here we present our preliminary our work on the coiled-coil domain of human SUN2. We describe the purification of various constructs that yield large quantities of pure protein. We also present evidence for how the coiled-coil contributes to SUN2 trimerization, and we discuss our structural work on the coiled-coil domain of SUN2. Structures of elongated coiled-coils are rare, especially of trimeric coiled-coils. Hence, we believe we have a unique opportunity to structurally characterize the coiled-coil

domain of SUN2 and enrich our understanding of trimeric coiled-coil structure, while also addressing various unresolved questions in the field of LINC complexes.

Results

Purification of the coiled-coil domain of SUN2

We first generated constructs of $SUN2_{(273-717)}$ that spanned the entire length of the coiled-coil domain and included the SUN domain. All SUN2 constructs described are N-terminally tagged with a 6x histidine affinity tag, followed by a 3C protease cleavage site. The purification of all our constructs follows the same general protocol. First we separate SUN2 from lysate through nickel affinity chromatography. The eluted protein is then separated from co-purified nucleic acids by size exclusion chromatography. Nucleic acids inhibit complete proteolytic cleavage of the histidine tag. The main peak from the size exclusion run is collected and the 6x histidine tag is removed using a 1:200 ratio of rhinovirus 3C protease. The protein complex is then separated from the solubility tag by a second size exclusion step. This yields pure, homogeneous SUN2.

In some cases, we co-expressed $SUN2_{(273-717)}$ as well as various shorter SUN2 constructs with MBP fused KASH2. For the shorter constructs, the complex can be separated from the apo form through size exclusion. For the larger SUN2 constructs, the hydrodynamic radius is not much larger in the presence or absence of MBP-KASH and size exclusion chromatography is largely ineffective at separating KASH bound from apo-SUN2. The best way to purify these is with an orthogonal amylose pull-down that separates SUN2 MBP-KASH complexes from apo-SUN2.

The SUN domain of SUN2 is not required for trimerization

We have been able to purify large amounts of different SUN2 construct that possess various elements of the coiled-coil domain. Despite considerable effort, these constructs resisted crystallization. During our analysis we noticed that a certain length of the coiled-coil is required for trimerization of SUN2. Constructs that contain the SUN domain, but lack the segment of the coiled-coil spanning residues 322-369 consistently form monomers during gel filtration. Our next strategy was to remove the SUN domain and to focus on purifying the coiled-coil domain of SUN2. Constructs of SUN2 lacking the SUN domain were well behaved in solution, and also yielded large amounts of protein (Figure 3). Interestingly, these constructs can be substantially shortened from the C terminus without affecting their oligomerization or stability. However, this trend is challenged when the N terminus is shortened. Constructs lacking the residue range 322-369 form large soluble aggregates that elute in the void volume during size exclusion chromatography, potentially by forming off-register oligomers (Figure 3). Since this residue range seems to be critical for trimer formation, we have dubbed it the trimerization domain. Because of its important function, our structural characterization of SUN2 is exclusively focused on constructs that include the trimerization domain.

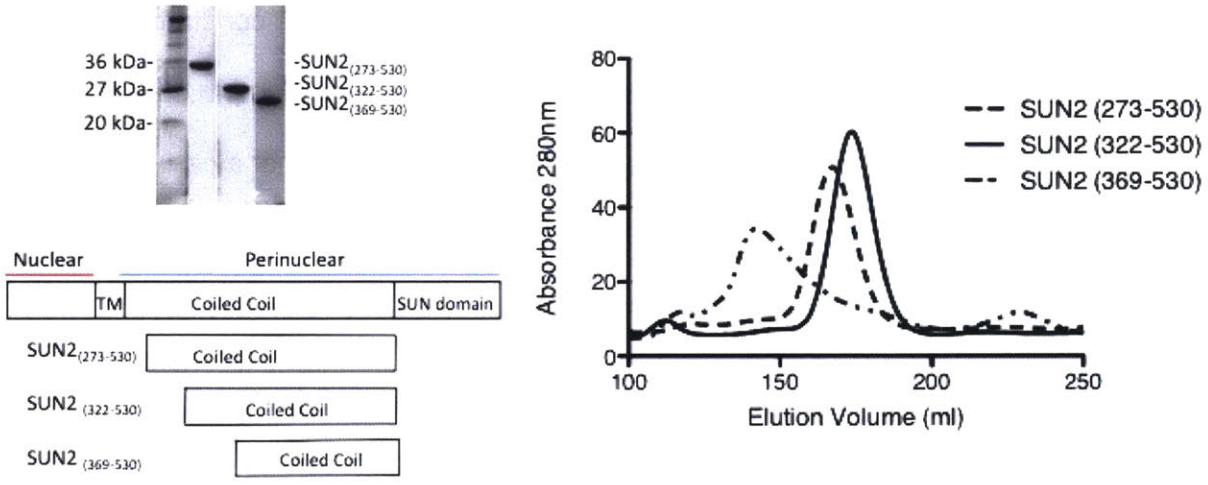


Figure 3. Identification of coiled-coil region critical for trimerization of SUN2. Left, SDS-PAGE analysis of three purified SUN2 coiled-coil fragments. The gel was stained with Coomassie Brilliant Blue. Schematic diagram of the domain architecture of human SUN2 and of the constructs used in this experiment. Right, size exclusion profiles of the three coiled-coil constructs.

SUN2 nanobodies

The variable domain of single-chain antibodies specific to camelids are called nanobodies. Nanobodies are small, can bind their target proteins with high affinity, and they can be produced recombinantly in high yield (Fridy, 2014). In numerous examples they have been shown to promote crystallization, particularly of conformationally flexible proteins. (Demircioglu, 2016; Pardon, 2014). To further our characterization of SUN2 we have generated nanobodies from alpacas. Nanobodies are generated towards the protein that is injected in the animal, or fragments of the protein if it is degraded in serum. Because of this we decided to inject the full perinuclear domain of $SUN2_{(273-717)}$. We must next obtain high affinity binders using phage display technology coupled with ELISA. For this we employed the use of two constructs, first the full trimeric coiled-coil

domain of $SUN2_{(273-530)}$ in order to obtain nanobodies that bind to the coiled-coil domain. The second construct was the exclusively monomeric $SUN2_{(369-717)}$ that includes the SUN domain in order to purify nanobodies that bind to SUN2 monomers. Using both constructs we were able to obtain nanobodies that form gel-filtration stable complexes with SUN2.

All nanobodies were expressed as C-terminal bdSUMO fusions (Frey, 2014). The nanobodies generated for the coiled-coil are insoluble upon cleavage of the solubility tag, we thus tested their binding by mixing uncut bdSUMO-nanobody to $SUN2_{(273-717)}$ or with $SUN2_{(273-530)}$ lacking the SUN domain. We mixed these at a 1:1 ratio and only then added the bdSUMO protease SENp, in this context the nanobody remains in solution. Employing size exclusion chromatography, we can readily separate $SUN2_{(273-717)}$ in complex with the nanobody from the bdSUMO solubility tag. This complex remains as a trimer, even upon nanobody binding. Next, we tested the binding of our other nanobodies to $SUN2_{(369-717)}$. Using the aforementioned purification scheme we have also been able to identify a nanobody that binds to the $SUN2_{(369-717)}$ monomer (Figure 4). With size exclusion chromatography of the complex we confirm that the nanobody which binds to SUN2 trimers cannot bind to SUN2 monomers, and vice versa. We have thus been able to generate specific binders to two different oligomeric states of SUN2.

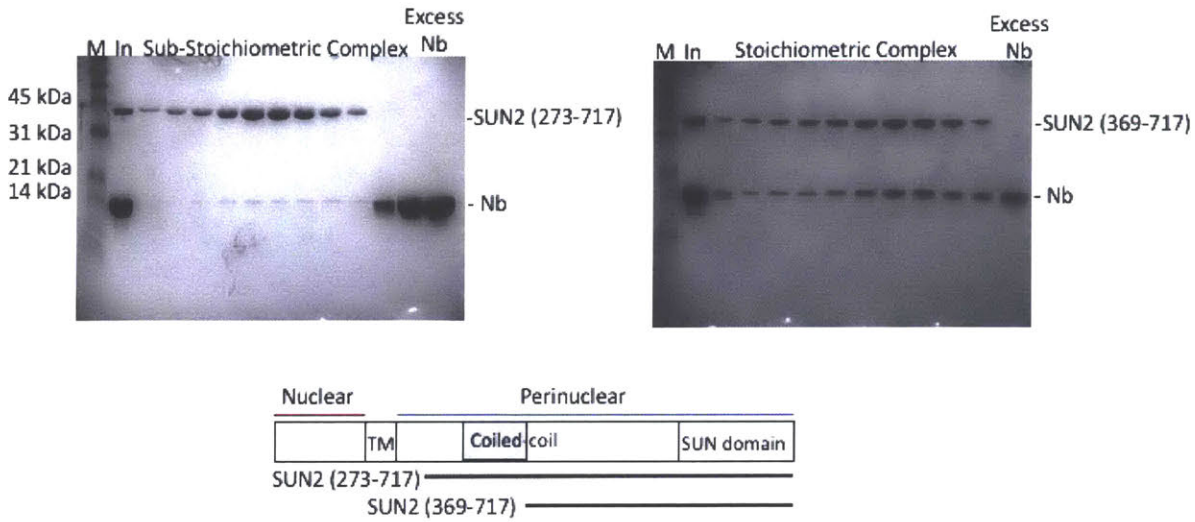


Figure 4. Nanobody SUN2 complex. Both gels correspond to fractions from gel filtration experiment. In both cases, nanobody is mixed with SUN2 at a 1 to 1.5 molar excess of nanobody. Left, nanobody does not bind to SUN2 trimers. Right, nanobody forms a stable complex with SUN2 monomers. Bottom, schematic diagram of SUN2 domain architecture.

Crystallization of SUN2 coiled-coil

We were able to obtained initial crystals of the coiled-coil domain of $SUN2_{(322-455)}$ which appeared after 3 weeks as clusters of needles (Figure 5). Optimization of these crystals proved challenging. The crystals can be reproduced and their size scaled in a 24 well hanging drop format, crystal nucleation was extremely slow. Despite extensive efforts to accelerate nucleation, crystals would appear only after about 4 months and would grow over a 1-week period in a 24-well format. Regardless of this difficulty, we have optimized these coiled-coil crystals to single large rods that reach a final size of about 200 μm x 30 μm x 30 μm. The crystallization condition contained a high percentage of MPD (50%) that conveniently also serves as a cryo-protectant.

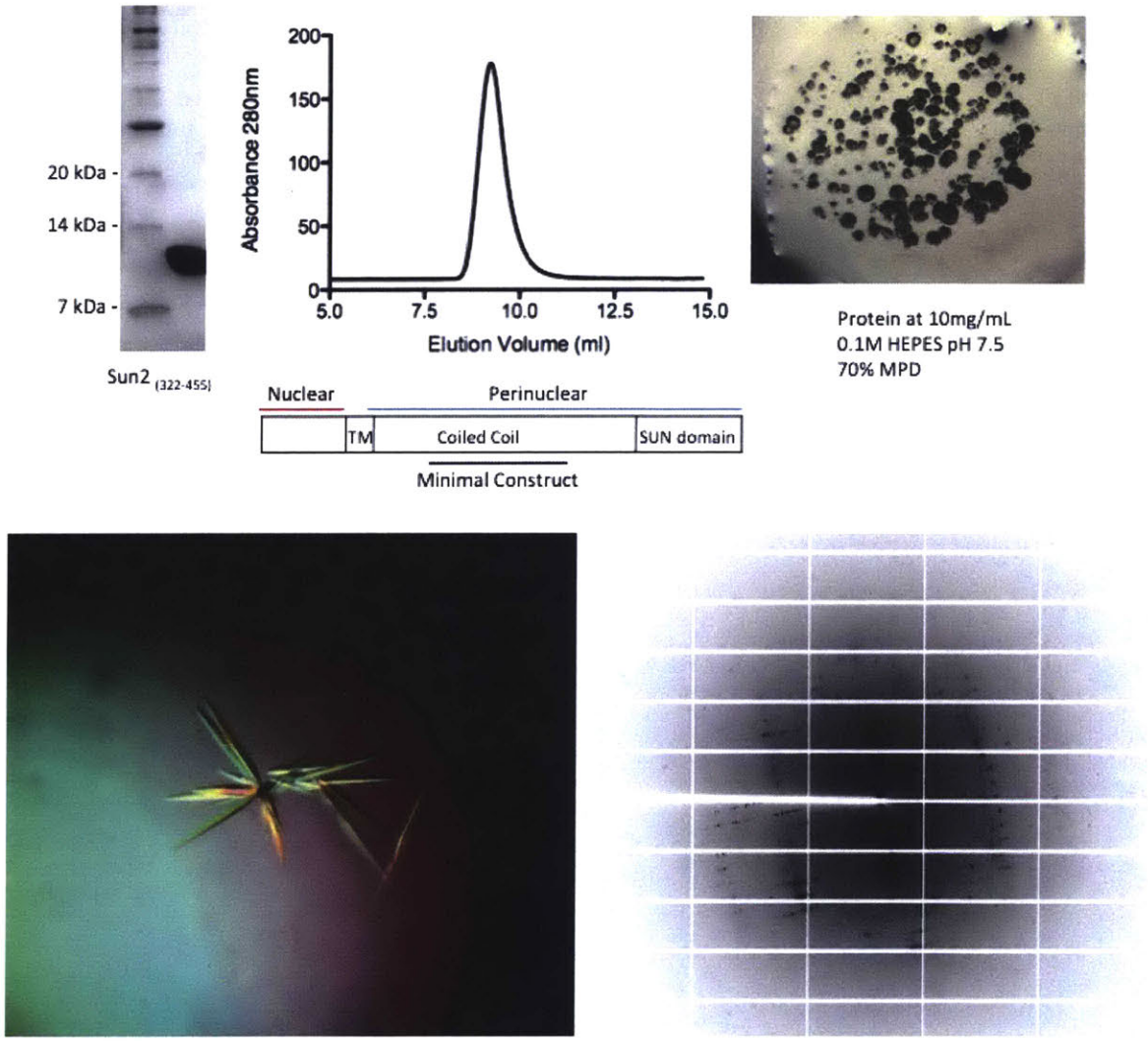


Figure 5. Purification, crystallization, and diffraction of SUN2 coiled-coil domain. Top left, purification of the SUN2 coiled-coil trimerization domain and size exclusion chromatograph. Schematic highlights the fragment of the SUN2 coiled-coil we have purified. Top right, initial crystals of SUN2 coiled-coil. Bottom left, optimized crystals of SUN2 coiled-coil, picture is taken 5 months after the crystallization experiment is set up. Bottom right, diffraction patter of tantalum soaked SUN2 coiled-coil crystals.

Data collection and structure solution

Crystals of the SUN2 coiled-coil are rather elongated and diffract well throughout their length, which allowed us to distribute radiation dosage over a large area of the crystal. Diffraction data was collected using the helical collection method over a 360° rotation that spanned the length of the crystal. Spots corresponding to one axis of the lattice were quite close together, to avoid spot overlap we collected small wedges of only 0.2° per frame (Figure 5). This is an indication that one crystal axis is quite large. Crystals withstood radiation without showing signs of decay as diagnosed by R_{merge} vs frame.

We also collected multiple data sets with various anomalous scatterers in order to solve the structure using the single wavelength anomalous diffraction (SAD) method (Table 1). All anomalous data sets were collected as described above. Additionally, we collected data on various selenomethionine derivatized crystals of high quality.

Data reduction was carried out using the HKL2000 suite. When indexed as an R3 space group, two axis (x and y) were consistently about 50 Å in length, while the third axis (z) is in the range of 600 Å. Data quality for both native and derivatized crystals is good, and resolution usually extended to the range of 2.1 – 2.7 Å (Table 1).

Data collection and refinement statistics

Protein	SUN2 322-455	SUN2 322-455 Tantalum	SUN2 322-455 SeMet	SUN2 (L414M) 322-455 SeMet
Organism	*H. sapiens*	*H. sapiens*	*H. sapiens*	*H. sapiens*
Data collection				
Space group	R 3	R 3	R 3	R 3
a, b, c (Å)	44.8, 44.8, 589.1	42.5, 42.5, 590.0	44.8, 44.8, 588.0	44.9, 44.9, 594.2
α, β, γ (°)	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 90.0, 120.0
Wavelength (Å)	0.9792	1.2543	0.9791	0.9792
Completeness (%)	100.0 (98.0)	100.0 (99.8)	100.0 (99.9)	100.0 (99.8)
Redundancy	15.5 (9.1)	2.1 (1.9)	8.4 (4.7)	4.4 (2.9)
$R_{p.i.m.}$ (%)	1.6 (48.4)	5.3 (20.4)	4.1 (21.2)	4.2 (78.0)
I/σ	69.6 (0.9)	25.8 (3.8)	15.79 (0.9)	31.0 (0.5)
CC1/2 (%)	99.9 (65.6)	100 (90.4)	99.9 (89.3)	99.9 (73.0)
Resolution Range (Å)	99.0 – 2.2	99.0 – 2.7	99.0 – 2.6	99.0 – 2.1

Table 1. Data collection and refinement statistics. The highest resolution shell is in parenthesis. CC1/2 was used as the criteria for resolution cutoff.

There are likely 3 or 4 copies of the coiled-coil present in the asymmetric unit, judged by the solvent content. Since there is an overlap between our crystallized SUN2 coiled-coil and the recently solved murine SUN2 coiled-coil we used this as a search model (PDB code 5ED9). While a solution is found for the search model, it is not well enveloped in electron density and further refinement did not yield electron density for the remainder of the protein. Further, we generated a model for the entire coiled-coil based on homology modeling (hhpred) and performed MR search but this did not yield a reasonable solution.

Since MR did not produce a reasonable solution, we next attempted to solve the structure using experimental phases generated from the selenomethionine dataset and from the multiple heavy atom soaks we employed. Using the SHELX program suite, we

searched for heavy atom sites using SAD. Using the tantalum soaked crystals, we are able to find 3 heavy atom sites. The sub-structure solution places the heavy atoms in special sites, positioned along the Z axis in the 3-fold symmetry axis of the unit cell. Electron density is clearly present at these sites up to an RMSD of 4. However, sub-structure refinement does not yield interpretable density, and generates R_{free} and R_{work} values in the order 45-50% after refinement. The same solution is found when performing sub-structure searches for the platinum derivatized dataset, but no interpretable electron density is generated after placing the heavy atoms (Figure 6). The selenomethionine derivatized crystals also generate similar electron density, and using molecular replacement yields a similar electron density map to what is observed in Figure 6.

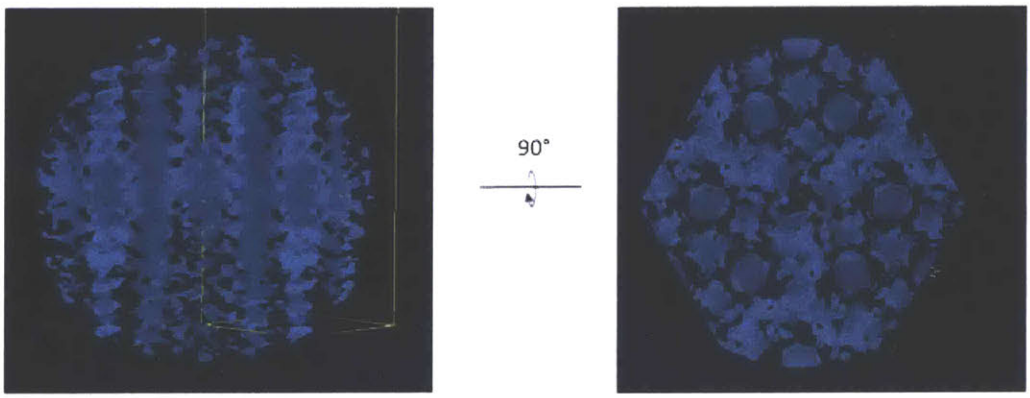


Figure 6. Electron density after tantalum cluster placement. Left panel shows tubular electron density extending along the z-axis, reminiscent of an elongated coiled-coil trimer. Right panel looks down the z-axis. Tantalum clusters are placed in the center of the observed density on the 3-fold symmetry axis.

Discussion

It is well established that SUN2 folds as a trimer, as evidenced by multiple crystal structures, ultracentrifugation, and SEC MALS analysis (Sosa, 2012; Nie et al., 2016; Zhou et al., 2012). It is interesting that SUN2 constructs lacking a fragment of the coiled-coil fail to trimerizes, and instead behave as monomers in solution. Since KASH peptides bind at the interface between SUN monomers they probably stabilize trimers containing the SUN domain. In the absence of KASH, however, SUN2 trimerization probably depends on interactions within the coiled-coil domain. It is still unclear if monomers of SUN2 exist in the PNS, since this behavior is only observed in constructs lacking a considerable fragment on the coiled-coil that includes the trimerization domain. We hypothesize that the coiled-coil domain mediates early trimerization of SUN2, which brings adjacent SUN domains into close proximity and poises them for KASH binding. Binding of SUN to KASH then "zips up" the SUN domain and generates fully trimerized SUN2. In this model, interactions between the SUN domain of SUN2 and the adjacent coiled-coil are either transient or only present in recombinant constructs lacking the full coiled-coil domain that do not undergo an early trimerization event mediated by the coiled-coil. Further characterization of the coiled-coil domain of SUN2 is required to properly address these questions.

The structural analysis of the coiled-coil element of SUN2 proved to be remarkably challenging. Neither with MR nor SAD phasing we were able to solve the structure yet. It is unclear to us what causes this, since in both cases the MR or sub-

structure solutions can be placed with high confidence and the crystals do not seem to have any serious pathologies. We speculate that the unusual unit cell and composition of the trimeric coiled-coil, which has an inherent 3-fold symmetry axis, may contribute to this issue.

From the preliminary electron density maps we obtained it is clear that the coiled-coil trimers are organized as parallel trimers that extend along the z-axis. The data is not clear enough to see whether parallel running trimers may exchange strands, which could explain why refinement proves to be difficult. Since MR using homologous structures do not yield solutions it is likely that the structure of the SUN2 coiled-coil contains features that cannot be modeled properly based on sequence homology to other coiled-coils. The coiled-coil may switch handedness throughout its length, or perhaps engage in lateral interactions or partial domain swaps in our crystal structure which could have far reaching consequences in our understanding of LINC complex oligomerization.

We are currently taking a two-pronged approach to further our efforts of structurally characterizing the coiled-coil of SUN2. First, we are exhausting all avenues in the data processing, MR, and phasing of our current crystals in order to obtain interpretable electron density. This approach focuses on finding a computational solution to this problem. Second, we will try to crystallize nanobody bound fragments of

the coiled-coil. We can then work back from these structure as a starting model in order to phase our current crystals and identify unique features present in these crystals.

Materials and Methods

Plasmids and protein expression

Recombinant proteins were all expressed in *E. coli*. All SUN2 constructs were expressed from a modified ampicillin resistant pET-DUET1 (Novagen) vector. SUN2 is fused to an N-terminal 6x histidine tag. All SUN2 truncations were generating using inverse PCR. Nanobodies used here are expressed as C-terminal fusions to a 14x-histidine-bdSUMO tag that is proteolytically cleaved using the bdSUMO protease SENp at a 1 to 1000 ratio for an hour at 4°C.

Transformed LOBSTR(DE3)-RIL (Andersen, 2013) bacterial expression cells were grown at 37 °C to an OD600 of 0.6, then shifted to 18 °C and induced with 0.2 M isopropyl β-D-1-thiogalactopyranoside for 16 hours. Cells were harvested by centrifugation at 6000 g, resuspended in lysis buffer (50 mM potassium phosphate, pH 8.0, 400 mM NaCl, 40 mM imidazole) and lysed using an LM20 Microfluidizer Processor (Microfluidics). The lysate was cleared by centrifugation at 10000 g for 25 minutes. The soluble fraction was incubated with Nickel Sepharose 6 Fast Flow beads (GE Healthcare) for 30 minutes at 4 °C in batch. Beads are then washed 3 times using 5x the nickel bed volume. SUN2 is then eluted using 5x the nickel bed volume of elution buffer (10 mM Tris/HCl pH 8.0, 150 mM NaCl, 250 mM imidazole). SUN2 nickel elution is concentrated to a volume of 5 ml and is purified on a size exclusion chromatography column on an S200 1660 (GE healthcare) pre-equilibrated in running buffer (10 mM Tris/HCl, pH 8.0, 150 mM NaCl, and 0.2 mM EDTA, 1 mM DTT). The main peak is

collected and the solubility tag is cut overnight with rhinovirus 3C protease at a 1 to 200 molar ratio. Cleavage of the affinity tags is confirmed by SDS PAGE stained by Coomassie Brilliant Blue.

Nanobodies are tagged with 14-histidine-bdSUMO and overexpressed and purified as described for SUN2 except that after the first gel filtration step the protein is flash frozen in liquid nitrogen and stored at -80 °C until complex formation.

Purification of SUN2-nanobody complexes

SUN2 and the nanobodies are purified separately as described above. Corresponding SUN2 constructs are mixed at a 1:1 molar ratio with nanobody that has not been treated with SENp in order to form a complex. The bdSUMO protease SENp is then added at a 1:1000 molar ratio of protease to nanobody, and is placed at 4 °C for 1 hour. The complex is then purified on an Superdex 200 10/300 (GE Healthcare) that has been pre-equilibrated in running buffer. Stoichiometric complex formation was verified by SDS-polyacrylamide that was stained with Coomassie Brilliant Blue. Complex was then concentrated to 10 mg/ml for further experiments, the remaining protein complex is flash frozen in liquid nitrogen and stored at -80 °C.

Crystallization and data processing of the SUN2 coiled-coil domain

Crystals of the coiled-coil domain of SUN2 were first obtained using sparse matrix screens in 0.1M HEPES pH 7.5, and 70% MPD, with the protein concentrated to

10 mg/ml. These crystals appeared after 3 weeks and grew over a 1-week period. The crystals can be replicated in a 24-well hanging drop format, but still grew as clusters of needles. The optimized condition has reduced protein concentration, now at 5 mg/ml, the protein is buffered in 10 mM Tris/HCl pH 7.4, 150 mM NaCl, 0.2 mM EDTA, and 1 mM DTT. The crystallization condition is composed of 0.1M HEPES pH 8.3, 50% MPD, and 0.1 M KCl as an additive. Optimization of these crystals drastically slowed nucleation with crystals appearing after 4 months but yielded single large crystals.

Data reduction is carried out using the HKL2000 suite (Otwinowski and Minor, 1997), molecular replacement is performed with PHASER from the PHENIX suite (Adams et al., 2010), using the 30 C-terminal residues of murine SUN2 coiled-coil structure (PDB code 5ED8) (Nie et al., 2016). Model refinement is carried out using phenix.refine. Anomalous sub-structure sites are found using the SHELX program suite HKL2map or with AutoSol from the PHENIX suite. Electron density and model placement is inspected using Coot (Emsley, 2010).

References

Alvarez, B. H., Gruber, M., Ursinus, A., Dunin-Horkawicz, S., Lupas, A. N., & Zeth, K. (2010). A transition from strong right-handed to canonical left-handed supercoiling in a conserved coiled-coil segment of trimeric autotransporter adhesins. *Journal of Structural Biology*, *170*(2), 236–245. http://doi.org/10.1016/j.jsb.2010.02.009

Andersen, K. R., Leksa, N. C., & Schwartz, T. U. (2013). Optimized E. coliexpression strain LOBSTR eliminates common contaminants from His-tag purification. Proteins: Structure, Function, and Bioinformatics, 81(11), 1857–1861. http://doi.org/10.1002/prot.24364

Cain, N. E., Tapley, E. C., McDonald, K. L., Cain, B. M., & Starr, D. A. (2014). The SUN protein UNC-84 is required only in force-bearing cells to maintain nuclear envelope architecture. *The Journal of Cell Biology*, *206*(2), 163–172. http://doi.org/10.1083/jcb.201405081

Chang, W., Worman, H. J., & Gundersen, G. G. (2015). Accessorizing and anchoring the LINC complex for multifunctionality. *The Journal of Cell Biology*, *208*(1), 11–22. http://doi.org/10.1083/jcb.201409047

Demircioglu, F. E., Sosa, B. S., Ingram, J. Ploegh, H. L., Schwartz, T. U. (2016). Structures of TorsinA and its disease- mutant complexed with an activator reveal the molecular basis for primary dystonia. *eLife.* 5:e17893. http://doi.org/10.7554/eLife.17983.001

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Cryst (2010). D66, 486-501 [Doi:10.1107/S0907444910007493]*, 1–16. http://doi.org/10.1107/S0907444910007493

Fridy, P. C., Li, Y., Keegan, S., Thompson, M. K., Nudelman, I., Scheid, J. F., et al. (2014). A robust pipeline for rapid production of versatile nanobody repertoires. *Nature Methods*, *11*(12), 1253–1260. http://doi.org/10.1038/nmeth.3170

Frey, S., & Görlich, D. (2014). A new set of highly efficient, tag-cleaving proteases for purifying recombinant proteins. *Journal of Chromatography A*, *1337*, 95–105. http://doi.org/10.1016/j.chroma.2014.02.029

Hildebrand, A., Remmert, M., Biegert, A., & Söding, J. (2009). Fast and accurate automatic structure prediction with HHpred. *Proteins: Structure, Function, and Bioinformatics*, *77*(S9), 128–132. http://doi.org/10.1002/prot.22499

Lindeman, R. E., & Pelegri, F. (2012). Localized Products of futile cycle/ lrmp Promote Centrosome-Nucleus Attachment in the Zebrafish Zygote. *Curbio*, *22*(10), 843–851.

http://doi.org/10.1016/j.cub.2012.03.058

Luxton, G. W. G., Gomes, E. R., Folker, E. S., Vintinner, E., & Gundersen, G. G. (2010). Linear Arrays of Nuclear Envelope Proteins Harness Retrograde Actin Flow for Nuclear Movement. *Science*, *329*(5994), 956–959. http://doi.org/10.1126/science.1189072

Nie, S., Ke, H., Gao, F., Ren, J., Wang, M., Huo, L., et al. (2016). Coiled-Coil Domains of SUN Proteins as Intrinsic Dynamic Regulators. *Structure/Folding and Design*, *24*(1), 80–91. http://doi.org/10.1016/j.str.2015.10.024

Otwinoski, Z., & Minor W. (1997). Processing of X-ray Diffraction Data in Oscillation mode. Methods in Enzymology, 276, 307–326.

Pardon, E., Laeremans, T., Triest, S., Rasmussen, S. O. R. G. F., nig, A. W. O., Ruf, A., et al. (2014). A general protocol for the generation of Nanobodies for structural biology. *Nature Protocols*, *9*(3), 674–693. http://doi.org/10.1038/nprot.2014.039

Sosa, B. A., Kutay, U., & Schwartz, T. U. (2013). Structural insights into LINC complexes. *Current Opinion in Structural Biology*, *23*(2), 285–291. http://doi.org/10.1016/j.sbi.2013.03.005

Sosa, B. A., Rothballer, A., Kutay, U., & Schwartz, T. U. (2012). LINC Complexes Form by Binding of Three KASH Peptides to Domain Interfaces of Trimeric SUN Proteins. *Cell*, *149*(5), 1035–1047. http://doi.org/10.1016/j.cell.2012.03.046

Starr, D. A., & Fridolfsson, H. N. (2010). Interactions Between Nuclei and the Cytoskeleton Are Mediated by SUN-KASH Nuclear-Envelope Bridges. *Annual Review of Cell and Developmental Biology*, *26*(1), 421–444. http://doi.org/10.1146/annurev-cellbio-100109-104037

Wang, W., Shi, Z., Jiao, S., Chen, C., Wang, H., Liu, G., et al. (2012). Structural insights into SUN-KASH complexes across the nuclear envelope. *Cell Research*, *22*(10), 1440–1452. http://doi.org/10.1038/cr.2012.126

Zhou, Z., Du, X., Cai, Z., Song, X., Zhang, H., Mizuno, T., et al. (2012). Structure of Sad1-UNC84 Homology (SUN) Domain Defines Features of Molecular Bridge in Nuclear Envelope. *Journal of Biological Chemistry*, *287*(8), 5317–5326. http://doi.org/10.1074/jbc.M111.304543

Chapter 3: Cytoplasmic Domains of LINC Complexes

Summary

The core of LINC complexes is formed by two proteins, SUN at the INM and the C-terminal domain of nesprins (KASH) at the ONM. The cytoplasmic domain of nesprins, predominantly serves as a structural protein that connects the nucleus to various elements of the cytoskeleton. Although the functional role of nesprins have been well studied, the large soluble domains of this family of proteins is structurally poorly characterized, especially in complex with other proteins. We set out to resolve the first structures of human nesprins and of Nesprin-2 in complex with various binding partners.

Here we present purification strategies for various fragments of Nesprin-2, FHOD1, Fascin-1, and Nesprin-3α that yield large amounts of pure, homogeneous protein. We first show that Nesprin-2 forms a stable complex with FHOD1, and we present our preliminary crystals and diffraction data of this complex. Next we discuss our attempts at forming a stable complex between Fascin-1 and Nesprin-2, as well as our attempts at crystallizing this complex.

Finally, we characterize the oligomeric state of Nesprin-3α in solution and using limited proteolysis assays we design minimal constructs that separate Nesprin-3α into two modules. With these minimal constructs we have been able to obtain preliminary crystals of various fragments of Nesprin-3α that together comprise the entire cytoplasmic domain of Nesprin-3α.

Introduction

The core of LINC complexes is formed within the perinuclear space (PNS) and consists of the conserved SUN domain of INM SUN proteins, and the C-terminal KASH peptide of Nesprins that traverse the ONM. SUN proteins are anchored at the INM by interactions with the nuclear lamina, predominantly with lamin A/C. While KASH domain proteins are anchored at the ONM through their interaction with the SUN domain (Starr, 2010; Sosa, 2013; Ungricht, 2015). In humans, 6 KASH domain proteins have been identified, four of which fall into the nesprin family of proteins (Nesprin-1/2/3/4) (Ketema, 2007; Lideman, 2012; Horn, 2013). The cytoplasmic domain of nesprins interact with various elements of the cytoskeleton (Chang, 2015). Of these, Nesprin-1 and Nesprin-2 giant (N2G) are the largest proteins, of 800 and 600 kDa, respectively. They are predominantly composed of repeats of a three helical bundle known as a spectrin repeat (SR) (Autore, 2013). At their N terminus both Nesprin-1 and N2G possess an actin binding domain (ABD). Their ABD is formed by two calponin homology domains (CH1/2), that are indispensable for nuclear anchoring to actin filaments (Zhang, 2002).

During cell migration in NIH3T3 fibroblasts rearward nuclear positioning takes place. This process depends on the formation of thick actin bundles and on the nucleus being physically connected to them as they move away from the leading edge. The nuclear connections are mediated by SUN2 and N2G that form linear arrays of LINC complexes known as transmembrane actin associated nuclear lines (TAN lines) (Luxton, 2010). The loss of the ABD of N2G abolishes nuclear migration, but does not affect the

formation of actin bundles. A small construct that includes the KASH domain of Nesprin-2 as well as the ABD is enough to rescue nuclear migration (Luxton, 2010). Recent work has shown that the SRs of N2G also play a role in TAN line function, specifically SR11-12 and SR51-53 (Figure 1). Both of these elements can bind to soluble actin binding proteins. SR11-12 bind to FHOD1, a formin homology protein that bundles actin filaments (Kutscheidt, 2014). While SR51-53 bind to Fascin-1, another actin bundling protein (Jayo, 2016). Loss of either FHOD1 or Fascin-1 affects nuclear polarization and results in reduced cell migration but does not abolish TAN line formation. This supports a model in which Nesprin-2 forms multiple points of contacts with actin filaments that enhance nuclear-cytoskeletal coupling in order to mediate TAN line formation and further stabilize the interaction.

FHOD1 is a unique formin homology protein. Typical formins cap the barbed end of growing actin filaments and promote elongation of existing actin polymers (Campellone & Welch, 2010.). FHOD1 however, bundles pre-existing actin filaments, a function similar to what is observed in actin stress fibers at the cell periphery and in TAN lines during fibroblast migration (Kutscheidt, 2014.). FHOD1, like other formins has a modular domain organization (Campellone & Welch, 2010). At the N terminus, formins have a GTPase binding domain (GBD), followed by a diaphanous inhibitory domain (DID) that serves as an actin binding domain (ABD). These domains form the N-terminal half of FHOD1 and have been shown to be sufficient for centrosome orientation during cell migration, and can partially rescue rearward nuclear positioning during TAN line

formation in FHOD1$^{-/-}$ cells (Kutscheidt, 2014). The C-terminal half of FHOD1 comprises a predicted short coiled-coil motif that has been shown to mediate dimerization of FHOD1 (Madrid, 2004), followed by the the eponymous actin binding formin homology (FH1 and FH2) domains. The FH1/FH2 domains serve the generic function of binding actin in this context, as they can be replaced with the CH1/2 ABD domains of α-actinin, and FHOD1 function remains unchanged (Jayo, 2016). Since FHOD1 can dimerize, with one Nesprin-2 binding site per monomer, it can potentially tether two neighboring LINC complexes together. Thus the FHOD1 Nesprin-2 interaction is a potential candidate for the formation of high order LINC complex arrays.

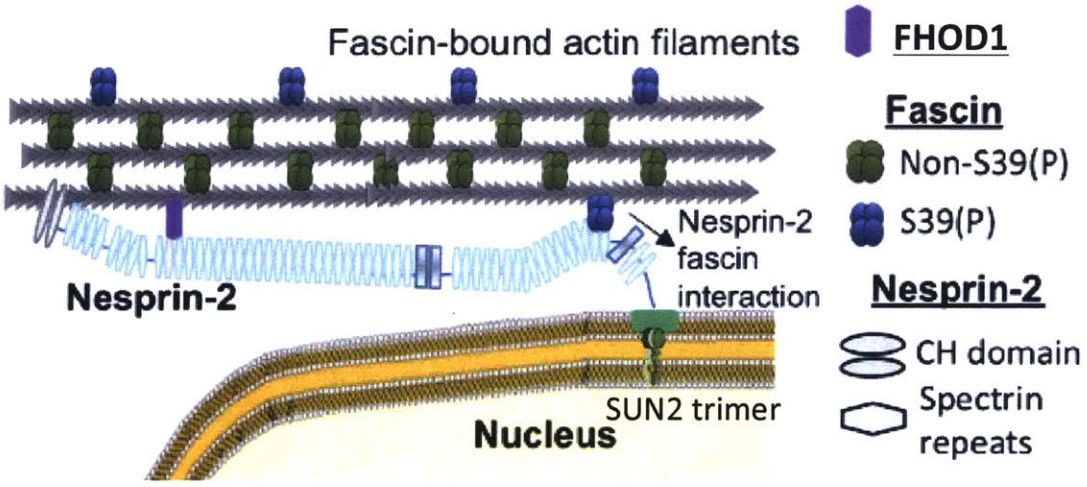


Figure 1. Nesprin-2 binds to actin directly and indirectly. Nesprin-2 can bind directly to actin through its N-terminal CH1/CH2 domain, and indirectly by associating with FHOD1 and Fascin-1. *Modified from Jayo, 2016.*

Fascin-1 is an actin binding protein that bundles parallel actin filaments (Figure 1) and is known to stabilize filopodia in migrating cells (Türmer, 2015). Fascin-1 is only found at

low levels in normal epithelia. High levels of fascin-1 mRNA in tumors however is correlated to a poor patient prognosis. Drugs that bind to the fascin-1ABD inhibit cell migration by interfering with fascin-1's actin bundling function, making fascin-1 a suitable and desirable target for drug discovery (Chen 2010; Zhang, 2015). Recent work has shown that Fascin-1 interacts directly with N2G and tethers it to actin during TAN line formation (Jayo, 2016). This interaction connects the C-terminal SR51-53 of N2G to actin, SR51-53 lie proximal to the INM as opposed to the FHOD1 connection at the N-terminal domain of N2G to actin (Figure 1). Fascin-1 is composed of four compact ß-trefoil domains, of these, ß-trefoil-3 has been shown to mediate Nesprin-2 binding (Jayo, 2016). This interaction is necessary for nuclear deformation during cell migration since cells lacking fascin-1 cannot traverse tight spaces that require the nucleus to be deformed (Pfisterer, 2017). Overexpression of the ß-trefoil-3 domain of Fascin-1 uncouples N2G SR51-53 from actin, and is enough to inhibit cell migration through pores smaller than 10μM. Taken together, the interaction between Fascin-1 and N2G adds another connection to actin that has far reaching implications in tumor metastasis.

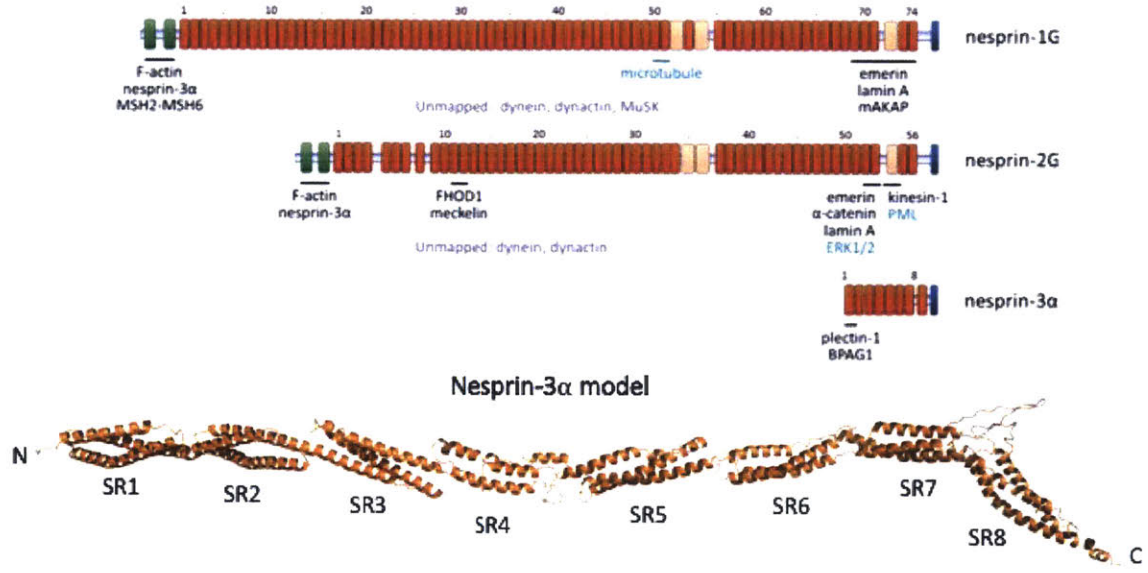


Figure 2. Schematic representation of Nesprin-1/2/3. Top, domain architecture of mammalian Nesprin-1/2/3 to scale. CH1/2 ABD are shown in green, SR are orange, KASH domain is colored blue. SRs are numbered, underlined segments interact with proteins denoted underneath *Adapted from Chang, 2015.* Bottom, model of Nesprin-3 based on structural homology to other spectrin repeat containing proteins. SR-Spectrin repeat.

Compared to Nesprin-1/-2, discussed above, Nesprin-3α is much smaller with a molecular weight of 100 kDa. A total of 8 spectrin repeats form the majority of the cytoplasmic domain. At its N terminus, SR1 of Nesprin-3α is a plectin binding domain, as opposed to a globular ABD like its homologs Nesprin-1/2 (Figure 2). Plectin, in turn can bind to intermediate filaments, and to actin through its own ABD domain, also composed of a bipartite CH1/CH2 domain (Ketema, 2007). Nesprin-3α has been shown to directly interact with the ABD of both Nesprin-1/2, potentially regulating their interaction with actin filaments (Lu, 2012). Since plectin bridges various cytoskeletal

elements together, and Nesprin-3α can interact with other nesprins it is likely that they form a meshwork around the nucleus that may regulate the size and deformability of the organelle (Morgan, 2011; Lu, 2012; Petrie, 2017).

In the present study we set out to structurally characterize the cytoplasmic domains of various nesprins. First through structural homology prediction, we designed Nesprin-2/3 constructs that yield large quantities of protein when expressed in *E. coli*. We were able to form complexes between N2G and FHOD1, and we determine the stoichiometry of this complex to be 1:1. We obtained crystals of the complex that diffract to moderate resolution. We also attempted to form a stable complex between N2G and Fascin-1, however, we observed a dynamic interaction, challenging structural characterization. Finally, we characterized the oligomeric state of Nesprin-3α in solution, and, using limited proteolysis, we designed optimal crystallization constructs. We were able to obtain large three dimensional crystals of overlapping fragments of Nesprin-3α that, in sum, provide complete structural coverage of the protein.

Results

Purification of Nesprin-2 spectrin repeats 11-12 and spectrin repeats 51-53

First, we generated a construct of N2G spanning residues (1340-1678), supposedly containing spectrin repeats 11 and 12 (Figure 3). The construct was fused N-terminally to a 6 histidine, 7 arginine, SUMO affinity/solubility handle. Initial attempts at overexpression of this construct yielded insoluble protein, even with the solubility tag. This result was somewhat surprising, given that recombinant expression of the same fragment yielded soluble protein before, according to the literature (Kutscheidt, 2014) To improve the construct design, we resorted to structural prediction (hhpred; Hildebrand, 2009). The resulting protein homology model indicated that the original construct contained spectrin repeats 11 and 12, but, in addition had a single helix from spectrin repeat 10 at the N terminus and two helices of spectrin repeat 13 at the C terminus, respectively. Since these partial domains most likely caused the protein to aggregate, We redesigned the protein construct to only include spectrin repeats 11 and 12, now N2G SR11-12 (residues 1425-1649). With these modifications, the expression levels of N2G SR11-12 significantly increased.

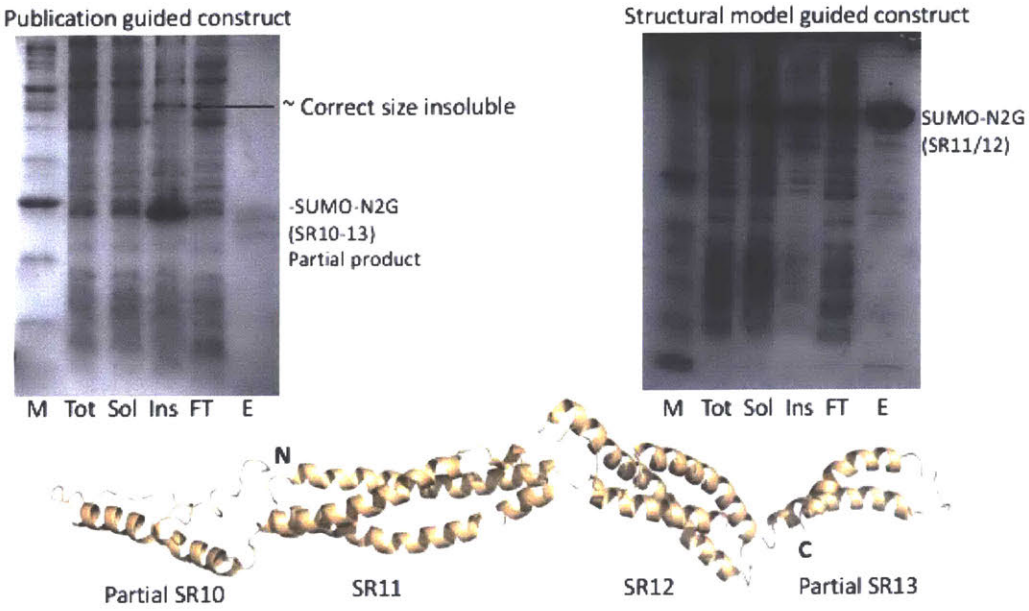


Figure 3. Optimization of N2G SR11-12 construct. Top, SDS-PAGE gels of nickel affinity purification. Top left, construct based on previously published purification; top right, optimized construct based on structural modeling. Bottom, structural model of construct that is insoluble (from top left gel). Tot-total lysate, Sol- soluble fraction, Ins-insoluble fraction, FT- flow through from nickel pull down, E-nickel elution.

The purification protocol consists of three chromatographic steps. The first purification step is Ni-affinity to isolate N2G from the soluble bacterial lysate. The second step is size exclusion chromatography, which yields homogenous protein. The solubility/affinity tag is cleaved with 3C protease and captured by cation-exchange chromatography due to the 7xArg element. With this protocol we were able to obtain high yields of pure protein (~2mg per liter of bacterial culture) (Figure 4).

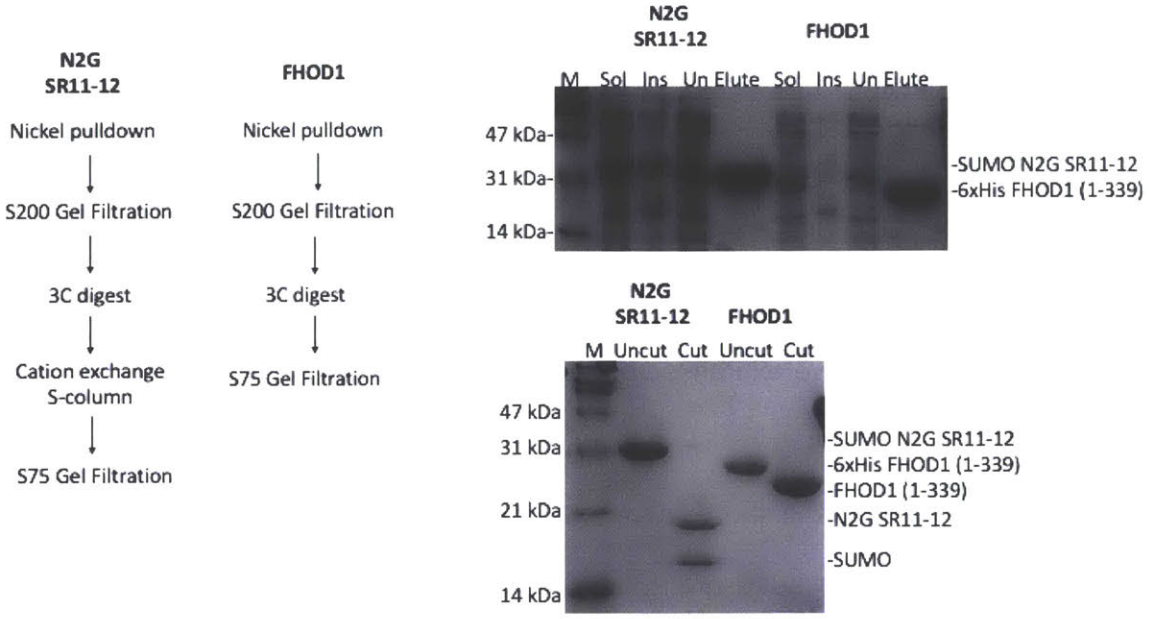


Figure 4. Purification scheme of N2G SR11-12 and FHOD1. Left, flowchart of N2G SR11-12 and FHOD1 purification scheme. Right, top gel shows nickel pull-down of N2G SR11-12 and FHOD1. M-protein marker, Sol-soluble fraction, Ins-insoluble fraction, Un-unbound to nickel, Elute-nickel elution. Bottom gel shows 3C protease digest.

Before purifying Nesprin-2 spectrin repeats 51-53 (N2G SR51-53) we again performed structural modeling to ensure the presence of complete spectrin repeat domains in the construct. We generated constructs of N2G SR51-53 that included residues 6017-6355. Using a similar purification strategy as with N2G SR51-53, except for the presence of an N-terminal 14 his-bdSUMO solubility tag, I was able to express large soluble quantities of the protein. This purification also starts with a Ni-affinity pull-down, but instead of eluting with imidazole we cut the solubility tag on the beads with the SENp protease and wash off our purified protein (Frey, 2014). The protein is further purified via size

exclusion chromatography, and finally yields about 15 mg pure protein per liter of bacterial culture (Figure 6).

Purification of FHOD1 and Fascin-1

Initially, we attempted to purify full length FHOD1$_{(1\text{-}1164)}$. While the expression levels were quite high and an overexpressed band is detected via SDS-PAGE analysis of the soluble fraction, there was severe C-terminal degradation of FHOD1$_{(1\text{-}1164)}$. From this result we decided to use the shortest construct that is known to bind N2G SR11-13, which consists of the previously crystallized GBD and DID domain encompassing FHOD1$_{(1\text{-}339)}$ (Figure 5). This construct was expressed with an N-terminal 14xHis-bdSUMO tag and was purified as described for N2G SR51-53.

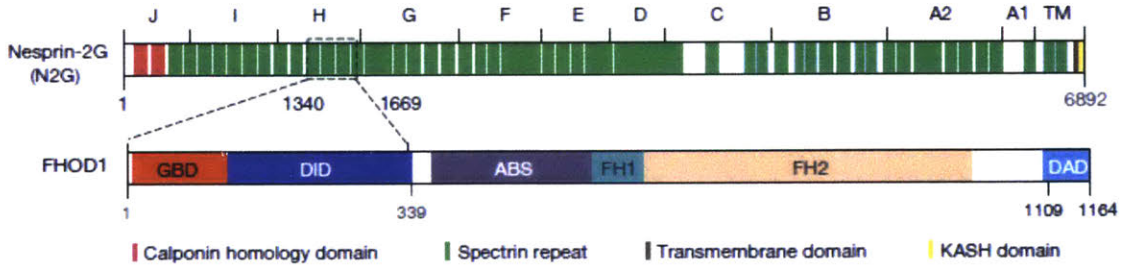


Figure 5. Schematic diagram of N2G and FHOD1. Domain architecture of N2G, and FHOD1. Highlighted fragments of FHOD1 and N2G are sufficient to mediate interaction. *Adapted from Kutscheidt, 2014.*

For Fascin-1, we used the same tagging strategy (N-terminal 14xHis-bdSUMO). Full-length Fascin-1$_{(1\text{-}493)}$ was expressed highly and was cut while bound to nickel beads

using the bdSUMO protease SENp. Further purification steps were as described for N2G SR11-13 (Figure 6).

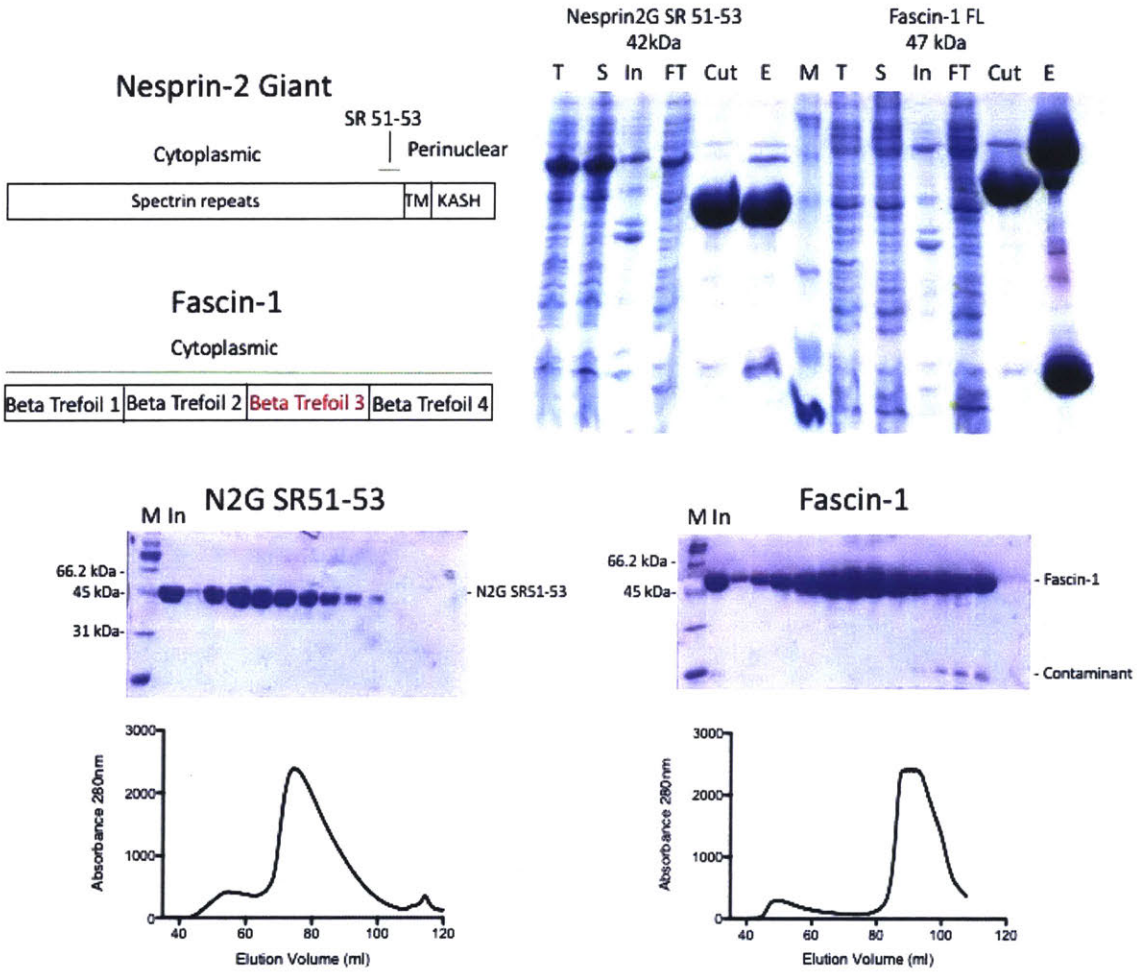


Figure 6. Purification of N2G SR51-53 and Fascin-1. Top left, diagram of purified proteins, green bar shows fragment of each protein used in this purification. Top right, SDS-PAGE gel of nickel affinity purification of fascin-1 and N2G SR11-12. T-total lysate, S-soluble fraction, In-insoluble fraction, FT- flow through, Cut-wash after 1-hour incubation with SENp, E-elution using imidazole. Bottom, gels correspond to each gel filtration chromatogram. Input for size exclusion is taken from the "cut" lane of the nickel affinity gel.

Complex formation of FHOD1 and N2G SR11-12

To form the complex, we first tried mixing FHOD1 and N2G SR11-12 at a 1 to 1 molar ratio to form a stoichiometric complex. Initially complex formation in our standard gel filtration buffer (10 mM Tris/HCl pH 7.4, 150 mM NaCl, 1 mM DTT, 0.2 mM EDTA) was unsuccessful. Both proteins eluted separately (data not shown). However, we noticed a broadening of the FHOD1 peak in the gel filtration profile, when run together with N2G SR11-12 rather than alone. This suggested that the proteins did interact, but the complex was not stable. We addressed this two ways, first the proteins were concentrated to 10 mg/ml before loading and instead of standard gel filtration buffer we employed a low salt version of this buffer to enhance protein-protein interactions with only 50 mM as opposed to 150 mM sodium chloride. With this adjustment, we were able to form a complex that was stable during our size exclusion chromatography step (Figure 7).

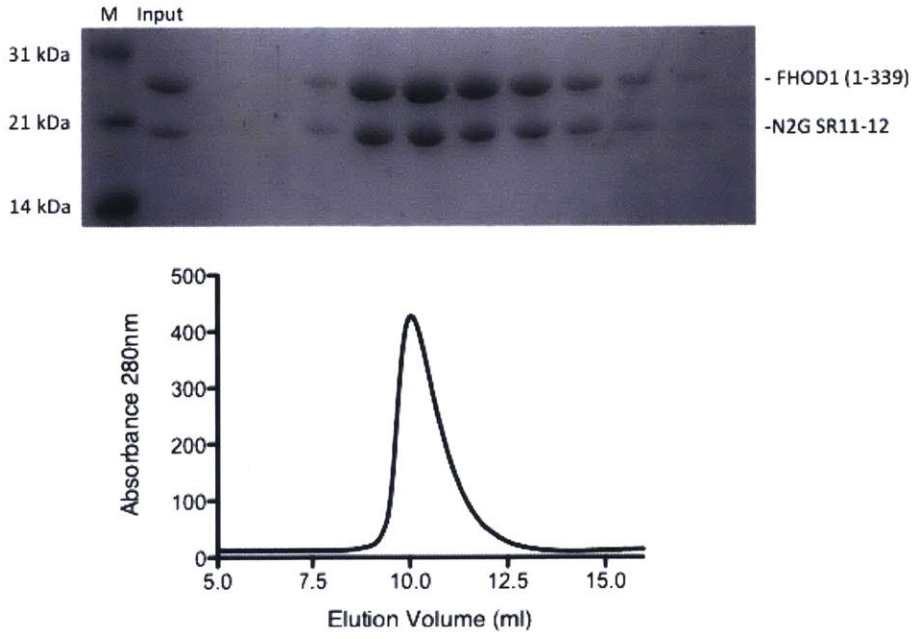


Figure 7. Complex formation of FHOD1$_{(1-339)}$ and Nesprin-2$_{(1425-1649)}$. Under low salt conditions, FHOD1 and N2G SR11-12 form a stable complex. SDS-PAGE corresponds to the main peak of the chromatogram.

We next asked what the stoichiometry of the FHOD1 and N2G SR11-12 complex was. To this end, using the low salt buffer we performed SEC-MALS analysis of the protein complex. The complex eluted as a single peak, and as expected the stoichiometry of the complex was 1:1, with an estimated molecular weight of 61.7 kDa (Figure 8). This closely matches the expected molecular weight of 63 kDa for 1:1 complex. We next attempted to crystallize the complex.

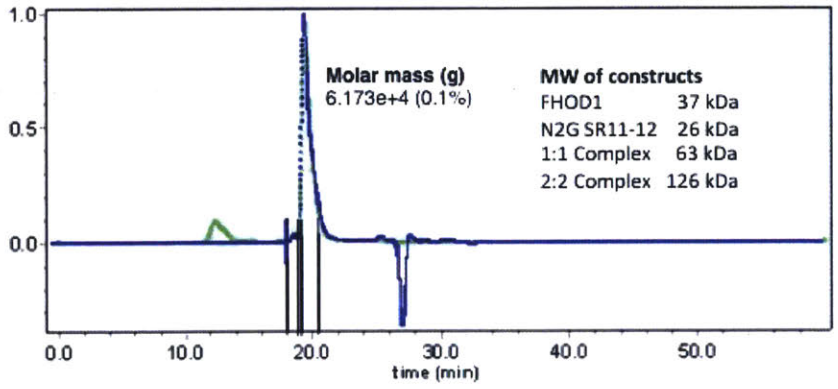


Figure 8. FHOD1 forms a 1:1 complex with N2G SR11-12. Size exclusion chromatography with multi-angle laser scattering (SEC-MALS). Chromatogram shows that the main peak corresponds to a molar mass of 61.7 kDa, closely matching a 1:1 complex that has an expected molecular weight of 63 kDa.

Crystallization, data collection, and data processing of FHOD1 and N2G SR11-12 complex

Protein crystals of FHOD1 and N2G SR11-12 complex appeared after 2 days in an Index HT screen. Initial crystals appeared in a sitting drop format, at 18 °C, in 15% PEG3350 and 0.1 M di-sodium succinate. The crystals were readily reproduced in a 24 well sitting drop format and the crystal size scaled with the drop size (Figure 9). The crystals were optimized and 1% glycerol was used as an additive which allowed the growth of single crystals of about 100 μM x 30 μM x 30 μM. These crystals were cryoprotected in either 30% PEG3350, or 15% glycerol. Some of the thinner crystals cracked or bent in the cryo-condition while harvesting them. The larger crystals however behaved well upon freezing. The best native crystal diffracted to 3.4 A (Figure 10).

Figure 9. Crystals of FHOD1 and Nesprin-2 complex. Left, initial crystals from a sparse matrix screen. Middle panel shows the optimized crystals in a 24-well hanging drop format. Right, washed crystals were crushed, dissolved and run on an SDS-PAGE stained with Coomassie Brilliant Blue. Two bands of comparable intensity are visible that correspond to FHOD1 and N2G SR11-12, confirming that crystals contain both proteins at a 1:1 ratio.

Initial data analysis using Xtriage suggested that a single copy of the complex was present in the asymmetric unit and no serious crystal pathologies were present. To

solve the phase problem, we employed molecular replacement using PHENIX MR (Adams, 2011). Our search model was the previously solved structure of $FHOD1_{(1-339)}$. An initial search yielded a solution with a TFZ score of 8.1 and an LLG score of 104.6. The armadillo repeats of FHOD1's DID domain was generally enveloped in density with moderate difference density suggesting some conformational change from the previously solved structure. The solution for the GBD domain however was clearly incorrect. Most of the GBD domain was not properly enveloped in electron density. Since both the GBD and DID domains have a linker between them it is possible that their positions relative to each other are different in our structure. To address this, we next searched for the armadillo repeat, and the GBD domain separately. As expected the same solution for the armadillo repeat was easily found, but no solution was found for the GBD domain on its own.

The structure of N2G SR11-12 has not been previously solved, however we are able to generate a homology model of the protein with fairly high confidence. We tried searching using our SR11-12 model, as well as the individual SRs, in the presence or absence of the armadillo repeat from our previous solution. However, no acceptable solution was found with any of these search procedures.

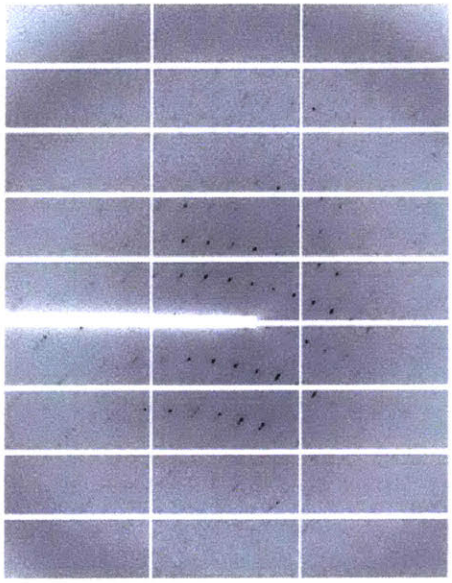


Table 1. Data collection and refinement statistics

Protein	FHOD1 N2G SR11-12	FHOD1 N2G SR11-12 Ta
Organism	*H. sapiens*	*H. sapiens*
Data collection		
Space group	P 21 21 2	P 21 21 2
a, b, c (Å)	202.6, 52.5, 81.5	203.2, 51.2, 80.9
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Wavelength (Å)	0.9791	1.2543
Completeness (%)	98.9 (94.4)	97.5 (96.0)
Redundancy	10.5 (7.8)	2.1 (1.9)
Rp.i.m. (%)	5.1 (65.0)	7.0 (53.8)
I/σ	68.9 (0.77)	24.7 (0.77)
CC1/2 (%)	99.9 (50.2)	99.2 (48.5)
Resolution Range (Å)	99.0 – 3.3	99.0 – 6.8

Figure 10. Diffraction pattern and data processing statistics. Diffraction pattern of native crystals of The FHOD1-N2G SR11-12 complex. Data collection and refinement statistics correspond to the native and derivatized dataset, respectively. Values in parenthesis correspond to the highest resolution shell.

In order to solve the structure, we decided to obtain experimental phases. We did this by soaking the crystals in mother liquor containing different heavy metals and with the use of a selenomethionine derivatized protein complex. These crystals diffracted poorly though, and the resolution was too low to find proper substructure solutions, or even detect signal for the heavy atoms employed. Because the cryo-conditions are still imperfect, we are currently screening multiple cryo-conditions in order to improve the quality of data obtained from both the native and derivatized crystals.

Purification of Nesprin-3α

We set out to establish a purification protocol that would yield large amounts of chromatographically pure Nesprin-3α in *E. coli* in order to pursue crystallographic, and biophysical characterization of Nesprin-3α. Nesprin-3α has a single transmembrane domain comprising residues 926-946 that anchors the cytoplasmic domain to the ONM (Figure 11). The transmembrane domain, including the KASH peptide, is excluded from all our constructs. Additionally, secondary structure predictions and structural homology modeling suggest that the residues adjacent to the transmembrane helix which connect SR7 to SR8 are unstructured (810-850). Therefore we also excluded this region from our constructs. All the Nesprin-3α constructs described here are cloned as N-terminal fusions to 14xHis-bdSUMO. We cut the protein while attached to the nickel beads and collect the supernatant, this yields about 10 mgs of fairly pure Nesprin-3α $_{(1\text{-}807)}$ starting from a 1-liter bacterial culture (Figure 5). The cut protein was then purified to homogeneity and separated from co-purifying nucleic acids on a Superdex 200 16/60 column. All fragments of Nesprin-3α that we employ are purified using the same method.

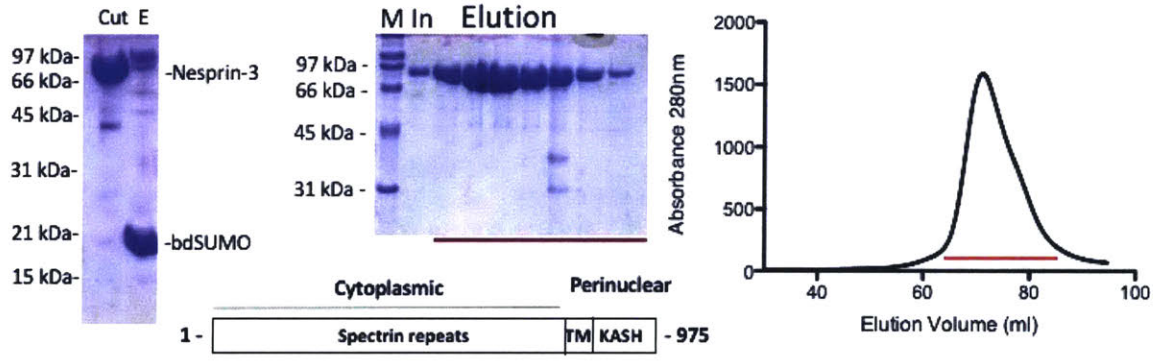


Figure 11. Purification of the cytoplasmic domain of Nesprin-3α. Left, the 14xHis-bdSUMO tag of Nesprin-3α $_{(1\text{-}807)}$ is proteolytically cleaved while bound to nickel (left lane). Subsequently, uncut protein and solubility tag are eluted with imidazole (right lane). Right, SDS-PAGE gel corresponds to the main peak from the chromatogram, cut fraction is used as input for gel filtration. Gel lanes and corresponding fractions from gel filtration are underlined in red. Bottom, schematic diagram of Nesprin-3α, green line represents fragment of protein used in this purification (domains not shown to scale).

Optimization of Nesprin-3α constructs via limited proteolysis

Our initial attempts at crystallizing the cytoplasmic domain of Nesprin-3α did not yield any crystals. Since our bioinformatics analysis of Nesprin-3α suggested a modular structure we decided to separate the protein into its various constituent modules and crystallize those individually. The first change we made was removing the last 60 C-terminal residues that although secondary structure prediction models helical, tertiary structure prediction indicates that this element does not form part of a SR, and is instead solvent exposed. We reasoned that this region is probably a linker that adds conformational flexibility between SR7 and SR8 which is already excluded from our constructs. Spectrin repeats are three-helix bundles and in some cases the third helix of one bundle can be quite long, such that it forms the first helix of the next bundle. Because of this we did not want to simply design constructs that removed single spectrin domains without knowing if there was structural relation between adjacent SRs. To address this, we used a limited proteolysis assay.

We first tested both V8 and Trypsin digestion of purified Nesprin-3α $_{(1\text{-}807)}$ (Figure 12). V8 digestion showed that even at very low concentrations a roughly 10 kDa piece of Nesprin-3α $_{(1\text{-}807)}$ is proteolytically cleaved, this most likely corresponds to the last ~60 residues that we had removed from most of our constructs already. Otherwise, the V8 digest results were not particularly informative. Trypsin digest, however yielded very clear results. In the presence of trypsin, Nesprin-3α $_{(1\text{-}807)}$ was digested into two major fragments of about 50 and 40 kDa, respectively. We next tested if these two fragments interacted. To this end we digested about 500 μg of Nesprin-3α $_{(1\text{-}807)}$ with trypsin, stopped the reaction with PMSF and injected the digested protein onto an Superdex 200 10/300 size exclusion column (Figure 12). While we observed one major peak, we observed the two fragments running at a distinct offset from one another, indicating that they do not interact.

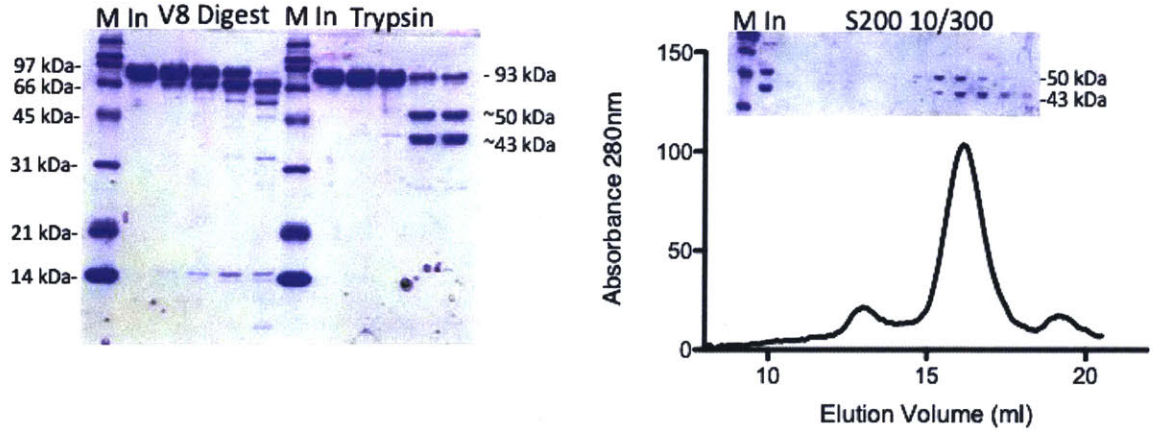


Figure 12. V8 and Trypsin digest of Nesprin-3. Left panel, protease digest using either V8 or Trypsin, protease concentration increases from left to right. Right panel, size exclusion chromatogram of trypsin digested Nesprin-3$_{(1\text{-}807)}$.

Trypsin cuts C-terminal to arginine or lysine. Homology models showed that there are two possible sites with a positive charge in a region that is not structured, and would yield fragments of about the correct size. One site is between SR3 and 4, and the second site is between SR4 and 5. To quickly determine at which of these two sites trypsin digested we submitted a series of C-terminal truncations of Nesprin-3α to limited proteolysis (Figure 13). In all these digests the smaller molecular weight fragment did not shift in size, while the larger molecular weight fragment was smaller as our construct receded from the C terminus. Clearly, this shows that the large molecular weight fragment formed the C-terminal portion of our protein, and therefore trypsin must digest between SR3 and 4.

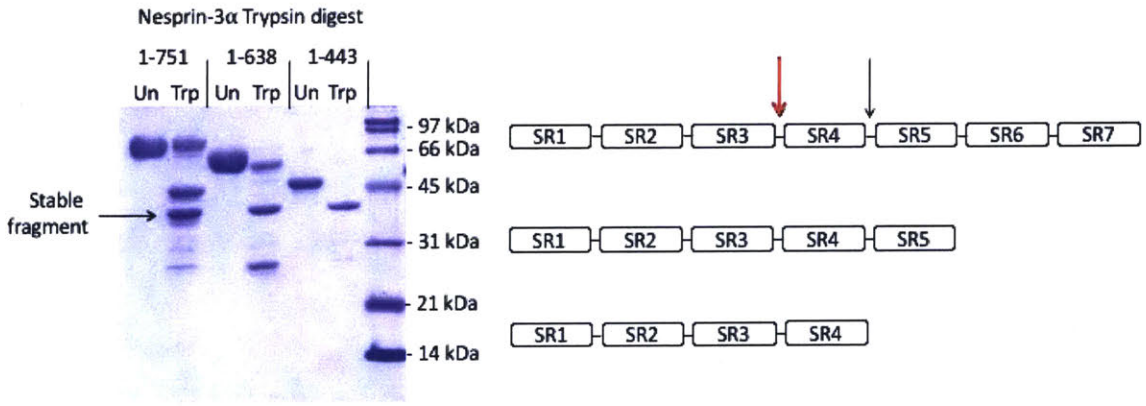


Figure 13. Identifying trypsin digest site. Left, Trypsin digests of various constructs that have been shortened from the C terminus. Arrow shows the fragment that is invariable in size. Right, cartoons show the two possible cut sites for trypsin. If the arrow between SR3-4 is the trypsin cut site, the smaller molecular weight fragment on the gel will be invariable. If the arrow between SR4-5 is the trypsin cut site, the larger molecular weight fragment will be invariable in size. Red arrow indicates trypsin cut site.

Based on this result we generated two Nesprin-3 constructs, one that contained the N-terminal portion comprised of SR1-SR3, and another that contained SR4-7. SR4-7 and SR5-7 yielded crystals in our initial screens that are readily reproducible in a 24 well format (Figure 14). The N-terminal SR1-SR3 also show promising preliminary crystals in initial screens that we will further optimize.

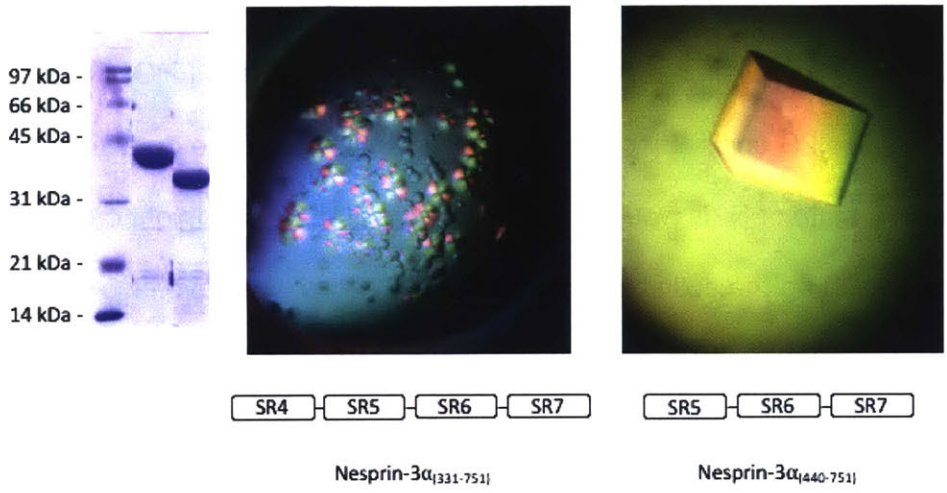


Figure 14. Crystals of Nesprin-3. Left, purification of Nesprin-$3_{(331-751)}$ and Nesprin-$3_{(440-751)}$, constructs based on limited proteolysis assay. Right, optimized crystals of purified Nesprin-$3_{(331-751)}$ and Nesprin-$3_{(440-751)}$.

Discussion

Our characterization of the oligomeric state of $FHOD1_{(1-339)}$ and N2G SR11-12 showed that the complex forms with a 1:1 stoichiometry. The oligomeric state of N2G has not been previously reported, although proteins with similar domains such as plectin have been reported to exist as dimers or tetramers (Fontao, 2001). It is possible that our short N2G fragment does not recapitulate the oligomeric state of an entire N2G protein. If N2G exists as a dimer in cells, or as higher oligomeric state, perhaps FHOD1 can interact with a single N2G chain. Since a single N2G chain can bind to an actin filament, a higher oligomeric state of N2G would juxtapose multiple ABDs and might cross-link actin filaments, stabilizing the actin bundles seen in TAN lines. We have also confirmed the direct interaction between N2G and FHOD1 that has been shown to provide a second contact point between N2G and actin. FHOD1 has been shown to dimerize through its coiled-coil domain proximal to its actin binding FH2 domain. Dimerization of FHOD1 could mediate clustering of N2G by bringing together adjacent N2G filaments. Moreover, since FHOD1 also has its own ABD, it can aid in bundling actin filaments already observed in TAN lines further stabilizing them. The oligomeric states of FHOD1 and potential oligomers N2G suggest that avidity can play a role in TAN line stabilization or function.

During our attempts at forming a complex between Fascin-1 and N2G SR51-53 the proteins did not stably bind, as judged by gel filtration. It is possible that this interaction has fast kinetics, or is of moderate affinity when compared to the interaction between

FHOD1 and N2G SR11-12. Despite our 1:1 mix of Fascin-1 and N2G SR51-53 the crystals we obtained in multiple conditions only contain Fascin-1, not the complex. These results do not preclude the formation of a stable complex, there are many variables that still remain unexplored and may yield a stable complex such as screening pH, different ions, different concentrations of salts or an additive that is present in a cell but lacking in our buffer. Currently, we have purified an Avi-tagged versions of both Fascin-1, and N2G SR51-53 in order to characterize complex formation by studying the kinetics of the interactions using biolayer interferometry. Identifying a condition in which Fascin-1 and N2G SR51-53 interact stably would pave the way for further structural and biochemical characterization of this complex.

Our purification of Nesprin-3α and gel filtration analysis suggests that Nesprin-3α exists as a monomer in solution. Furthermore, despite the presence of SRs, Nesprin-3 elutes at the expected volume for a ~90 kDa protein, suggesting that Nesprin-3α may not be elongated in solution but perhaps adopts some more compact fold. Additionally, the limited proteolysis shows that there are stable domains in the protein and that the two main fragments composed of the N-terminal and C-terminal domain are non-interacting. Therefore, if the fold is compact it most likely consists of an intrachain interaction within the very N-terminal and SR1-3 or within SR4-7. Our purification and characterization of Nesprin-3 in solution has provided insight that was invaluable in optimizing our constructs and obtaining crystals. Further structural and biophysical characterization of

Nesprin-3α will pave the way for studying Nesrpin-3α in complex with its various binding partners and provide a clear structural picture of nuclear/cytoplasmic anchoring.

Materials and Methods

Plasmids and protein expression

Recombinant proteins were all expressed in *E. coli*. FHOD1, Fascin-1, N2G SR51-53 and all Nesprin-3 fragments were expressed from a modified ampicillin resistant pET-DUET1 (Novagen) vector. All constructs are expressed with an N-terminal 14x-histidine-bdSUMO tag that is cleavable using the bdSENp protease at a 1 to 1000 ratio for an hour at 4°C (ref.). N2G SR 11-13 is expressed in another modified Kanamycin resistant pET-DUET vector with an N-terminal 7x-histidine-10x-arginine-SUMO affinity/solubility tag. The tag is removable by proteolytic cleavage with rhinovirus 3C protease. FHOD1 and N2G SR11-13 was amplified from human cDNA and inserted into the corresponding vector by Gibson assembly (Gibson, 2009). Fascin-1, N2G SR51-53, and Nesprin-3 were sub-cloned from mammalian expression vectors into bacterial expression vectors via Gibson assembly. All Nesprin-3α truncations were generating using inverse PCR.

Transformed LOBSTR(DE3)-RIL bacterial expression cells (Andersen, 2013) were grown at 37 °C to an OD600 of 0.6, then shifted to 18 °C and induced with 0.2 M isopropyl β-D-1-thiogalactopyranoside for 16 hours. Cells were harvested by centrifugation at 6000 g, resuspended in lysis buffer (50 mM potassium phosphate, pH 8.0, 400 mM NaCl, 40 mM imidazole) and lysed using an LM20 Microfluidizer Processor (Microfluidics). The lysate was cleared by centrifugation at 10000 g for 25 minutes. The soluble fraction was incubated with Nickel Sepharose 6 Fast Flow beads (GE Healthcare) for 30 minutes at 4 °C in batch. After the nickel beads were washed with

lysis buffer, for bdSUMO tagged proteins, they were resuspended in 4 bed volumes of lysis buffer and 15ug of bdSENp was added. Reaction was allowed to proceed for 1 hour at 4 °C. Cut protein is purified by collecting the supernatant. Uncut protein, and excess solubility tag that was bound to nickel was eluted in elution buffer (10 mM Tris/HCl pH 8.0, 150 mM NaCl, 250 mM imidazole). Protein that possessed a 3C protease cleavage site was directly eluted from nickel followed by a size exclusion chromatography column on an Superdex 75 16/60 (GE healthcare) pre-equilibrated in running buffer (10 mM Tris/HCl, pH 8.0, 150 mM NaCl, and 0.2 mM EDTA, 1 mM DTT). The main peak was collected and the solubility tag cut overnight with rhinovirus 3C protease at a 1 to 200 molar ratio. Cut protein was applied to a cation exchange column to remove the cut solubility tag and to separate any uncut protein. Cut protein is collected in the flow through, while both uncut protein and cut solubility tag bind to the cation exchange column, and elute at around 600 mM NaCl. After this the purification of all proteins was identical. All proteins were further purified by size exclusion chromatography using an Superdex 75 or Superdex 200 16/60 column (GE Healthcare) equilibrated in running buffer.

Complex formation of FHOD1 and N2G SR11-12

FHOD1 and N2G SR11-12 were purified separately and mixed at a 1:1 ratio to form a complex. The complex was dialyzed into low salt buffer (10 mM Tris/HCl pH 7.4, 50 mM NaCl, 0.2 mM EDTA, 1 mM DTT) overnight and was then purified to homogeneity on an Superdex 75 column (GE Healthcare). Stoichiometric complex formation was verified by

SDS-polyacrylamide that was stained with Coomassie Brilliant Blue. Complex was then concentrated for further characterization and crystallization to 10 mg/ml, the remaining protein complex was flash frozen in liquid nitrogen and stored at -80 °C.

Limited proteolysis assay

Limited proteolysis assays were initially performed with V8 protease and trypsin using final concentration range of 1 – 30 ng of protease to digest 200 μg in order to identify stable fragments of Nesprin-3. Protease degradation reactions were allowed to proceed for 1 hour at room temperature and were stopped by adding 5x SDS loading dye. The samples were boiled and loaded onto an SDS-polyacrylamide gel that was stained using Coomassie Brilliant Blue. For size exclusion chromatography experiments of the proteolysed fragments, the limited proteolysis reaction was stopped by adding PMSF to a final concentration of 50 uM. The samples were then loaded onto an Superdex 75 10/300 (GE Healthcare) column, eluted samples were analyzed by SDS-polyacrylamide gels and Coomassie Brilliant Blue staining.

Crystallization and data processing

$FHOD1_{(1\text{-}339)}$ and N2G SR11-12 complex crystallized in 15% PEG3350, 0.1 M sodium succinate, and 1% glycerol. Crystals grew overnight and were rod shaped with dimensions of about 150 μm x 30 μm x 30 μm. Crystals were cryoprotected in 15% glycerol or 30% PEG3350 in three steps of increasing glycerol concentration. Data reduction was carried out using HKL2000 (Otwinowski and Minor, 2007). Molecular

replacement and experimental phasing was performed using the PHENIX suite (Adams et al., 2010). Protein models and density maps were visualized with Coot (Emsley et al., 2010).

Various fragments of purified Nesprin-3 were subjected to multiple crystallization screens at a concentration of 10 mg/ml. Nesprin-3$\alpha_{(331-751)}$ and Nesprin-3$\alpha_{(440-751)}$ both produces initial crystal hits that grew as round crystals or needle-clusters respectively. Nesprin-3$\alpha_{(331-751)}$ crystallized in 19% PEG3350 and 0.3 M potassium formate. While Nesprin-3$\alpha_{(440-751)}$ crystallized in 0.1 M MES pH 6.5, 10% isopropanol, 0.1 M magnesium chloride and 5% PEG4000. Nesprin-$3_{(440-751)}$ morphology was improved to single diamond shaped crystals by 0.02 M EDTA as an additive. Both crystals were replicated on a 24-well hanging drop format, where crystal size scaled with drop size.

References

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Echols, N., Headd, J. J., et al. (2011). The Phenix software for automated determination of macromolecular structures. *Methods*, *55*(1), 94–106. http://doi.org/10.1016/j.ymeth.2011.07.005

Andersen, K. R., Leksa, N. C., & Schwartz, T. U. (2013). Optimized E. coliexpression strain LOBSTR eliminates common contaminants from His-tag purification. *Proteins: Structure, Function, and Bioinformatics*, *81*(11), 1857–1861. http://doi.org/10.1002/prot.24364

Autore, F., Pfuhl, M., Quan, X., Williams, A., Roberts, R. G., Shanahan, C. M., & Fraternali, F. (2013). Large-Scale Modelling of the Divergent Spectrin Repeats in Nesprins: Giant Modular Proteins. *PLoS ONE*, *8*(5), e63633–14. http://doi.org/10.1371/journal.pone.0063633

Campellone, K. G., & Welch, M. D. (2010). A nucleator arms race: cellular control of actin assembly. *Nature Reviews Molecular Cell Biology*, *11*(4), 237–251. http://doi.org/10.1038/nrm2867

Chang, W., Antoku, S., Östlund, C., Worman, H. J., & Gundersen, G. G. (2015a). Linker of nucleoskeleton and cytoskeleton (LINC) complex-mediated actin-dependent nuclear positioning orients centrosomes in migrating myoblasts. *Nucleus*, *6*(1), 77–

88. http://doi.org/10.1080/19491034.2015.1004947

Chang, W., Worman, H. J., & Gundersen, G. G. (2015b). Accessorizing and anchoring the LINC complex for multifunctionality. *The Journal of Cell Biology*, *208*(1), 11–22. http://doi.org/10.1083/jcb.201409047

Chen, L., Yang, S., Jakoncic, J., Zhang, J. J., & Huang, X.-Y. (2010). Migrastatin analogues target fascin to block tumour metastasis. *Nature*, *464*(7291), 1062–1066. http://doi.org/10.1038/nature08978

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Cryst (2010). D66, 486-501 [Doi:10.1107/S0907444910007493]*, 1–16. http://doi.org/10.1107/S0907444910007493

Frey, S., & Görlich, D. (2014). A new set of highly efficient, tag-cleaving proteases for purifying recombinant proteins. *Journal of Chromatography A*, *1337*, 95–105. http://doi.org/10.1016/j.chroma.2014.02.029

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, *6*(5), 343–345. http://doi.org/10.1038/nmeth.1318

Hannemann, S., Madrid, R., Stastna, J., Kitzing, T., Gasteier, J., Schonichen, A., et al. (2008). The Diaphanous-related Formin FHOD1 Associates with ROCK1 and Promotes Src-dependent Plasma Membrane Blebbing. *Journal of Biological Chemistry*, *283*(41), 27891–27903. http://doi.org/10.1074/jbc.M801800200

Hildebrand, A., Remmert, M., Biegert, A., & Söding, J. (2009). Fast and accurate automatic structure prediction with HHpred. *Proteins: Structure, Function, and Bioinformatics*, *77*(S9), 128–132. http://doi.org/10.1002/prot.22499

Horn, H. F., Kim, D. I., Wright, G. D., Wong, E. S. M., Stewart, C. L., Burke, B., & Roux, K. J. (2013). A mammalian KASH domain protein coupling meiotic chromosomes to the cytoskeleton. *The Journal of Cell Biology*, *202*(7), 1023–1039. http://doi.org/10.1083/jcb.201304004

Jayo, A., Malboubi, M., Antoku, S., Chang, W., Ortiz-Zapater, E., Groen, C., et al. (2016). Fascin Regulates Nuclear Movement and Deformation in Migrating Cells. *Developmental Cell*, *38*(4), 371–383. http://doi.org/10.1016/j.devcel.2016.07.021

Ketema, M., & Sonnenberg, A. (2011). Nesprin-3: a versatile connector between the nucleus and the cytoskeleton. *Biochemical Society Transactions*, *39*(6), 1719–1724. http://doi.org/10.1042/BST20110669

Ketema, M., Wilhelmsen, K., Kuikman, I., Janssen, H., Hodzic, D., & Sonnenberg, A. (2007). Requirements for the localization of nesprin-3 at the nuclear envelope and its interaction with plectin. *Journal of Cell Science*, *120*(19), 3384–3394. http://doi.org/10.1242/jcs.014191

Kutscheidt, S., Zhu, R., Antoku, S., Luxton, G. W. G., Stagljar, I., Fackler, O. T., & Gundersen, G. G. (2014). FHOD1 interaction with nesprin-2G mediates TAN line formation and nuclear movement. *Nature Cell Biology*, *16*(7), 708–715. http://doi.org/10.1038/ncb2981

Lindeman, R. E., & Pelegri, F. (2012). Localized Products of futile cycle/ lrmp Promote Centrosome-Nucleus Attachment in the Zebrafish Zygote. *Current Biology*, *22*(10), 843–851. http://doi.org/10.1016/j.cub.2012.03.058

Lu, W., Schneider, M., Neumann, S., Jaeger, V.-M., Taranum, S., Munck, M., et al. (2012). Nesprin interchain associations control nuclear size. *Cellular and Molecular Life Sciences*, *69*(20), 3493–3509. http://doi.org/10.1007/s00018-012-1034-1

Luxton, G. W. G., Gomes, E. R., Folker, E. S., Vintinner, E., & Gundersen, G. G. (2010). Linear Arrays of Nuclear Envelope Proteins Harness Retrograde Actin Flow for Nuclear Movement. *Science*, *329*(5994), 956–959.

http://doi.org/10.1126/science.1189072

Luxton, G. W. G., Gomes, E. R., Folker, E. S., Worman, H., & Gundersen, G. G. (2014). TAN lines. *Nucleus*, *2*(3), 173–181. http://doi.org/10.4161/nucl.2.3.16243

Madrid, R., Gasteier, J. E., Bouchet, J., Schröder, S., Geyer, M., Benichou, S., & Fackler, O. T. (2004). Oligomerization of the diaphanous-related formin FHOD1 requires a coiled-coil motif critical for its cytoskeletal and transcriptional activities. *FEBS Letters*, *579*(2), 441–448. http://doi.org/10.1016/j.febslet.2004.12.009

Morimoto, A., Shibuya, H., Zhu, X., Kim, J., Ishiguro, K.-I., Han, M., & Watanabe, Y. (2012). A conserved KASH domain protein associates with telomeres, SUN1, and dynactin during mammalian meiosis. *The Journal of Cell Biology*, *198*(2), 165–172. http://doi.org/10.1083/jcb.201204085

Morgan J. T., Pfeiffer E. T., Thirkill T. W., Kumar P., Peng G., Fridolfsson H. N., Douglas G. C., Starr D. A., & Barakat A. I. (2011). Nesprin-3 regulates endothelial cell morphology, perinuclear cytoskeletal architecture, and flow-induced polarization, *Molecular Biology of the Cell. 22, 4324—4334.* http://doi.org/10.1091/mbc.E11-04-0287)

Otwinoski, Z., & Minor W. (1997). Processing of X-ray Diffraction Data in Oscillation mode. Methods in Enzymology, 276, 307–326.

Petrie, R. J., Harlin, H. M., Korsak, L. I. T., & Yamada, K. M. (2017). Activating the nuclear piston mechanism of 3D migration in tumor cells. *The Journal of Cell Biology*, *216*(1), 93–100. http://doi.org/10.1083/jcb.201605097

Pfisterer, K., Jayo, A., & Parsons, M. (2017). Control of nuclear organization by F-actin binding proteins. *Nucleus*, *8*(2), 126-133. http://doi.org/10.1080/19491034.2016.1267093

Schonichen, A., Mannherz, H. G., Behrmann, E., Mazur, A. J., Kuhn, S., Silvan, U., et al. (2013). FHOD1 is a combined actin filament capping and bundling factor that selectively associates with actin arcs and stress fibers. *Journal of Cell Science*, *126*(8), 1891–1901. http://doi.org/10.1242/jcs.126706

Schulte, A., Rak, A., Pylypenko, O., Ludwig, D., & Geyer, M. (2007). Purification, crystallization and preliminary structural characterization of the N-terminal region of the human formin-homology protein FHOD1. *Acta Cryst (2007). F63, 878-881 [Doi:10.1107/S1744309107043400]*, 1–4. http://doi.org/10.1107/S1744309107043400

Schulte, A., Stolp, B., Schönichen, A., Pylypenko, O., Rak, A., Fackler, O. T., & Geyer, M. (2008). The Human Formin FHOD1 Contains a Bipartite Structure of FH3 and GTPase-Binding Domains Required for Activation. *Structure*, *16*(9), 1313–1323. http://doi.org/10.1016/j.str.2008.06.008

Sosa, B. A., Kutay, U., & Schwartz, T. U. (2013). Structural insights into LINC complexes. *Current Opinion in Structural Biology*, *23*(2), 285–291. http://doi.org/10.1016/j.sbi.2013.03.005

Starr, D. A., & Fridolfsson, H. N. (2010). Interactions Between Nuclei and the Cytoskeleton Are Mediated by SUN-KASH Nuclear-Envelope Bridges. *Annual Review of Cell and Developmental Biology*, *26*(1), 421–444. http://doi.org/10.1146/annurev-cellbio-100109-104037

Fontao, L., Geerts, D., Kuikman, I. Koster, J., Kramer, D., & Sonnenberg, A. (2001). The interaction of plectin with actin: evidence for cross-linking of actin filaments by dimerization of the actin-binding domain of plectin, *Journal of Cell Science*. 114(11) 2065—2076.

Türmer, K., Orbán, J., Gróf, P., & Nyitrai, M. (2015). FASCIN and alpha-actinin can regulate the conformation of actin filaments. *BBA - General Subjects*, *1850*(9),

1855–1861. http://doi.org/10.1016/j.bbagen.2015.05.018

Ungricht, R., & Kutay, U. (2015). ScienceDirect Establishment of NE asymmetry — targeting of membrane proteins to the inner nuclear membrane. *Current Opinion in Cell Biology*, *34*, 135–141. http://doi.org/10.1016/j.ceb.2015.04.005

Zhang, Q., Ragnauth, C., Greener, M. J., Shanahan, C. M., & Roberts, R. G. (2002). The Nesprins Are Giant Actin-Binding Proteins, Orthologous to *Drosophila melanogaster* Muscle Protein MSP-300. *Genomics*, *80*(5), 473–481. http://doi.org/10.1006/geno.2002.6859

Zhang, T., Ling, X.-L., Hou, X.-M., & Zhao, D. (2015). Clinicopathological significance of fascin-1 expression in patients with non-small cell lung cancer. *OncoTargets and Therapy*, 1589–7. http://doi.org/10.2147/OTT.S84308

Conclusion and Future Direction

Summary

The work presented here has led to the high-resolution structures of *H. sapiens* SUN2 in complex with KASH3, KASH4, and KASH5. In these structures we identify an alternate mode of binding between SUN2 and KASH than what has been reported previously (Sosa, 2012). From these structures we conclude that the identity of residue at position -11 may determine which binding mode any KASH peptide will prefer when bound to SUN2. The binding site of each KASH peptide is unaffected by the status of the neighboring binding sites, using this premise we show that a single SUN2 trimer can bind to multiple different KASH peptides simultaneously adding an unexpected layer of complexity to LINC complexes. KASH1 and KASH2 are able to covalently bind to SUN2 via a disulfide bond, but KASH3, KASH4, KASH5, and KASH6 cannot form that covalent bond, when bound in the alternate binding mode. We propose that while SUN2 binding to KASH is promiscuous, forces generated at the cytoplasm by cytoskeletal components may selectively disassemble non-covalent SUN2-KASH pairs (Jahed, et al., 2015). This mechanism would allow SUN2 to preferably bind KASH1 or KASH2 when SUN2 LINC complexes must withstand large forces.

We have extensively characterized the oligomeric state of the coiled-coil domain of SUN2 in solution, we have crystallized a major portion of the coiled-coil, and have collected high resolution datasets for a segment of the coiled-coil required for trimerization. It is clear that the segment we have crystallized is important for function of LINC complexes because the coiled-coil mediates trimerization of SUN2, and SUN2

trimerization is a pre-requisite for KASH binding (Sosa, 2012; Demirciouglu, 2015; Nie, et al., 2016). Until we solve the structure of the SUN2 coiled-coil domain, details of how trimerization is mediated remain speculative.

We have also performed preliminary structural characterization of the cytoplasmic domain of KASH proteins. We focus on Nesprin-2 (KASH2) for its role in TAN lines, and Nesprin-3α (KASH3) which has been shown to regulate nuclear size (Luxton, 2010; Kutscheidt, 2014; Jayo, 2016; Lu, 2012) through interactions with Nesprin-1/2. First, we present our purification and crystallization of Nesprin-2 in complex with FHOD1 which will provide a significant step forward in our understanding of how TAN lines are stabilized. Additionally, we present our preliminary work with Nesprin-3α. Here, we establish a generic purification method that yields large quantities of pure Nesprin-3α. We next identify stable fragments of Nesprin-3α which we proceed to crystallize.

Future directions

SUN-KASH complexes

With our current repertoire of SUN2-KASH structures we have a good understanding of the interactions between SUN2 and the core of KASH peptides. KASH1 and KASH2 are known to form a disulfide bond with SUN2 (Sosa, 2012). KASH3 and KASH4, have the conserved cysteine at position -23 but do not form a disulfide bond with SUN2, raising the question of why this cysteine is conserved in these KASH peptides. From the work presented here, we now know that a proline at position -11 is required for the formation

of a disulfide with SUN2. If a proline at position -11 is required for KASH peptides to form a disulfide with SUN1, however, is unclear. Because SUN1 is known to interact with both KASH3 and KASH4, it is tempting to speculate if these KASH peptides can indeed covalently bind to SUN1 (Ketema, et al., 2007; Horn, et al 2013a; Horn, et al., 2013b). Structural and biochemical characterization of SUN1 in complex with KASH peptides is required to determine the role that proline -11 plays when bound to SUN1 as opposed to SUN2.

Coiled-coil domain of SUN2

Currently we have high resolution X-ray data of the coiled-coil domain of SUN2, both native and anomalous datasets. Despite having a highly homologous partial protein model that we can use for molecular replacement, and being able to place heavy atoms in our maps, we cannot generate a reasonable electron density map. We will focus on finding solutions for the coiled-coil that generate clear electron density and a buildable map. If a computational solution is not achieved in a timely fashion, we will focus on crystallizing the coiled-coil in complex with a nanobody. The main advantage of using a nanobody is that the packing of the coiled-coil in the unit cell will be different from our current crystals, likely bypassing the technical problems we are currently trying to overcome.

Cytoplasmic domains of LINC complexes

The cytoplasmic domain of nesprins (KASH proteins) is predominantly composed of spectrin repeats (SR), that form the highly modular structural core of nesprins. Despite the repetitive nature of these modules certain SR can specifically interact with other proteins. Here we focused on interactions between Nesprin-2 and the actin binding proteins FHOD1 and Fascin-1. Our immediate objective is to obtain high resolution anomalous data of our FHOD1 and Nesprin-2 complex crystals in order to obtain accurate experimental phases to solve the structure of the complex. This structure will provide the first structural glimpse into a SR in complex with another protein, and will provide a molecular basis for how TAN lines are formed. To fully understand how TAN lines are stabilized it is critical that all bipartite interactions are well characterized. To this end we have begun work on the structural and biochemical characterization between Fascin-1 and Nesprin-2. Currently we are focused on stabilizing the complex for further structural characterization. Fascin-1 is known to play a role in nuclear deformation, which is required for cell migration during metastasis and expression of Fascin-1 is correlated to poor prognosis in cancer patients (Wang, 2016; Min, 2015). Therefore, a detailed structural understanding of how Fascin-1 interacts with cytoskeletal elements will provide a basis for therapeutic drug design.

A long term goal of this project is to determine how Nesprin-2 interacts with actin. This occurs through two mechanisms. First through the N-terminal actin binding (ABD) calponin homology (CH1 and CH2) domains of Nesprin-2, and second through interactions mediated by FHOD1 and Fascin-1. It is likely that this pursuit will require the use of cryo-electron microscopy given the size of Nesprin-2 (~800 kDa) and that these

interactions will probably require using actin filaments as opposed to globular actin. In addition to this, by using nanobodies generated towards Nesprin-2 (ABD) or the SR which bind to FHOD1 and Fascin-1 we can finely probe the molecular requirements for assembling and stabilizing TAN lines.

References

Demircioglu, F. E., Cruz, V. E., & Schwartz, T. U. (2016). Purification and Structural Analysis of SUN and KASH Domain Proteins. *Methods in Enzymology*, Volume 569, 63-78. http://doi.org/10.1016/bs.mie.2015.08.011

Horn, H. F., Brownstein, Z., Lenz, D. R., Shivatzki, S., Dror, A. A., Dagan-Rosenfeld, O., et al. (2013a). The LINC complex is essential for hearing. *Journal of Clinical Investigation*, 1–11. http://doi.org/10.1172/JCI66911

Horn, H. F., Kim, D. I., Wright, G. D., Wong, E. S. M., Stewart, C. L., Burke, B., & Roux, K. J. (2013b). A mammalian KASH domain protein coupling meiotic chromosomes to the cytoskeleton. *The Journal of Cell Biology*, *202*(7), 1023–1039. http://doi.org/10.1083/jcb.201304004

Jahed, Z., Shams, H., & Mofrad, M. R. K. (2015). A Disulfide Bond Is Required for the Transmission of Forces through SUN-KASH Complexes. *Biophysj*, *109*(3), 501–509. http://doi.org/10.1016/j.bpj.2015.06.057

Jayo, A., Malboubi, M., Antoku, S., Chang, W., Ortiz-Zapater, E., Groen, C., et al. (2016). Fascin Regulates Nuclear Movement and Deformation in Migrating Cells. *Developmental Cell*, *38*(4), 371–383. http://doi.org/10.1016/j.devcel.2016.07.021

Ketema, M., Wilhelmsen, K., Kuikman, I., Janssen, H., Hodzic, D., & Sonnenberg, A. (2007). Requirements for the localization of nesprin-3 at the nuclear envelope and its interaction with plectin. *Journal of Cell Science*, *120*(19), 3384–3394. http://doi.org/10.1242/jcs.014191

Kutscheidt, S., Zhu, R., Antoku, S., Luxton, G. W. G., Stagljar, I., Fackler, O. T., & Gundersen, G. G. (2014). FHOD1 interaction with nesprin-2G mediates TAN line formation and nuclear movement. *Nature Cell Biology*, *16*(7), 708–715. http://doi.org/10.1038/ncb2981

Lu, W., Schneider, M., Neumann, S., Jaeger, V.-M., Taranum, S., Munck, M., et al. (2012). Nesprin interchain associations control nuclear size. *Cellular and Molecular Life Sciences*, *69*(20), 3493–3509. http://doi.org/10.1007/s00018-012-1034-1

Luxton, G. W. G., Gomes, E. R., Folker, E. S., Vintinner, E., & Gundersen, G. G. (2010). Linear Arrays of Nuclear Envelope Proteins Harness Retrograde Actin Flow for Nuclear Movement. *Science*, *329*(5994), 956–959.

Min, K. W., Chae, S., Kim, D. H., Do, S. I., Kim, K., Lee, H., et al. (2015). Fascin expression predicts an aggressive clinical course in patients with advanced breast cancer. *Oncology Letters*, 1–10. http://doi.org/10.3892/ol.2015.3191

Nie, S., Ke, H., Gao, F., Ren, J., Wang, M., Huo, L., et al. (2016). Coiled-Coil Domains of SUN Proteins as Intrinsic Dynamic Regulators. *Structure/Folding and Design*, *24*(1), 80–91. http://doi.org/10.1016/j.str.2015.10.024

Sosa, B. A., Rothballer, A., Kutay, U., & Schwartz, T. U. (2012). LINC Complexes Form by Binding of Three KASH Peptides to Domain Interfaces of Trimeric SUN Proteins. *Cell*, *149*(5), 1035–1047. http://doi.org/10.1016/j.cell.2012.03.046

Wang, C.-Q., Tang, C.-H., Chang, H.-T., Li, X.-N., Zhao, Y.-M., Su, C.-M., et al. (2016). Fascin-1 as a novel diagnostic marker of triple-negative breast cancer. *Cancer Medicine*, *5*(8), 1983–1988. http://doi.org/10.1002/cam4.746

Purification and Structural Analysis of SUN and KASH Domain Proteins

This section is adapted from Demircioglu FE, Cruz VE, Schwartz TU. "Purification and Structural Analysis of SUN and KASH Domain Proteins" Methods in Enzymology 2016; 569:63-78. PMCID: 26778553

Abstract

Molecular tethers span the nuclear envelope to mechanically connect the cytoskeleton and nucleoskeleton. These bridge-like tethers, termed LInkers of Nucleoskeleton and Cytoskeleton (LINC) complexes, consist of SUN proteins at the inner nuclear membrane and KASH proteins at the outer nuclear membrane. LINC complexes are central to a variety of cell activities including nuclear positioning and mechanotransduction, and LINC-related abnormalities are associated with a spectrum of tissue-specific diseases, termed laminopathies or envelopathies. Protocols used to study the biochemical and structural characteristics of core elements of SUN-KASH complexes are described here to facilitate further studies in this new field of cell biology.

Introduction

The nuclear envelope (NE) physically separates the nucleus from the cytoplasm, generating two distinct compartments. Molecular exchange between the nucleoplasm and cytoplasm is mediated by nuclear pore complexes, which act as selective permeability barriers. Mechanical communication between the nucleus and cytoplasm involves specific tethers, termed **LI**nkers of **N**ucleoskeleton and **C**ytoskeleton (LINC) complexes, that span the NE. LINC complexes are formed by a family of KASH (**K**larsicht, **A**NC-1, and **S**yne **H**omology) proteins embedded in the outer nuclear membrane (ONM) that interact within the NE lumen with SUN (**S**ad1 and **UN**C-84) proteins, which span the inner nuclear membrane (INM). SUN and KASH proteins each project from the NE and directly bind to components of the nucleoskeleton and the

cytoskeleton, respectively. These mechanical connections are critically important in a wide range of activities such as nuclear migration and anchorage, meiotic chromosome movements, the centrosome-nucleus connection, signal transduction and DNA repair (Burke & Roux, 2009; Chang, Worman, & Gundersen, 2015; Luxton & Starr, 2014; Rothballer & Kutay, 2013; Starr & Fridolfsson, 2010).

KASH proteins are tail-anchored, single-span transmembrane proteins, mostly found at the ONM. The C-terminal "KASH motif" comprises the transmembrane helix and the adjacent luminal segment, which consists of 8-30 residues, depending on the specific nesprin gene and species (Starr & Han, 2002). Vertebrate KASH proteins are often called nesprins (NE spectrin repeat proteins), since their cytoplasmic portions typically contain numerous spectrin repeats. The cytoplasmic extensions vary greatly in size due to alternative splicing and transcription initiation of multiple nesprin genes (Zhang et al., 2001). The two longest ('giant'; 0.8-1.0 MDa) nesprin isoforms, Nesprin-1G and Nesprin-2G, each bind to actin filaments via calponin homology (CH) domains at their N terminus. Coupling of actin filaments to LINC complexes has been best visualized in migrating fibroblasts, where SUN2-Nesprin-2G complexes assemble into linear arrays at the NE and form so-called TAN (transmembrane actin-associated nuclear) lines. Formation of TAN lines is instrumental in moving the nucleus rearward by coupling to retrogradely-moving actin cables (Luxton, Gomes, Folker, Vintinner, & Gundersen, 2010; Luxton, Gomes, Folker, Worman, & Gundersen, 2011). The much shorter protein Nesprin-3α binds plectin, which in turn binds to cytoplasmic intermediate filaments and/or actin. Since Nesprin-3α also binds to the CH domains of Nesprin-1G

and Nesprin-2G, a nesprin scaffold is proposed to form around the nucleus that might play a role in regulating nuclear size (Lu et al., 2012). Nesprin-4 interacts with microtubules through kinesin-1, and is proposed to function specifically in ear development and hearing (Horn, Brownstein, et al., 2013a). Yet another tissue-specific KASH protein, Nesprin-5, binds to microtubules through dynein and functions during meiotic chromosome pairing in germ cells (Horn, Kim, et al., 2013b). Finally, a recently recognized sixth KASH protein in zebrafish, lymphoid restricted membrane protein (LRMP), is involved in pronuclear congression during fertilization (Lindeman & Pelegri, 2012).

Similarly most organisms also encode several SUN homologs. Of the five known mammalian SUN proteins, SUN1 and SUN2 are widely expressed (Crisp et al., 2006; Padmakumar et al., 2005), whereas SUN3, SUN4, and SUN5 are expressed during spermatogenesis in testis (Göb, Schmitt, Benavente, & Alsheimer, 2010). SUN proteins have at least one transmembrane helix, which typically anchors them in the INM. The C-terminal ~20 kDa SUN domain, preceded by a predicted coiled-coil segment of variable length, are both located in the NE lumenal space. The N terminus of SUN proteins extends into the nucleoplasm, and binds lamins (nuclear intermediate filament proteins). Mammalian SUN proteins are known to bind to A-type lamins, while interaction with B-type lamins is relatively weak (Crisp et al., 2006; Haque et al., 2006). Although lamina attachment restricts diffusion of the SUNs (Ostlund et al., 2009), lamins do not seem to be the only factors anchoring LINC complexes. Indeed, both SUN1 and SUN2 are properly localized in the absence of A- and B-type lamins (Crisp et al., 2006; Haque et

al., 2006; Padmakumar et al., 2005). Lamin associated proteins such as emerin and SAMP1 likely help anchor the LINC complexes (Borrego-Pinto et al., 2012; Chang, Folker, Worman, & Gundersen, 2013), and the intricate interplay among these proteins is not yet understood. Mutations in A-type lamins, emerin, SUNs and nesprins can each disrupt nucleocytoskeletal coupling and are also genetically linked to laminopathies such as skeletal and/or cardiac muscular dystrophies, lipodystrophy, dysplasia or segmental progeroid ('accelerated aging') disorders (Worman, 2012). The striated muscle disease EDMD (Emery-Dreifuss muscular dystrophy), and the premature aging syndrome HGPS (Hutchinson-Gilford progeria syndrome) are examples where perturbed functioning of LINC complexes and associated factors contributes to pathology (Bione et al., 1994; Bonne et al., 1999; Zhang et al., 2007) (Puckelwartz et al., 2009) (Haque et al., 2010) (Chen et al., 2012). Further structural and biochemical characterization of LINC complexes is needed to understand the molecular basis of these disorders.

Recent crystallographic studies established the structure of the core element of the LINC complex, namely the SUN-KASH interaction (Wang et al, 2012; Zhou et al, 2012). Human SUN2 proteins form a triple-stranded coiled-coil stalk to generate a trimeric structure that positions the adjacent, C-terminal β-sandwich-shaped SUN domains to form a globular trefoil. The helical stalk assumes an unusual, right-handed supercoil that positions the SUN domains in a KASH-binding competent state. Three KASH peptides are bound at the three interfaces between adjacent SUN protomers in the trefoil, immediately explaining why monomeric SUN does not bind a KASH peptide.

Altogether, the core of the LINC complex is a heterohexameric SUN_3-$KASH_3$ complex. The carboxyl group of the terminal residue in the KASH peptide is specifically recognized by the SUN domain, thereby explaining why KASH peptides are always located at the C terminus of a protein. A cysteine residue at the N terminus of the KASH peptide can form a disulfide bridge with a conserved cysteine on the SUN domain, presumably enhancing the mechanical strength of the complex (Sosa, Rothballer, Kutay, & Schwartz, 2012).

Although these SUN2-KASH1/2 structures revealed crucial information about LINC complex assembly, much more work is needed to fully understand the SUN-KASH interactome. Specific methods for the purification, biochemical analysis, and structure determination of apo-SUN2, and SUN2-KASH complexes, are detailed below. In addition to SUN2-KASH1/2 complexes, which have been crystallized, we also describe strategies for purifying SUN2 complexes with KASH3, KASH4, KASH5 or KASH6.

Purification of SUN proteins and SUN-KASH complexes

Construct design

We use the pETDuet-1 bacterial expression system (EMD Biosciences) to produce SUN and KASH domain proteins from two different multiple cloning sites (MCS). The cDNA encoding each SUN domain protein is cloned into the first MCS (MCS1), and the cDNA encoding the KASH domain protein is cloned into the second MCS (MCS2), either as single open reading frames for isolation of the apo proteins or in tandem for isolation of SUN-KASH complexes. When cloned together, the dual cassette expression enables

SUN-KASH interaction already in the bacterial cell, facilitating isolation of stoichiometric complexes.

We generated N-terminally 6xHis-tagged SUN2 fragments for expression from MCS1. These fragments contain lumenal portions of SUN2. Since SUN2 fragments that lack significant portions of the predicted coiled-coil regions might become monomeric in solution, prohibiting KASH binding, short SUN2 fragments are fused to a coiled-coil fragment of engineered tri-GCN4 (Ciani et al., 2010) at their N termini to restore KASH-binding competence (Fig. 1A). The crystallized part of SUN2 (residues 522-717) is such an example (Sosa et al., 2012). This engineering is not required for longer SUN2 constructs with a native extended coiled-coil stalk, which form stable trimers in solution. A cleavage site for 3C protease is inserted near the N terminus of each SUN2 fragment to facilitate removal of the fusion tag after purification.

For purification of SUN2-KASH complexes, we clone lumenal portions of KASH motifs into the MCS2 site of the pETDuet-1 vector (Fig. 1B). KASH1/2/3/4 peptides can be attached to 3C-cleavable maltose binding protein (MBP) tags (di Guan, Li, Riggs, & Inouye, 1988) at their N termini. MBP-tagging helps in various ways. First, it typically yields super-stoichiometric expression of the MBP-KASH fusion protein compared to SUN2, which enables the isolation of stoichiometric SUN2-KASH complexes in large amounts. Second, the MBP moiety provides an orthogonal affinity-tag after Ni^{2+}-pulldown to isolate stoichiometric complexes in high purity (see Section 2.3).

We faced problems removing the fusion tags during the purification of SUN2-KASH complexes. To increase the efficiency of tag removal, we introduced flexible

Gly/Ser-rich linkers on either side of the 3C-cleavage sites (Fig. 1). This linker strategy proved useful for cutting off the 6xHis-triGCN4 tags, but did not improve the removal of MBP tags. As explained in Section 2.3, this problem was solved in some cases by cleaving the MBP tag in multiple steps, followed by chromatography. For example, the SUN2-KASH1/2 crystals were obtained from MBP-tagged complexes (Sosa et al., 2012). However, isolation of the other SUN-KASH complexes (SUN2-KASH3/4/5/6) was more problematic. For these complexes we tested different tags including superfolder (sf) GFP (Pédelacq, Cabantous, Tran, Terwilliger, & Waldo, 2006), thioredoxin (LaVallie, Lu, Diblasio-Smith, Collins-Racie, & McCoy, 2000) and GB1 (Huth et al., 1997). The GB1 tag significantly enhanced cleavage efficiency, and was used to purify SUN2-KASH5/6 complexes (Fig. 1B). Advantages of using different tags during purification will be explained in detail in Section 2.3.

Purification of human apo-SUN2

We purified human apo-SUN2 (residues 335-717) and apo-SUN2 (residues 522-717) in *E. coli* strain LOBSTR-BL21(DE3)-RIL (Kerafast, Inc, Boston MA) (Andersen, Leksa, & Schwartz, 2013) for increased purity, using the following protocol.

1. Inoculate 3-6 ml Lysogeny Broth (LB) with a single LOBSTR-BL21(DE3)-RIL colony that was heat-shock transformed with the SUN2-expressing plasmid. Include ampicillin (100 µg/ml) and chloramphenicol (34 µg/ml) to select for the pETDuet-1 derived SUN2-expressing plasmid and the RIL plasmid (Agilent Technologies), respectively. Grow this starter culture overnight at 30°C.

2. The next morning, inoculate 1 L of LB medium containing 0.4% (w/v) glucose, ampicillin (100 µg/ml) and chloramphenicol (34 µg/ml) with the overnight culture. Grow these bacteria in a 2-L baffled Erlenmeyer flask at 37°C in a shaker to an OD_{600} of 0.6-0.8, then transfer to 18°C and incubate 20 minutes longer. Then add IPTG (0.2 mM final) to induce protein expression, and shake overnight at 18°C.

3. The next morning, record the OD_{600} (usually between 6-8) and then harvest. Pellet cells by centrifugation at 6,000 rpm for 6 min (e.g., Sorvall SLA-3000 rotor). Resuspend the bacterial pellet (20 ml lysis buffer per 1000 OD_{600}) in ice-cold lysis buffer (50 mM potassium phosphate pH 8.0, 400 mM NaCl, and 40 mM imidazole). Note that lysis and all subsequent steps should be done at 4°C with pre-chilled solutions.

4. Resuspend the bacteria homogenously to obtain a clump-free cell suspension, then process using a cell homogenizer (Constant Systems) at 25 kpsi. Mix the collected lysate immediately with 0.1 M PMSF (50 µl per 10 ml lysate) and add 250 units of TurboNuclease (Eton Bioscience).

5. Centrifuge the lysate at 9,500 rpm for 25 min (e.g., Sorvall SLA-600TC rotor), and recover the supernatant. Mix the supernatant with Ni^{2+} Sepharose 6 Fast Flow (GE Healthcare) slurry equilibrated with lysis buffer. Use approximately 1 ml Ni^{2+} resin per 1000 OD_{600} of cells.

6. Gently stir the mixture for 30 min, collect the Ni^{2+}-resin in a 50 ml conical tube via several spins using a tabletop centrifuge, and then batch-wash the Ni^{2+}-resin three times with 40 ml lysis buffer. Pour the Ni^{2+}-Sepharose slurry into a disposable Pierce

column (Thermo Scientific), and wash with 6x resin bed volumes of lysis buffer via gravity flow.

7. Once the column is drained, elute proteins using a 6x resin bed volume of elution buffer (10 mM Tris/HCl pH 8.0, 150 mM NaCl and 250 mM imidazole).

8. Concentrate that eluted protein to a final volume of ~10 ml using a centrifugal concentrator, and then purify by size exclusion chromatography on a HiLoad 26/60 Superdex S200 column (GE Healthcare) in a buffer containing 10 mM Tris/HCl pH 8.0 and 150 mM NaCl. Pool the peak corresponding to His_6-tagged SUN2 is pooled, and mix with 3C protease at an enzyme:protein ratio of 1:50 (w/w). Overnight incubation with 3C protease is generally sufficient to remove the fusion tags, but this should be verified by SDS-PAGE analysis of a small aliquot.

9. Concentrate the cleaved SUN2 protein and purify again by size exclusion chromatography on a HiLoad 26/60 Superdex S200 column in 10 mM Tris/HCl pH 8.0, and 150 mM NaCl. Note that apo-SUN2 (residues 335-717) and apo-SUN2 (residues 522-717) elute differently without their fusion tags. Apo-SUN2 (residues 335-717) includes the entire predicted coiled-coil region and elutes as a homotrimer, as confirmed by analytical ultracentrifugation (Sosa et al., 2012). By contrast, apo-SUN2 (residues 522-717) behaves as a monomer after the His_6-tri-GCN4 tag is removed, although this behavior is buffer dependent (see Section 3.2). These observations support the notion that the coiled-coil region (residues 335-540) adjacent to the SUN domain (residues 540-717) helps stabilize the SUN homotrimer and, hence, the KASH-binding-competent oligomeric state.

Purification of human SUN2-KASH complexes

Protocols for purifying SUN2 (residues 522-717)-KASH1-6 complexes are described here and shown schematically (Fig. 2).

1. Steps 1-7 described in section 2.2 remain essentially the same, with one important difference: we now include 1 mM KCl in the Ni^{2+} elution buffer (and all subsequent buffers) since a K^+ ion is likely to be coordinated in the cation loop of SUN2-KASH complexes (Sosa et al., 2012).

2. Similar to apo-SUN2, purify SUN2-KASH Ni^{2+}eluates by size exclusion chromatography using a Superdex S200 column equilibrated with 10 mM Tris/HCl pH 8.0, 150 mM NaCl, and 1 mM KCl. This step helps remove aggregates, unbound SUN2, and KASH. Alternatively, the Ni^{2+} eluate can be subjected to a second purification on amylose resin if KASH is MBP-tagged. In this scenario, imidazole is removed by dialyzing the Ni^{2+} eluate against 10 mM Tris/HCl pH 8.0, 150 mM NaCl and 1 mM KCl prior to binding to amylose, and the SUN2-KASH complexes are then eluted in the presence of 10 mM maltose.

3. Mix the resulting SUN2-KASH complexes with 3C protease at an enzyme:substrate ratio of 1:25-1:50 (w/w). Retain an aliquot of uncut sample for SDS-PAGE analysis. After incubating 16 hours, resolve a small aliquot of SUN-KASH complexes by 15% SDS-PAGE to estimate cleavage efficiency. If necessary, add fresh 3C protease and incubate for at least 8 hours.

4. (a) The above procedure yields MBP-tagged SUN-KASH complexes that are often incompletely cleaved. Cleaved MBP can also form soluble aggregates of ill-defined size

that partially co-elute with SUN-KASH complexes during gel filtration. To overcome both problems, and improve SUN-KASH complex homogeneity, we cleave the MBP tag in two steps (Fig. 2A). First, load the partially cleaved SUN-KASH complex from step 2.3.3 on a Superdex S200 column (in standard gel filtration buffer, e.g., 10 mM Tris/HCl pH 8.0, 150 mM NaCl, and 1 mM KCl) and purify. To avoid protein precipitation before column loading, the protein concentration should not exceed 1-2 mg/ml at this step. The partially cleaved SUN-KASH complex elutes in a fairly non-homogeneous peak during gel filtration, and is collected with contaminants. Second, pool the SUN-KASH complex-containing fractions, and cleave again with 3C protease at an enzyme:substrate ratio of approximately 1:50 (w/w). This second cleavage step should completely remove the fusion tags from SUN2-KASH1/2/3/4 complexes (verify by analytical SDS-PAGE). Re-concentrate these purified complexes, and then load onto a HiLoad 16/60 Superdex S200 column and purify using standard gel filtration buffer. SUN-KASH complexes elute in a single peak, and are now suitable for crystallization studies or biochemical assays.
(b) The above strategies failed to yield pure and homogeneous complexes in the case of SUN2-KASH5/6. We tested alternatives to MBP-tags that could sustain super-stoichiometric expression of KASH5/6 peptides, relative to SUN, and could be removed more efficiently by 3C protease in step 2.3.3. Two candidates, sfGFP and thioredoxin, both reduced KASH-fusion protein expression to sub-stoichiometric levels compared to SUN (data not shown). However, the GB1 tag met both criteria, and was completely removed by proteolysis as observed by SDS-PAGE analysis. In a final step, the resulting complexes are concentrated and purified via gel filtration (Fig. 2B). We have

not tried the GB1 tag with other SUN-KASH complexes, but suspect it may have consistent advantages over the MBP tag.

Structural analysis of human SUN2 and SUN2-KASH1/2 complexes

Crystallization and structure determination

We deposited three crystal structures in the Protein Data Bank representing human apo-SUN2 (PDB ID: 4DXT), and SUN2-KASH1/2 complexes (PDB ID: 4DXR/ 4DXS) (Sosa et al., 2012). Independently, apo-SUN2 (PDB ID: 3UNP) and SUN2-KASH2 (PDB ID: 4FI9) structures were solved by the Wang and Zhou labs (Wang et al, 2012; Zhou et al, 2012). Crystallization in all labs was performed by the hanging-drop vapor diffusion method in distinct experiment drop compositions. Apo-SUN2 crystals have grown in 16% (w/v) polyethylene glycol (PEG) 3350 and 200 mM potassium thiocyanate at 18°C in our lab, whereas the Wang group crystallized it in 100 mM imidazole, 1 M sodium acetate pH 6.5 and 10 mM YCl_3 at 4°C. The resulting structures are overall very similar, except for a varying conformation of the unstructured KASH 'lid' (SUN2 residues 567-587) in the apo-form.

The KASH lid of the SUN domain becomes ordered, and adopts a β-hairpin form, only upon binding to KASH. Reflecting the high similarity between KASH1 and KASH2 in length and amino acid composition, the lid adopts an identical conformation in SUN2-KASH1/2 complexes. Although KASH1/2 peptides are not involved in crystal-packing contacts, SUN2-KASH1 and SUN2-KASH2 complexes crystallize in different conditions.

In our lab, the SUN2-KASH1 complex was crystallized in 100 mM HEPES pH 7.4, 7% (w/v) PEG 4000, 10% 1,6-hexanediol and 0.25% n-decyl-b-D-maltoside (DM), whereas SUN2-KASH2 complex crystals were grown in 100 mM HEPES pH 7.5, 200 mM ammonium acetate, 25% 2-propanol and 0.3% DM. The Zhou lab, on the other hand, crystallized the SUN2-KASH2 complex in 50 mM $MgCl_2$, 100 mM HEPES pH 7.5, 6% (w/v) polyethylene glycol monomethyl ether 5000, and 19.5 mM methyl-6-O-(N-heptylcarbamoyl)-a-D-glucopyranoside (HECAMEG).

Interestingly, apo-SUN2 as well as SUN2-KASH complexes, pack in rhombohedral crystals such that a trefoil-to-trefoil assembly occurs between neighboring SUN2 homotrimers, with the coiled-coil stalks pointing away in opposite directions. The major difference between these crystals is that the distance between neighboring apo-SUN2 trimers is much smaller than that of the SUN2-KASH1/2 complexes, due to the conformational change in the KASH-lid upon binding to KASH. Therefore, apo-SUN2 crystals have a lower solvent content, which might explain why they tend to diffract to higher resolution. Because the KASH peptide is not directly involved in crystal packing, there is a good chance that other SUN2-KASH complexes can be structurally characterized under similar crystallization conditions.

In vitro binding experiments

Interactions between SUN and KASH proteins have been studied using *in vitro* binding protocols established for apo-SUN2 and KASH2. For these studies apo-SUN2 (residues 522-717) and KASH2 (residues 6863-6885) are each purified separately. Apo-SUN2 is

purified as described in section 2.2. We purify KASH2 as a fusion to the C-terminus of 6xHis-sfGFP; this 6xHis-sfGFP-KASH2 polypeptide ('sfGFP-KASH2') is first purified by Ni^{2+}-affinity as described for apo-SUN constructs. The Ni^{2+}-eluate is further purified by gel filtration using a Superdex S75 column equilibrated into 10 mM Tris/HCl pH 8.0 and 150 mM NaCl. sfGFP-KASH2 purified in this manner yields two approximately equal-intensity bands on SDS-PAGE, with a small size difference. Only the larger species binds SUN, suggesting the truncation is C-terminal and eliminates (part of) the KASH-peptide. Since these two forms of sfGFP-KASH2 are present at a ratio of about 1:1 (by SDS-PAGE), for qualitative binding assays we use a molar stoichiometry of 2:1 (sfGFP-KASH2 to apo-SUN2) to ensure a roughly 1:1 ratio of apo-SUN2 and binding-competent sfGFP-KASH2.

Binding is measured by analytical gel filtration on a Superdex S200 HR10/300 column (Fig. 3). Prior to injecting the sample, apo-SUN2, and apo-SUN2 incubated with sfGFP-KASH2, are dialyzed into the gel filtration buffer overnight. For each gel filtration run, a total volume of 500 μl is loaded onto the column; each sample contains 50 μM SUN2 and 100 μM sfGFP-KASH2 (to achieve a 1:1 ratio of SUN2 to binding-competent undegraded sfGFP-KASH). In the chromatogram, the sfGFP-KASH2 fusion protein is monitored by absorbance at 488 nm, and total protein absorption is measured at 280 nm.

The results of such binding experiments are shown in Figure 3. When apo-SUN2 (residues 522-717) is run in buffer 1 (20 mM HEPES/NaOH pH 8.0, 100 mM KCl), it elutes primarily as a trimer ('S3'; Fig. 3A); a small fraction elutes as a monomer ('S1';

Fig. 3A). Buffer 1 was previously used for *in vitro* SUN-KASH binding assays, however the effect of monomer-trimer equilibrium on KASH binding was not discussed (Zhou et al, 2012). In contrast, apo-SUN2 elutes as a monomer when using buffer 2 (10 mM Tris/HCl pH 8.0, 150 mM NaCl, 1 mM KCl) (Fig. 3B). SUN trimerization is a prerequisite for KASH binding, as shown by gel filtration analysis of apo-SUN2 / sfGFP-KASH2 mixtures. Using buffer 1, we obtain a stoichiometric SUN2 / sfGFP-KASH2 complex, and excess sfGFP-KASH2 elutes as a separate peak ('K'; Fig. 3C). Using buffer 2, only a fraction of SUN2 elutes as an assembled complex with sfGFP-KASH2, whereas the majority remains unbound (Fig. 3D). Thus, buffer conditions play an important role in SUN-KASH binding experiments. This buffer dependence is particularly critical for SUN constructs that lack portions of the coiled-coil element.

Conclusions and pitfalls

Our biochemical understanding of SUN-KASH assemblies is still incomplete, and many interesting questions remain open. The buffer-dependence of SUN proteins *in vitro* is not yet understood; indeed the mechanisms of LINC complex assembly and disassembly may be strongly influenced by the unique microenvironment of the NE/ER lumen, which current *in vitro* conditions do not reproduce. The specificity and strength of SUN-KASH interactions are relatively unexamined, yet further analysis of SUN-KASH proteins from different species can potentially help resolve this problem. Finally, apart from the SUN-KASH core, structural knowledge about most other regions of LINC

complexes is limited— an open area of research with fascinating implications for understanding the nuclear envelope and mechanisms of human laminopathy diseases.

Acknowledgements

We thank Brian A. Sosa for developing the initial protocols for purification and crystallization of human SUN2-KASH1/2 complexes. This work was supported by a grant from the NIH (1RO1-AR065484).

Figures

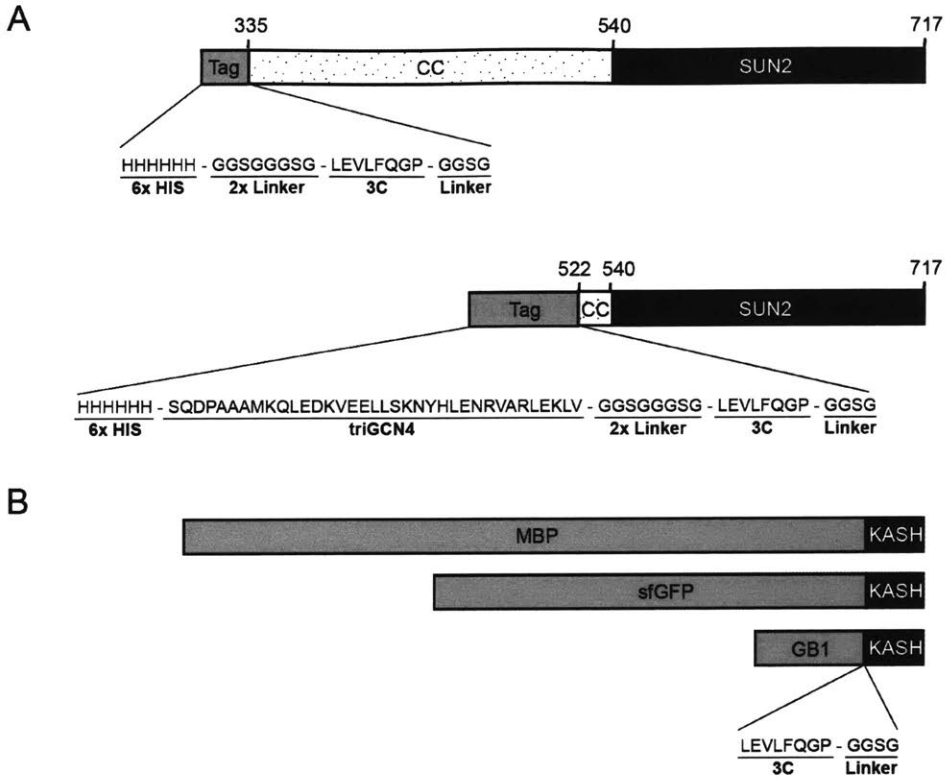


Figure 1:

Schematic drawing of the expression constructs for human SUN2 (**A**), and KASH (**B**). Each SUN2 fragment includes the SUN domain (residues 540-717) and preceding lumenal segments of different length, predicted to form coiled-coils (CC). The SUN2 (residue 335-717) fragment is N-terminally fused to a 3C-cleavable 6xHis tag, whereas a 6xHis:triGCN4 tag is used for a shorter SUN2 fragment (residues 522-717). KASH motifs are C-terminally attached to 3C-cleavable MBP, superfolder GFP (sfGFP) or GB1 tags. Short flexible linkers are incorporated around the 3C cleavage sites to enhance removal of the fusion tags.

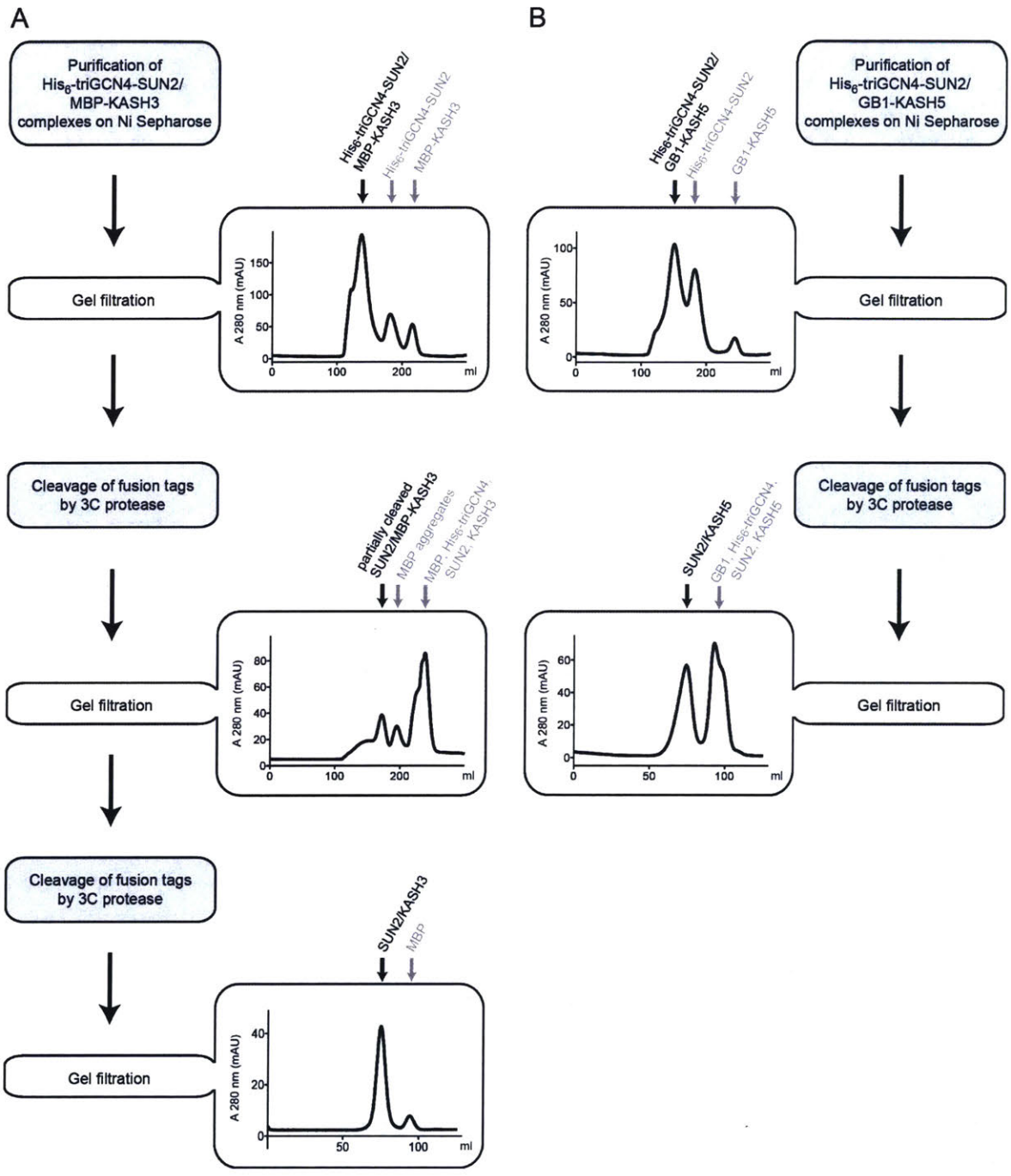


Figure 2:

Purification schemes for SUN2-KASH complexes. **(A)** Purification of the SUN2-KASH3 complex is illustrated to exemplify the strategy used for SUN2 / MBP-KASH constructs.

A representative gel filtration elution profile at each step is shown, and the proteins eluted under each peak are indicated. **(B)** Purification of the SUN2-KASH5 complex is illustrated to demonstrate the strategy used for SUN2 / GB1-KASH constructs.

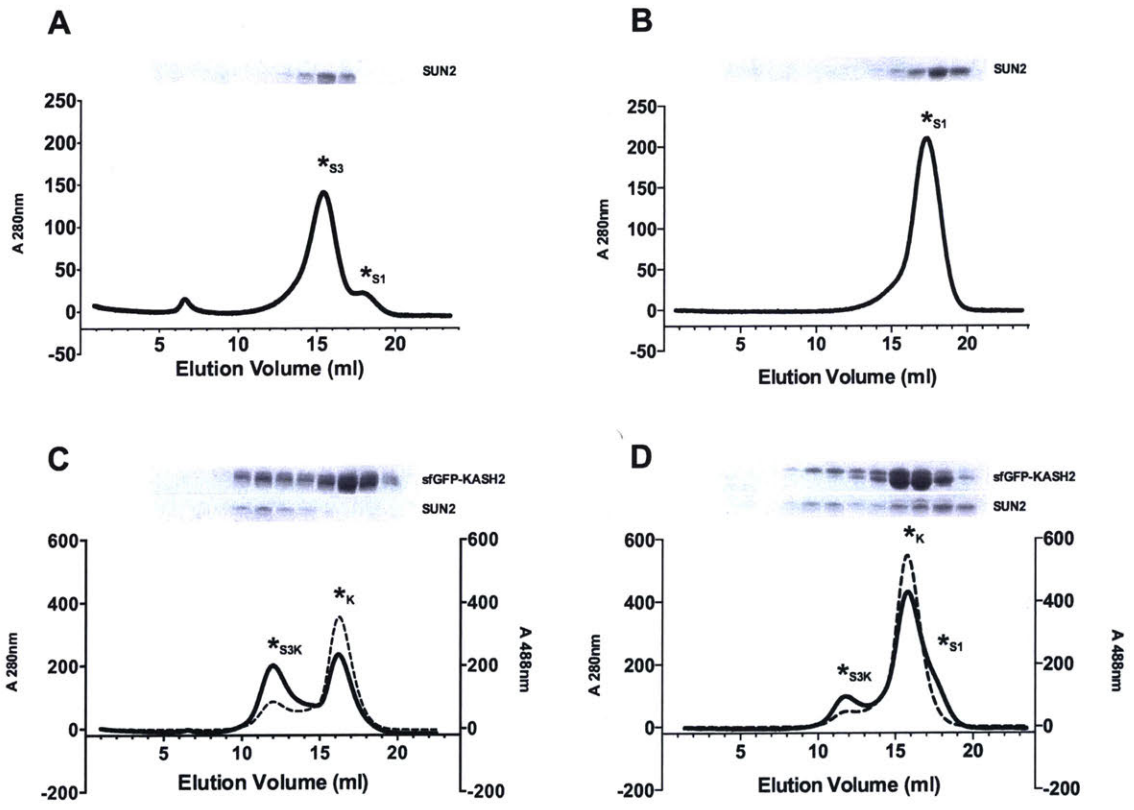


Figure 3:

In vitro SUN-KASH binding experiments. SUN-KASH binding was analyzed by gel filtration on a Superdex S200 HR10/300 column. Each panel shows a representative gel filtration profile and corresponding SDS-PAGE as follows: **(A)** Apo-SUN2 in buffer 1. **(B)** Apo-SUN2 in buffer 2. **(C)** Apo-SUN2 and sfGFP-KASH2 in buffer 1. **(D)** Apo-SUN2 and sfGFP-KASH2 in buffer 2. Asterisks are used to mark each peak, denoting the SUN2 monomer peak as S1, the SUN2 trimer as S3, the SUN2 trimer bound to sfGFP-KASH2

as S3K, and unbound sfGFP-KASH2 as K. Solid lines on each chromatogram represent the 280 nm trace; dashed lines represent the 488 nm trace. Buffer 1 contains 20 mM HEPES pH 8.0 and 100 mM KCl. Buffer 2 contains 10 mM Tris/HCl pH 8.0, 150 mM NaCl and 1 mM KCl. Each Coomassie-stained SDS-PAGE gel shows 1 ml fractions covering the 7-18 ml elution segment. SUN oligomerization is buffer-dependent. Buffers that enable SUN trimerization are required for SUN-KASH binding experiments.

References

Andersen, K. R., Leksa, N. C., & Schwartz, T. U. (2013). Optimized E. coli expression strain LOBSTR eliminates common contaminants from His-tag purification. *Proteins*, *81*, 1857–1861.

Bione, S., Maestrini, E., Rivella, S., Mancini, M., Regis, S., Romeo, G., & Toniolo, D. (1994). Identification of a novel X-linked gene responsible for Emery-Dreifuss muscular dystrophy. *Nat Genet*, *8*, 323–327.

Bonne, G., Di Barletta, M. R., Varnous, S., Becane, H. M., Hammouda, E. H., Merlini, L., et al. (1999). Mutations in the gene encoding lamin A/C cause autosomal dominant Emery-Dreifuss muscular dystrophy. *Nat Genet*, *21*, 285–288.

Borrego-Pinto, J., Jegou, T., Osorio, D. S., Auradé, F., Gorjánácz, M., Koch, B., et al. (2012). Samp1 is a component of TAN lines and is required for nuclear movement. *J Cell Sci*, *125*, 1099–1105.

Burke, B., & Roux, K. J. (2009). Nuclei take a position: managing nuclear location. *Dev Cell*, *17*, 587–597.

Chang, W., Folker, E. S., Worman, H. J., & Gundersen, G. G. (2013). Emerin organizes actin flow for nuclear movement and centrosome orientation in migrating fibroblasts. *Mol Biol Cell*, *24*, 3869–3880.

Chang, W., Worman, H. J., & Gundersen, G. G. (2015). Accessorizing and anchoring the LINC complex for multifunctionality. *J Cell Biol*, *208*, 11–22.

Chen, C.-Y., Chi, Y.-H., Mutalif, R. A., Starost, M. F., Myers, T. G., Anderson, S. A., et al. (2012). Accumulation of the inner nuclear envelope protein Sun1 is pathogenic in

progeric and dystrophic laminopathies. *Cell*, *149*, 565–577.

Crisp, M., Liu, Q., Roux, K., Rattner, J. B., Shanahan, C., Burke, B., et al. (2006). Coupling of the nucleus and cytoplasm: role of the LINC complex. *J Cell Biol*, *172*, 41–53.

di Guan, C., Li, P., Riggs, P. D., & Inouye, H. (1988). Vectors that facilitate the expression and purification of foreign peptides in Escherichia coli by fusion to maltose-binding protein. *Gene*, *67*, 21–30.

Göb, E., Schmitt, J., Benavente, R., & Alsheimer, M. (2010). Mammalian sperm head formation involves different polarization of two novel LINC complexes. *PloS One*, *5*, e12072.

Haque, F., Lloyd, D. J., Smallwood, D. T., Dent, C. L., Shanahan, C. M., Fry, A. M., et al. (2006). SUN1 interacts with nuclear lamin A and cytoplasmic nesprins to provide a physical connection between the nuclear lamina and the cytoskeleton. *Mol Cell Biol*, *26*, 3738–3751.

Haque, F., Mazzeo, D., Patel, J. T., Smallwood, D. T., Ellis, J. A., Shanahan, C. M., & Shackleton, S. (2010). Mammalian SUN protein interaction networks at the inner nuclear membrane and their role in laminopathy disease processes. *J Biol Chem*, *285*, 3487–3498.

Horn, H. F., Brownstein, Z., Lenz, D. R., Shivatzki, S., Dror, A. A., Dagan-Rosenfeld, O., et al. (2013a). The LINC complex is essential for hearing. *J. Clinical Investigation*, *123*, 740–750.

Horn, H. F., Kim, D. I., Wright, G. D., Wong, E. S. M., Stewart, C. L., Burke, B., & Roux,

K. J. (2013b). A mammalian KASH domain protein coupling meiotic chromosomes to the cytoskeleton. *J Cell Biol*, *202*, 1023–1039.

Huth, J. R., Bewley, C. A., Jackson, B. M., Hinnebusch, A. G., Clore, G. M., & Gronenborn, A. M. (1997). Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR. *Protein Science : a Publication of the Protein Society*, *6*, 2359–2364.

LaVallie, E. R., Lu, Z., Diblasio-Smith, E. A., Collins-Racie, L. A., & McCoy, J. M. (2000). Thioredoxin as a fusion partner for production of soluble recombinant proteins in Escherichia coli. *Methods in Enzymology*, *326*, 322–340.

Lindeman, R. E., & Pelegri, F. (2012). Localized products of futile cycle/lrmp promote centrosome-nucleus attachment in the zebrafish zygote. *Current Biology, 22*, 843–851.

Lu, W., Schneider, M., Neumann, S., Jaeger, V.-M., Taranum, S., Munck, M., et al. (2012). Nesprin interchain associations control nuclear size. *Cell. Molec. Life Sciences, 69*, 3493–3509.

Luxton, G. W. G., & Starr, D. A. (2014). KASHing up with the nucleus: novel functional roles of KASH proteins at the cytoplasmic surface of the nucleus. *Curr Opin Cell Biol*, *28*, 69–75.

Luxton, G. W. G., Gomes, E. R., Folker, E. S., Vintinner, E., & Gundersen, G. G. (2010). Linear arrays of nuclear envelope proteins harness retrograde actin flow for nuclear movement. *Science*, *329*, 956–959.

Luxton, G. W. G., Gomes, E. R., Folker, E. S., Worman, H. J., & Gundersen, G. G.

(2011). TAN lines: a novel nuclear envelope structure involved in nuclear positioning. *Nucleus (Austin, Tex.)*, *2*, 173–181.

Ostlund, C., Folker, E. S., Choi, J. C., Gomes, E. R., Gundersen, G. G., & Worman, H. J. (2009). Dynamics and molecular interactions of linker of nucleoskeleton and cytoskeleton (LINC) complex proteins. *J Cell Sci*, *122*, 4099–4108.

Padmakumar, V. C., Libotte, T., Lu, W., Zaim, H., Abraham, S., Noegel, A. A., et al. (2005). The inner nuclear membrane protein Sun1 mediates the anchorage of Nesprin-2 to the nuclear envelope. *J Cell Sci*, *118*, 3419–3430.

Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C., & Waldo, G. S. (2006). Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology*, *24*, 79–88.

Puckelwartz, M. J., Kessler, E., Zhang, Y., Hodzic, D., Randles, K. N., Morris, G., et al. (2009). Disruption of nesprin-1 produces an Emery Dreifuss muscular dystrophy-like phenotype in mice. *Hum Mol Genet*, *18*, 607–620.

Rothballer, A., & Kutay, U. (2013). The diverse functional LINCs of the nuclear envelope to the cytoskeleton and chromatin. *Chromosoma*, *122*, 415–429.

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

Starr, D. A., & Fridolfsson, H. N. (2010). Interactions between nuclei and the cytoskeleton are mediated by SUN-KASH nuclear-envelope bridges. *Annu Rev Cell Dev Biol*, *26*, 421–444.

Starr, D. A., & Han, M. (2002). Role of ANC-1 in tethering nuclei to the actin cytoskeleton. *Science*, *298*, 406–409.

Wang, W., Shi, Z., Jiao, S., Chen, C., Wang, H., Liu, G., Wang, Q., Zhao, Y., Greene, M. I., & Zhou, Z. (2012). Structural insights into SUN-KASH complexes across the nuclear envelope. *Cell Res*, *22*, 1440–1452.

Worman, H. J. (2012). Nuclear lamins and laminopathies. *J Pathology*, *226*, 316–325.

Zhang, Q., Bethmann, C., Worth, N. F., Davies, J. D., Wasner, C., Feuer, A., et al. (2007). Nesprin-1 and -2 are involved in the pathogenesis of Emery Dreifuss muscular dystrophy and are critical for nuclear envelope integrity. *Hum Mol Genet*, *16*, 2816–2833.

Zhang, Q., Skepper, J. N., Yang, F., Davies, J. D., Hegyi, L., Roberts, R. G., et al. (2001). Nesprins: a novel family of spectrin-repeat-containing proteins that localize to the nuclear membrane in multiple tissues. *J Cell Sci*, *114*, 4485–4498.

Zhou, Z., Du, X., Cai, Z., Song, X., Zhang, H., Mizuno, T., Suzuki, E., Yee, M. R., Berezov, A., Murali, R., Wu, S. L., Karger, B. L., Greene, M. I., & Wang, Q. (2012). Structure of Sad1-UNC84 homology (SUN) domain defines features of molecular bridge in nuclear envelope. *J Biol Chem*, *287*, 5317–5326.

The *Caenorhabditis elegans* Protein FIC-1 is an AMPylase that Affects Susceptibility to *Pseudomonas aeruginosa* Infections

This section is adapted from Truttmann MC, Cruz VE, Guo X, Engert C, Schwartz TU, Ploegh HL. "The Caenorhabditis elegans Protein FIC-1 Is an AMPylase That Covalently Modifies Heat-Shock 70 Family Proteins, Translation Elongation Factors and Histones." PLOS Genetics 2016; 12(5):e1006023. PMCID: 27138431

Abstract
Protein AMPylation by Fic domain-containing proteins (Fic proteins) is an ancient and conserved post-translational modification of mostly unexplored significance. Here we characterize the *Caenorhabditis elegans* Fic protein FIC-1 *in vitro* and *in vivo*. FIC-1 is an AMPylase that localizes to the nuclear surface and modifies core histones H2 and H3 as well as heat shock protein 70 family members and translation elongation factors. The three-dimensional structure of FIC-1 is similar to that of its human ortholog, HYPE, with 38% sequence identity. We identify a link between FIC-1-mediated AMPylation and susceptibility to the pathogen *Pseudomonas aeruginosa*, establishing a connection between AMPylation and innate immunity in *C. elegans*.

Introduction
How post-translational modifications regulate protein activity is a fundamental question in biology. Phosphorylation, methylation or acetylation reversibly control cellular signaling pathways, the misregulation of which is often associated with pathologies, including cancer or autoimmune diseases [1, 2]. AMPylation, the covalent addition of AMP to a target protein, has recently been described as a new post-translational modification found in both prokaryotes and eukaryotes. However, protein AMPylation is far less well understood as a post-translational modification and its implications for cellular physiology remain largely unknown. Protein AMPylation in metazoans is catalyzed by fic-domain containing proteins (Fic proteins). Fic proteins are an evolutionarily conserved protein family, numbering approximately 2700 members distributed over most kingdoms of life, with the exception of fungi and plants [3, 4]. While many bacterial species encode a number of different Fic proteins, most eukaryotes,—including *Caenorhabditis elegans*, *Drosophila melanogaster*, *Mus musculus* and *Homo sapiens*—carry only a single gene that specifies a Fic family member. All Fic proteins share a conserved sequence motif (HxFx(D/E)GN(G/K)R) found in their respective fic domains, including an invariant histidine required for catalysis [5]. Fic proteins accept a variety of nucleotide substrates, including ATP and UTP, to covalently AMPylate (adenylylate), UMPylate or phosphorylate their targets. However, protein AMPylation—the covalent addition of an AMP moiety to the target protein at the expense of a single ATP—is their predominant activity [6–9]. In bacteria, Fic protein-mediated AMPylation of Gyrase and Topoisomerase IV has been linked to toxin-antitoxin systems such as the VbhT-VbhA pair found in *Bartonella schoenbuchensis* [5, 10]. In addition, several pathogens evolved effector proteins equipped with Fic-domains that—upon translocation into the host cell—interfere with host cell signaling. They do so by covalently AMPylating and thus inactivating small GTPases of the Rho and Rab family [11, 12].
In eukaryotes, AMPylation by Fic proteins may regulate the unfolded protein response (UPR), as well as carry out covalent modification of histones [13–16]. The Drosophila Fic protein dfic as well as the human Fic protein HYPE target and modify the ER-resident HSP70 protein BiP/GRP78 [14, 15]. While ER stress

increases intracellular dFic / HYPE as well as BiP levels, the consequences of ER stress for AMPylation and activity of BiP remains controversial. Induction of the UPR may lessen BiP AMPylation, whereas a competing model infers an increase in AMPylation of BiP, resulting in increased ATPase activity [14–16]. Addressing Fic protein biology in *C. elegans* may help to resolve this paradox. Here, we characterize the *C. elegans* Fic protein FIC-1 *in vitro* and *in vivo*. We show that FIC-1 modulates antimicrobial defense responses of *C. elegans* against *Pseudomonas aeruginosa*, often used as a simple eukaryotic model of infectious disease and innate immunity. We show that FIC-1 is ubiquitously expressed throughout the nematode body. We demonstrate that FIC-1 acts as an AMPylase and covalently modifies core histones, HSP 70 family members and translation elongation factors. Finally, we determine the crystal structure of FIC-1 and its constitutively active mutant form FIC-1 E274G and identify a potential binding site for endogenous regulators. Our results provide the first evidence for a role of Fic protein-mediated AMPylation in protection of the host.

Results

FIC-1 alters susceptibility to killing by *P. aeruginosa*

C. elegans, which carries a single gene encoding a fic-domain containing protein (FIC-1), is a versatile model to study longevity, stress responses or innate immunity. We asked if changes in FIC-1 levels or activity resulted in global fitness defects in *C. elegans*. Therefore we created a *fic-1* deletion allele using CRISPR technology. The deletion allele (*n5823*) contains a 7 bp deletion in *fic-1*'s Exon IV, resulting in a pre-mature stop codon (S1A and S1B Fig). We also expressed a presumably constitutively active form of *fic-1*, FIC-1[E274G](nIs733) under the control of its endogenous promotor. First, we performed longevity assays to evaluate whether changes in FIC-1 activity might affect lifespan and observed no significant differences between *fic*-1(n5823) mutants, and FIC-1[E274G](nIs733) constitutively active animals or wild type controls at 20°C or 25°C (Fig 1A–1D, additional independent replicate shown in S8A–S8C Fig). When grown on a *P. aeruginosa* lawn, however, *fic*-1(n5823) mutants displayed increased susceptibility to killing by *P. aeruginosa* as compared to wild type (Fig 1F, additional independent replicate shown in S9A–S9H Fig). This was not due to a defect in pathogen sensing, as *fic-1*(n5823) mutant animals appeared to be unaffected in avoidance assays (Fig 1E). Expression of FIC-1 E274G slightly increased pathogen tolerance and enhanced relative infection outcome. Further, we could rescue survival of *fic-1*(n5823) mutants by expressing wild type *fic-1* in the mutant animals. Together, these results establishing a role for Fic proteins in innate immunity.

FIC-1 is not a master regulator of ER stress responses in *C. elegans*.

Studies in *D. melanogaster* as well as in human cells suggested a connection between Fic proteins and the regulation of the unfolded protein response (UPR) [14–16]. In *C.elegans*, animal development on *P. aeruginosa* requires the

presence of the X-box binding protein 1 (XBP-1), an immediate downstream target of the IRE-1-regulated UPR branch, while other UPR branches were dispensable [17].

To investigate if the observed *fic-1*-dependent changes in pathogen susceptibility were linked to changes in ER stress signaling, we examined involvement of *fic-1* in the induction of stress responses upon exposure to specific stress cues. The reporter constructs *hsp-4*::GFP (ER stress) or *hsp-6*::GFP (mitochondrial stress) revealed no apparent difference in either the ER (tunicamycin) or mitochondrial (ethidium bromide) stress response when tested in *fic-1*(n5823) mutants or animals expressing FIC-1[E274G](nIs733) (S2A and S2B Fig), using *E.coli* as a food source. When we transferred embryos onto NGM plates containing various concentrations of tunicamycin and scored development as well as adult survival, *fic-1*(n5823) loss of function nor the constitutively active form (nIs733) affected the outcomes (Fig 2A). Thus, we conclude that FIC-1 is not essential in inducing or sustaining the UPR in *C. elegans*. Next, we repeated our development assays, exposing animals to *Pseudomonas aeruginosa*. Changes in FIC-1 activity did not affect nematode development on *P. aeruginosa* (Fig 2B). As *C. elegans* encodes two Grp78/BiP homologues, *hsp-3* and *hsp-4*, assumed to cross-compensate for each other in their roles as ER-residing protein chaperones, we repeated these assays in the presence of *P. aeruginosa* or tunicamycin in *fic-1;hsp-3* mutants to possibly render them more sensitive to changes in FIC-1 activity. We also examined the impact of FIC-1 in a *xbp-1*-deficient background, because in human cells the xbp-1-linked branch of the UPR is not known to be modulated by HYPE, while the PEK-1 and ATF-6-linked branches are [14]. None of the conditions tested showed a significant difference between *fic-1*(n5823) mutants or constitutive active (nIs733) animals and their respective controls (Fig 2C and 2D, S2C and S2D Fig). *fic-1* is therefore not a key regulator of the UPR in *C. elegans*. Of note, *hsp-3* nematodes were very sensitive to *P. aeruginosa* exposure during early development and only few reached adulthood, highlighting a role for HSP-3 in the tolerance of chronic ER stress and innate immunity.

FIC-1 is expressed throughout *C. elegans* and enriched at the nuclear interface

To get a better understanding of the cell types and body parts that express FIC-1 *in vivo*, we applied a single molecule fluorescence *in situ* hybridization assay (smFISH) to detect endogenous *fic-1* mRNA [18]. Our results showed the presence of *fic-1* mRNA throughout the animal's body, including the germline, and at all developmental stages (Fig 3A and 3B). Embryonic expression levels appeared comparatively high and were further confirmed in animals expressing GFP under the control of the putative *fic-1* promotor (P*fic-1*::*gfp*) (S3A Fig). To characterize localization of FIC-1, we analyzed by fluorescence microscopy embryos of animals that express C-terminally HA-tagged FIC-1 under the control of a strong heat-shock promotor. Inducible FIC-1 expression was confirmed by immunoblotting (S3B Fig), and FIC-1-HA was detected using either an anti-HA antibody or a mouse anti-FIC-1 serum previously validated using recombinant

FIC-1 proteins (S3C Fig). We detected multiple FIC-1-HA species by immunoblotting with anti-HA antibody in *C.elegans* lysates, ranging in size from 58 kDa (full length) to approximately 30 kDa, indicating N-terminal processing of a fraction of FIC-1 (S3C Fig) either by proteolysis or alternative translation initiation. While we were unable to detect FIC-1 prior to induction (Fig 3C, uninduced), we observed low levels of FIC-1 expression throughout the cell with a notable accumulation of FIC-1 at the nuclear interface or the nuclear envelope / ER upon heat-shock (Fig 3C, induced). This enrichment at the nucleus-ER interface is reminiscent of intracellular HYPE localization [13, 14]. We further probed FIC-1 localization by sub-cellular fractionation and immunoblotting and confirmed significant enrichment of FIC-1 in the nuclear and ER fractions, with less presence in the cytoplasm (Fig 3D). We also over-expressed GFP-tagged FIC-1 E274G in HeLa cells and analyzed its intracellular distribution pattern by microscopy. We likewise observed GFP-FIC-1 accumulation in the nuclear envelope as well as in associated structures, similar to the pattern seen for GFP-HYPE E234G (S3D Fig). We conclude that FIC-1 is ubiquitously expressed in *C. elegans* and shows an intracellular localization pattern similar to that of HYPE, with an additional presence of FIC-1 in the cytoplasm.

FIC-1 is an AMPylase

To test a possible catalytic activity of FIC-1, we expressed and purified from *E. coli* a truncated version of FIC-1, as well as two additional mutant versions, FIC-1 E274G and FIC-1 H404A. While substituting the corresponding glutamic acid in FIC-1's human ortholog HYPE with glycine results in a hyper-active enzyme (E234G), exchanging the conserved histidine with an alanine (H363A) diminishes HYPE's target AMPylation activity. We first assessed self-AMPylation using αP^{33}-ATP (Fig 4A). As expected, wild type FIC-1 showed only weak self-AMPylation, FIC-1 E274G exhibited a massive increase in self-modification, while FIC-1 H404A did not show detectable self-AMPylation. We also tested the ability of FIC-1 E274G to accept nucleotides other than ATP as a substrate for self-modification (Fig 4B). Self-AMPylation as well as self-GMPylation proceeded at high rates, but we also observed self-CMPylation and self-UTPylation, indicating that FIC-1 E274G is even more promiscuous in its preference for nucleoside triphosphates than HYPE E234G. To test for target AMPylation, we repeated our FIC-1 *in vitro* AMPylation assay using recombinant histone H3 as substrate (Fig 4C). FIC-1 as well as FIC-1 E274G AMPylated histone H3 while FIC-1 H404A showed no such activity. To map the modified residue(s) on FIC-1, we performed self-AMPylation assays using ATP and subjected the samples to LC-MS/MS. We identified two auto-AMPylation sites, T352 and T476, AMPylated in $FIC\text{-}1_{258-508}$, $FIC\text{-}1_{258-508}$ E274G, $FIC\text{-}1_{134-508}$ and $FIC\text{-}1_{134-508}$ E274G samples (S4A and S4B Fig). Mutation of these sites individually in a FIC-1 E274G background did not drastically affect auto- or target AMPylation levels (Fig 4D). We also explored the ability of FIC-1 to AMPylate other histone family proteins previously identified as targets for its human homologue HYPE [13]. FIC-1 E274G modified histones H2A, H2B, H3.1, H3.2 H3.3 but not H1 or H4. Thus, the set of *in vitro* FIC-1

targets overlaps greatly with that of HYPE, yet differs at least in its ability to modify histone H4 (Fig 4E).

Structure and conservation of FIC-1

The crystal structure of HYPE was described recently [19]. The differences in target specificity of HYPE and FIC-1 as well as their sequence divergence (38% amino acid sequence identity) prompted us to solve the structures of FIC-1 and FIC-1 E274G crystallographically. Both proteins crystallized in the same space group and the structures could readily be solved by molecular replacement using apo-HYPE as a search model. The asymmetric unit contains a dimer of FIC-1, arranged in a similar way as seen with human HYPE (Fig 5A and S5A and S5B Fig). Like HYPE, FIC-1 is also a tripartite, entirely helical protein, with an N-terminal TPR element followed by a linker helix connecting to the Fic domain. The TPR element contains two stacked TPR motifs, each composed of two antiparallel helices. The fic domain of FIC-1 and HYPE superpose very well with a root mean square deviation of 1.4 Å. Only the linker helix and the TPR element are slightly shifted with respect to another, likely influenced by the different crystal packing environments. The fic domain in FIC-1 consists of 8 α-helices in total, where three helices precede the fic core and the ATP binding site. The first of these helices, known as "α-inh", contains the auto-inhibitory glutamate at position 274, while the following two helices have been dubbed the pre-A and pre-B helices. The fic core itself is a four-helix bundle (α1-α4), containing the conserved catalytic motif (HxFx(D/E)(A/G)N(GK)R), represented in *C. elegans* by the sequence **HPFT**D**GNGR** [20]. The next conserved feature is a loop located between helices 2 and 3 of the fic core, called the flap. The flap is not visible in our structure, which suggests inherent flexibility of this motif. After the fic core, the final two helices called helices post-A and post-B pack against both the TPR motif and the linker helix, positioning the C terminus close to the loop between the linker helix and the auto inhibitory helix (S5C Fig, top panels). FIC-1 crystallized with a sulfate occupying the β-phosphate site of ATP, to highlight similar features with HYPE we modeled an ATP molecule onto our Fic-1 structure (S5C Fig, bottom panels), the resulting hydrogen-bonding pattern between the fic core of FIC-1 and ATP are nearly identical to that of HYPE and ATP. Our data also suggest that eukaryotic Fic proteins are constitutive dimers. Similar to the human fic domain-containing protein HYPE, FIC-1 also crystallizes as a dimer. The dimer is held together by two discrete interfaces that together bury 676.4 $Å^2$ of solvent-accessible surface area (Fig 5B). The first interface is more extensive (384.9 $Å^2$) and is composed of helix pre-B and its preceding loop. This interaction is likely driven by the hydrophobic effect of burying V292 and I298 of both monomers in the nearly symmetrical binding interface. In the second interface (291.4 $Å^2$) the hydrogen bonding between the side chains of N341 is the most noticeable feature (Fig 5B, inset).

To test if FIC-1 is a dimer in solution and if dimerization is required for enzymatic activity, we generated point mutants designed to disrupt the dimerization interface. Based on our structure we predicted that an I298D mutation should

inhibit dimer formation without affecting the protein fold. We cloned and purified FIC-1 I298D as well as FIC-1 E274G/I298D. As expected, the mutant protein behaved as a monomer in solution as shown by size exclusion chromatography while FIC-1 wild-type eluted significantly earlier (Fig 5E). *In vitro* AMPylation testing FIC-1 E274G/I298D indicated a direct connection between FIC-1 dimerization and its ability to modify targets: Relative to FIC-1 E274G's capacity to AMPylate histone H3, FIC-1 E274G/I298D self-AMPylation was reduced to 46% (Fig 5F, upper panel and Fig 5G) and target AMPylation activity to 11% (Fig 5F, lower panel and Fig 5G), respectively.

To further compare FIC-1 with other eukaryotic Fic representatives, we performed an alignment of fic domain-containing proteins in highly divergent metazoans and mapped the conserved features to the surface of a FIC-1 monomer (Fig 5C). As expected, the ATP binding site is conserved throughout, while the TPR domain and the linker helix show less conservation. The dimer interface itself is highly conserved, suggesting that the FIC-1 dimer is the active form of the proteins in metazoans (Fig 5D). On the opposite side from the ATP-binding pocket, we observe another conserved region that forms a deep groove (Fig 5D, bottom left). This groove may accommodate binding partners, or provide an assembly point for (a) larger complex(es).

FIC-1 AMPylates heat-shock protein 70 family members as well as translation elongation factors

While we were able to demonstrate that FIC-1 AMPylates core histones *in vitro*, we hypothesized that there might be additional FIC-1 targets potentially linking AMPylation to innate immunity. To identify these proteins, we adapted a click-chemistry based approach, in which we spiked *C. elegans* total cell lysate with recombinant FIC-1 protein in the presence of N^6-propargyl-ATP as nucleotide substrate (S6A Fig) [21]. Following AMPylation, we completed the reaction with biotin-$(PEG)_3$-azide to covalently couple a biotin handle to the ATP-bound propargyl group. We recovered AMPylated and thus biotinylated proteins from total *C. elegans* lysate on Streptavidin-modified agarose beads. Bound proteins were eluted and analyzed by LC/MS/MS. A comparison of the hit list with two independent controls to eliminate false positives led to identification of two classes of proteins over-represented amongst the AMPylated fraction of proteins: HSP 70 proteins (HSP-1, HSP-3) as well as translation elongation factors (eEF-1A, eEF-1G, eEF-2) (Fig 6A). Heat-shock proteins possess chaperone activity: they bind to unfolded or misfolded targets and support their proper refolding or—when beyond repair—shuttle them towards degradation [22]. HSP-1 is predominantly cytosolic, whereas HSP-3 is retained within the ER lumen. Interestingly, HSP-3 is the *C. elegans* Bip/Grp78 ortholog. HSP-3—together with its close homolog, HSP-4—form a complex with IRE-1, ATF-6 and PEK-1, to preclude activation of UPR-related signaling events by preventing either oligomerization of IRE-1 and PEK-1 or proteolytic processing of ATF-6 in the golgi [23]. BiP/Grp78 and HSP-3 / HSP-4 share more than 70% sequence similarity, with most of the divergence localized to the very N and C termini (S6B

Fig). The second class of identified targets—translation elongation factors—regulate protein translation by coordinating the selection and binding of aminoacyl-tRNA to the ribosome's A-site (eEF-1A / EF-Tu) and by controlling the translocation of the peptidyl-tRNA from the A-site to the P-site of the ribosome (eEF-2) [24, 25]. *C. elegans* eEF-1A and eEF-2 share more than 80% sequence similarity with their human orthologs (S6C Fig). To validate these targets, we recombinantly expressed and purified *C. elegans* representatives of each family, HSP-1, HSP-3 as well as eEF-1A2. We tested them for modification in an *in vitro* AMPylation reaction, which confirmed HSP-1, HSP-3 and eEF-1A2 as substrates for FIC-1 E274G (Fig 6B). The human ortholog HYPE E234G modified these new FIC-1 targets as well (Fig 6C), indicative of their functional similarities. Our results thus reveal new Fic protein targets and suggest a role for AMPylation in the regulation of the HSP-3-dependent branch of the UPR as well as protein translation, a combination of which might account for the observed changes in pathogen susceptibility.

AMPylation site mapping on *C. elegans* proteins highlights poly-modifications

AMPylation by eukaryotic FIC proteins is a site-specific process where threonines represent the preferred sites of modification. To characterize AMPylation of *C. elegans* targets by FIC-1 E274G or HYPE E234G, we used a combined approach of LC-MS/MS analysis and *in vitro* AMPylation assays, testing recombinantly expressed targets with specific mutations that alter presumptive sites of AMPylation. We previously showed that histone H3 is not modified on tyrosines by HYPE E234G [13]. As eukaryotic FIC proteins preferentially modify threonines, we constructed, purified and tested three additional histone H3 mutants: histone $H3_{allTtoA}$ has all threonines replaced with alanines, $H3_{24\text{-}145}$ misses the N-terminal 24 amino acids among which are four threonines and $H3_{STtoAA}$ has two serine/threonine (S/T) motifs mutated to alanine/alanine (A/A). We observed that HYPE E234G was unable to AMPylate $H3_{allTtoA}$ while overall AMPylation levels of $H3_{24\text{-}145}$, and $H3_{STtoAA}$ were decreased, as compared to wild type H3 (Fig 7A). In contrast, FIC-1 E274G did not modify any of these histone H3 mutant proteins (S7A Fig). To address eEF-1A2 AMPylation, we first purified a truncated eEF-1A2 version ($eEF\text{-}1A2_{244\text{-}463}$), modified it in *in vitro* AMPylation reactions with FIC-1 E274G and analyzed the sample by LC-MS/MS. We detected two AMPylation sites, T269 and T432, with high confidence. Since AMPylation of human eEF-1A2 was recently mapped to T261, we recombinantly purified $eEF\text{-}1A2_{244\text{-}463}$ T261A as well as $eEF\text{-}1A2_{244\text{-}463}$ T432A, and tested these proteins in *in vitro* AMPylation assays using $\alpha\text{-}P^{33}$-ATP as nucleotide source. In reactions with FIC-1 E274G, we observed a significant reduction in $eEF\text{-}1A2_{244\text{-}463}$ T432A AMPylation while $eEF\text{-}1A2_{244\text{-}463}$ T261A modification was indistinguishable from $EEF\text{-}1A2_{244\text{-}463}$ wild type (Fig 7B and 7C). We did not observe a significant difference in AMPylation levels when using HYPE E234G as AMPylator in reactions with $EEF\text{-}1A2_{244\text{-}463}$ constructs (S7B Fig). To map the sites of AMPylation on HSP-1 and HSP-3, we first subjected *in vitro* modified

proteins to LC-MS/MS analysis and identified T176 on HSP-3 as the only site of modification. The human HSP-3 orthologue BiP was previously shown to be modified on T366 and T518, respectively [14, 16]. This prompted us to clone and purify HSP-1 and HSP-3 versions with mutations in the respective BiP S365/T366 or T518 orthologue residues (HSP-1 T342A, HSP-3 S370A/T371A, HSP-3 T523A). We first tested these mutant HSP-1 and HSP-3 versions in *in vitro* AMPylation assays using FIC-1 E274G as AMPylator and observed no significant changes in AMPylation levels as compared to wild type HSP-1 or HSP-3 (Fig 7D and 7E). Contrasting, HYPE E234G-medatied AMPylation of HSP-1 T342A was reduced as compared to wild type HSP-1 while HSP-3 AMPylation levels didn't fluctuate (S7C and S7D Fig). Together, our results highlight that *C. elegans* proteins are AMPylated on multiple sites, which do not necessarily overlap with modified human orthologues.

Discussion

Fic protein-mediated target AMPylation is a conserved post-translational modification that may serve to regulate target activity. The ER-resident *D. melanogaster* Fic protein, dFic, as well as its human ortholog, HYPE, AMPylate BiP, an ER chaperone involved in the regulation of the UPR [15, 16]. Further, HYPE also modifies core histones H2-H4 *in vitro*, suggesting a possible role for HYPE in stress and DNA damage control [13, 14]. However, the intracellular localization of eukaryotic Fic proteins, their regulation, the identity of the physiologically most relevant targets, as well as their role in modulating cellular signaling events remain elusive.

Here we biochemically characterized the *C. elegans* Fic protein FIC-1, solved its structure and investigated its role in *C. elegans* stress tolerance *in vivo*. As expected based on the level of conservation of the catalytic and regulatory domains, FIC-1 acts as an AMPylase, capable to add an AMP entity to itself (auto-AMPylation) or to a target protein (target AMPylation). We map the auto-AMPylation sites to T352 and T476, two surface-exposed amino acids that are far from the active site of either monomer in the dimeric structure. T352 sits on the highly flexible flap structure that is likely to be accessible for self-modification. However, auto-AMPylation of T476 is hard to rationalize with respect to its physical position relative to the active site. Thus, AMPylation of T476 might therefore represent trans-AMPylation events rather than self-modifications. Since we purified and tested FIC-$1_{134-508}$ and FIC-$1_{258-508}$ only, we were not able to confirm any modifications on the orthologues of mapped HYPE self-modification sites T76, T80 or T183 [14]. However, $HYPE_{aa187-437}$ E234G, which lacks all 3 presumable auto-AMPylation sites, remains fully active, displays self-modification and is capable of target modification.

Similar to HYPE and dFic, FIC-1 is hardly active in *in vitro* assays, while a single point mutation in the regulatory site (FIC-1 E274G) renders the enzyme more active without obviously altering its target specificity or increasing its promiscuity. We hypothesize that *in vivo*, FIC-1 is an efficient AMPylator, comparable to FIC-1

E274G, if a proper activation signal is provided, for example through interaction with (a) relevant partner protein(s). However, no such Fic protein regulators have been described yet.

We identify HSP-1, HSP-3 as well as eEF-1A, eEF-1G and eEF-2 as novel FIC-1 targets, all belonging to conserved protein families. HSP-1 and HSP-3 are heat-shock family 70 proteins and share >80% amino acid similarity with their human counterparts HSC-70 and BiP/Grp78, respectively. The human heat-shock 70 family protein BiP is AMPylated by HYPE on S365/T366 or T518 [15, 16]. Residues S365/T366 are highly conserved and present in BiP homologs found across species. *C. elegans* HSP-1, HSP-3 and HSP-4 all contain the very same amino acids as part of a strictly conserved amino acid stretch near these proteins' C termini (LVGGS**T**RIPK). In contrast, T518 of human BiP is present only in HSP-3 and absent from HSP-1 and HSP-4. We map HSP-3 AMPylation to T176, yet another amino acid that is part of a strictly conserved sequence motif shared among BiP homologs (AVV**T**VPAYFND). We further confirm that neither of the orthologous BiP AMPylation sites, S365/T366 and T518, are modified on HSP-1 or HSP-3 by FIC-1. Modification of different sites on orthologous proteins might reflect minor selectivity distinctions between HYPE and FIC-1. Notably, we observe that HYPE E274G preferentially modifies T342 on HSP-3, the equivalent to T366 in BiP. Whether AMPylation of different—or even multiple—sites on Heat shock 70 family proteins results in diverse changes of their activities remains to be studied.

The regulation of the HSP-3 / HSP-4 ortholog BiP by post-translational modifications is not a new concept: first, BiP is modified by ADP-ribosyltransferases on R470 and 492 in the substrate binding domain, resulting in reversible BIP inactivation [26]. Second, BiP is auto-phosphorylated on a threonine residue in close proximity to the catalytic cleft, presumably as a consequence of BiP's inherent ATPase activity [27]. The consequences of BIP AMPylation are a matter of debate. Cells that experience ER stress up-regulate HYPE, presumably increasing AMPylation of BIP and other target proteins. BIP AMPylation was hypothesized to facilitate its dissociation from a set of substrates that include IRE-1, PEK-1 and ATF-6 to support the initiation of the UPR. In this study, we explored a possible link between AMPylation and the induction of an ER stress response, but we failed to observe a strong connection between AMPylation activity and the UPR, nor did we see connections between AMPylation activity and early development or survival under conditions of ER stress. Our findings that AMPylation had no measurable effect on the IRE-1 / XBP-1 branch of the UPR are in accordance with published results, demonstrating that under acute ER stress conditions, HYPE-mediated target AMPylation is required for the activation of PERK and ATF6-based UPR cascades but not for IRE-1 activation [14]. Thus, the inbuilt redundancy in the regulation of the UPR may mask the consequence of Hypo- and Hyper-AMPylation on a organismic level. While we hypothesize that AMPylation of HSP-3 and other factors could affect ER signaling, we conclude from our data that

FIC-1 does not obviously regulate the UPR in *C. elegans* and may represent a "soft" rather than a major regulator of the UPR.

The cytosolic heat shock protein HSP-1 is involved in the regulation of nuclear export of DAF-16 following physiological stress. Similar to DAF-16, HSP-1 is exclusively cytosolic but upon stress partially relocalizes to the nucleus [28].

Like heat shock proteins, the identified translation elongation factors eEF-1A, eEF-1G and eEF-2 are > 75% identical to their human orthologs. Modifications of translation elongation factors by Fic proteins also occur in bacteria. The P1 bacteriophage-encoded Fic protein Doc phosphorylates *E. coli* EF-Tu at position T382, inhibiting translation [8, 9]. AMPylation of elongation factors may therefore alter translation in affected cells. We mapped eEF-1A2 poly-AMPylation by FIC-1 to T269 and T432. HYPE modifies human eEF-1A on T261 [29]. We find that T261 is not altered by FIC-1 or HYPE in *in vitro* reactions using *C. elegans* eEF-1A2 as substrate. *C. elegans* T432 is the residue that corresponds to *E.coli* eIF-Tu T382, indicating overlapping target site preferences between *E.coli* Doc and FIC-1. Of particular interest is the fact that eEF-1A2 is modified on two distinct sites. Whether or not these sites are simultaneously engaged, resulting in target oligo-AMPylation, or modified independently remains to be investigated.

Finally, FIC-1 also reliably AMPylates core histones H2 and H3 at least *in vitro*. Histones are among the most conserved proteins found in nature, with the human and *C. elegans* versions sharing ~ 80% and up to 97% (H3) or 98% (H4) sequence identity [30]. HYPE AMPylates core histones H2-H4 but not H1 [13]. FIC-1 AMPylates histones H2-H3 but not H1 or H4. In contrast to AMPylation, histone methylation occurs on lysines or arginines contained in the N-terminal flexible histone tail, while histone acetylation is restricted to lysine residues [31, 32]. Histones may also be phosphorylated on serines, threonines and tyrosines [33], an important modification during the DNA damage response, when phosphorylated histone H2A assembles in chromatin domains at sites of DNA breakage. Further, Chk-1-mediated histone H3 phosphorylation on T11 rapidly decreases in cells experiencing DNA damage, resulting in the repression of a set of genes including Cdk1 and cyclin B1 [34]. Threonine residues are the preferred targets for HYPE and FIC-1. Removal of the N-terminal 24 amino acids or mutation of the two serine/threonine motifs in histone H3 greatly reduced HYPE-mediated and abolished FIC-1-mediated H3 AMPylation, strongly suggesting oligo-AMPylation by HYPE. However, we cannot exclude the possibility that these mutations distort the overall fold of histone H3, thus preventing AMPylator-H3 interactions and subsequent modifications. Threonine oligo-AMPylation may mimic histone phosphorylation and play a role in the DNA damage response, too. Despite our best efforts, the ability to recover the modified endogenous targets in amounts that would enable an *in vivo* validation of our *in vitro* data has so far exceeded our experimental capabilities. Thus, it remains to be tested which proteins may represent the primary *in vivo* targets of FIC-1 in *C. elegans*.

To fully characterize the *C. elegans* protein FIC-1, we solved the atomic structure of FIC-1 and FIC-1 E274G. FIC-1 and HYPE are structurally very similar despite

their sequence divergence (38% amino acid sequence identity), which may explain the partially overlapping specificity of these two enzymes. As observed for HYPE E234G, removal of the auto-inhibitory glutamate in FIC-1 did not cause obvious structural changes that could account for the massive increase in enzyme activity. While the TPR domain—a known protein-protein interaction module—is likely to bridge interactions of other proteins to FIC-1 and HYPE, we identified a second potential interaction site for binding partners opposite the ATP binding groove [35]. Given its location in close proximity to the active site, this particular site might be occupied by FIC-1 activators / inhibitors and thus fulfill an important role in enzyme regulation. The recruitment of different modulators to this potential interaction site may also explain the observed differences in target specificities of the two enzymes. Further, like HYPE, FIC-1 crystallized as a dimer; the high level of conservation at the dimerization interface among metazoan Fic proteins suggests that FIC-1 might preferentially exist as a dimer in intact cells, too. Recent work on bacterial FIC proteins showed the existence of an inhibitory tetrameric NmFic complex in solution [36]. We failed to obtain evidence for FIC-1 tetramerization in our experiments. However, the disruption of the dimerization interface dramatically decreases self- and target AMPylation of FIC-1 E274G, suggesting that dimerization might enhance its activity. Whether FIC-1 dimerization has consequences on target specificity remains to be investigated.

Complementary to our biochemical and structural investigation of FIC-1, we assayed its role in *C. elegans* physiology. We observed FIC-1 transcripts (mRNA) in all embryonic, larval and adult stages, suggesting a role for FIC-1 throughout development. The intracellular distribution of FIC-1 resembles the localization pattern of HYPE. We also detect a fraction of FIC-1 in the cytoplasm. We hypothesize that this cytoplasmic FIC-1 fraction modifies cytosolic target proteins such as HSP-1. Although the presence of HYPE in the cytoplasm of human cells has not been previously reported, at least one study reported HYPE-dependent AMPylation of cytosolic proteins, likewise arguing for the presence of HYPE in the cytoplasm, [29]. Of note, FIC-1-HA displays at a slightly lower apparent molecular weight in the nuclear and ER fractions than in the cytoplasmic fraction. These small differences might be attributable either to the prevailing self-AMPylation status of FIC-1 in the distinct cellular compartments or reflect other, as yet uncharacterized, regulatory modifications of FIC-1. Despite the presence of FIC-1 at early developmental stages, *fic-1* deficient animals develop normally and show no defects in brood size, egg viability or egg development even in the presence of acute or chronic ER stress. Instead, we observed a moderately increased susceptibility to lethal *P. aeruginosa* infections in *fic-1* deficient nematodes and, accordingly, slightly enhanced pathogen tolerance upon hyper-AMPylation as induced either by FIC-1 E274G or several extra copies of the *fic-1* wild type gene as present in our *fic-1*(n5823;nIs734) rescue line. Unexpectedly, *hsp-3* nematodes showed a hyper-sensitivity phenotype with regard to egg development on *P. aeruginosa* as a food source.

These results support a model in which certain branches of UPR are essential for survival in the presence of bacterial pathogens. Innate immunity in *C. elegans* is controlled by DAF-2, PMK-1 and, to a lesser degree, HSF-1 [37]. The p38 mitogen-activated protein kinase PMK-1 pathway orchestrates the up-regulation of CUB-like proteins as well as C-type lectins in response to pathogen infections [37]. In contrast, HSF-1, a transcription factor activated by physiological stresses such as elevated temperatures, was proposed to act downstream of DAF-2 / DAF-16 and independently of PMK-1 [28]. Among the genes regulated by HSF-1 are molecular chaperones of the heat-shock protein family (HSPs) that may indirectly add to the innate immune response by supporting the refolding of unfolded proteins resulting from the release of ROS as a measure to fight the bacterial infection. Whether FIC-1 is involved in the regulation of DAF-2, HSF-1 or PMK-1 based immunity traits will require more extensive genetic and biochemical analysis.

Recent work highlighted a novel link between post-translational histone modifications and innate immunity in *C. elegans*: RNAi ablation of the H3K4 methyltransferase set-16/MLL decreased H3K4me3 levels at infection-associated gene promotors, leading to reduced transcription of these genes [38]. Further, global K14me1 levels as well as the intracellular distribution of K14me1-modified linker histone H1 (HIS-24) change in response to bacterial infections in *C. elegans*, suggesting an epigenetic component in the induction of an innate immune response [39]. In addition to a potential role in the regulation of DNA damage responses, histone AMPylation might alter the transcription of infection-related genes. We are currently exploring such possible connections.

We propose the following model to explain how FIC-1 mediated target AMPylation might alter protein functions that result in the phenotypes observed (Fig 8A): Infection of *C. elegans* with *P. aeruginosa* triggers an innate immune response, resulting in the release of reactive oxygen species (ROS). These molecules damage both DNA and proteins and indirectly initiate cellular repair mechanisms [40, 41]. In wild type animals exposed to stress, FIC-1 AMPylates HSP-1, HSP-3 and presumably HSP-4, thereby increasing dissociation of these chaperones from their intrinsic binding partner(s) and supporting the activation of the UPR in the ER lumen. Similar to phosphorylation, AMPylation of elongation factors could further contribute to UPR activation by inhibiting eEF-2, attenuating translation and increasing the specific expression of ATF-4, a transcription factor positively controlling the transcription of UPR-linked genes [42]. Histone AMPylation might limit the transcription of infection-associated genes and—together or in parallel with phosphorylation—trigger and support the repair of DNA damage introduced by ROS. Consequently, absence of AMPylation activity would result in slower induction of, or a less pronounced UPR and less efficient DNA damage repair. Vice versa, hyper-AMPylation could place cells in a primed state, able to immediately deal with any newly imposed stress. The observed changes in pathogen susceptibility, although reproducible and statistically significant, are minor, may be indirect and may not reflect the major cellular

process(es) regulated by AMPylation. Thus we propose that FIC-1 should be seen as a soft modulator, rather than a master regulator of the signaling processes discussed above.

In this study, we have described how FIC-1 AMPylates heat shock 70 family proteins, histones H2 and H3, as well as translation elongation factors and provided evidence for a link between FIC-1 mediated target AMPylation and innate immunity in *C. elegans*. Whether FIC-1 mediates target AMPylation modulates antimicrobial defense mechanisms and contributes to DNA and protein damage repair will require more extensive genetic and biochemical analysis.

Material and Methods

C. elegans worms were maintained at 20 °C on nematode growth medium (NGM) agar plates seeded with OP50 *E.coli* bacteria unless stated otherwise [40]. The following strains, mutations, integrations and extra-chromosomal arrays were used in this study:

SJ4005 (*zcIs4*[P*hsp-4*::*gfp*] V), RB1104 (*hsp-3*(*ok1083*) X), MT22519 (*nEx2219*[P*fic-1*::*GFP,* P*lin-44*::*GFP*]), MT22798 (*nEX2237*[P*fic-1*::FIC-1 E274G, P*myo-3*::mCherry]), MT22849 (*n5823* IV [*fic-1* KO]), MT23529 (*n5823* IV; *nIs734*), MT23188 (*n5823*; *nEx2318*), MT23262 (*n5823* IV; muIs109), MT23307 (*n5823* IV; *zcIs4* V), MT23265 (*n5823* IV; *zcIs13* V), MT23494 (*n5823* IV; *hsp-3*(*ok1083*) X), MT23495 (*hsp-3*(*ok1083* X; *nEx2237*)), MT23497 (*xbp-1*(*tm2482*) III; *n5823* IV), MT23498 (*xbp-1*(*tm2482*) III; *nEx2237*), MT23503 (*nIs733* [P*fic-1*::FIC-1 E274G, P*myo-3*::mCherry]), MT23506 (*zcIs4* V; *nIs733*), MT23508 (*zcIs13* V; *nIs733*), MT23527 (*daf-2*(*e1370ts*) III; *n5823* IV), MT23528 (*daf-2*(*e1370ts*) III; nIs733), MT23530 (*nEx2396*[P*hsp16.2*::*fic-1*; P*myo-3*::*mCherry*]), CB1370 (*daf-2*(*e1370ts*)) III, ZD418(*xbp-1*(*tm2482*) III, RB1104 (*hsp-3*(*ok1083*)), SJ4005 (*zcIs4* V [P*hsp-4*::GFP; lin-15(n765)]), SJ4100 (*zcIs13* V [P*hsp-6*::GFP])

Plasmid construction

P*fic-1*::FIC-1 E274G was built in a two step cloning process. First, a 1.6 kBps PCR fragment spanning the putative *fic-1* promoter was used to replace the eat-4 promoter sequence in pNB7. Next, the *fic-1* gene was amplified from *C. elegans* total cDNA and cloned into the new vector. Point mutations in *fic-1* were introduced by SOEing PCR with the mutations encoded in the respective primers. The promotor trap construct (P*fic-1*::GFP) was cloned by replacing the *fic-1* gene with a *gfp* orf. For inducible protein expression, the *fic-1* gene was cloned into pPD49.78 (Fire lab vector kit) with an additional C-terminal HA-tag introduced as part of the primer Recombinant FIC-1 protein purification for *in vitro* AMPylation assays was based on *fic-1*$_{258\text{-}508}$ cloned into a single orf of a pDUET vector. Primer sequences are available upon request. For *fic-1*(n5823) rescue, a 4.5 kbps linear DNA fragment was amplified using genomic DNA as template.

For crystallization, *fic-1*$_{134\text{-}508}$ was cloned with a non-cleavable N-terminal 6x His tag into an Ampicillin resistant T7 based bacterial expression plasmid. The transmembrane helix and membrane proximal domain of FIC-1 were excluded from both the wild type and the E274G construct.

CRISPR-mediated genome editing

CRISPR-mediated genome editing was performed essentially as previously described [41, 42]. 3 days after injection, mCherry-positive animals were separated and their F2 generation tested for homozygous *fic-1* modifications by sequencing. Lines of interest were back-crossed to wild type at least twice prior to use. MT22849 and derivate lines were routinely genotyped by PCR using 3 primers at once resulting in a 750 bps (wild type) or 950 bps (KO) product (Fig. S1c and d).

Germline transformation

Germline transformation was performed as described. The *gfp* reporter transgene (P*fic-1::GFP*) was injected at 50 ng/µl into *lin-15(n765ts)* animals with 50 ng/µl of pL15EK as a co-injection marker [43]. The rescue construct (4.5 kBps PCR fragment) was injected at 50 ng/µl into MT22433 with 10 ng/µl each of P*myo-3::mCherry* and P*rab-3::mCherry* as a co-injection markers. P*hsp-16::fic-1* was injected at 50 ng/µl into MT22433 with 10 ng/µl each of P*myo-3::mCherry* and P*rab-3::mCherry* as a co-injection markers.

Small molecular Fluorescence in situ hybridization (smFISH)

Fluorescence *in situ* hybridization was performed as described [17]. The *fic-1* smFISH probes (Biosearch Technologies, Inc) were conjugated to Cy5 fluorophores using the Amersham Cy5 Mono-reactive Dye pack (GE Healthcare). Probe sequences are available upon request. Images in Fig. 4 are maximum intensity projections of Z-stacks processed with the FFT Bandpass Filter operations in the image processing program Fiji [44].

Immunoblotting

Nematodes were washed off NGM plates, snap-frozen in liquid nitrogen and resuspended in 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM EDTA, 1% NP-40, 0.1% SDS supplemented with protease inhibitor mix (Roche). Animals were further cracked by sonication (15 pulses, 30 % maximal power), rested on ice for 30 minutes, centrifuged for 10 minutes at maximal speed, supplemented with SDS running buffer and subjected to SDS-PAGE. Proteins were transferred to a polyvinylidene difluoride (PVDF) membrane and probed with appropriate antibodies or sera; table S1 lists all antibodies used in this study. Chemiluminescent signal was detected using a Western Lightning ECL detection kit (Perkin Elmer Life Sciences) and exposure to XAR-5 films (Kodak).

Immunofluorescence staining

Adult nematodes were heat-shocked for 2 hours at 34 °C and thereafter treated with hypochlorite solution. Eggs were processed as described in [45]. All images were collected on a PerkinElmer Ultraview Multispectral Spinning Disk Confocal Microscope equipped with a Zeiss 1.4 NA oil immersion 63x objective lens and a Prior piezo-electric objective focusing device for maintaining focus. Images were acquired with a Hamamatsu ORCA ER-cooled CCD camera controlled with Metamorph software. Post-acquisition image manipulations were made using Fiji software [44].

Generation of FIC-1-specific mouse serum

BL52/B6 mice were subcutaneously primed with 100 mBL52/B6 mice were subcutaneously primed with adjuvant, boosted by intraperiotenal injection 4 weeks later.

Reporter assays
ER as well as mitochondria stress reporter tests were performed on NGM plates supplemented with 10 mR/ml tunicamycin and 35 mm/ml ethidium bromide, respectively, and scored after 24 hours (ER stress) or 72 hours (mitochondrial stress).

Survival/development assays
For tunicamycin assays, regular NGM plates were supplemented with various amounts of tunicamycin (EMD Milipore) in DMSO at a concentration of 1 mg/ml. Total DMSO volumes were adjusted to exclude solvent effects on worm development. Tunicamycin plates were incubated at 20 °C for 24 hours prior to use. *Pseudomonas aeruginosa* PA14 was grown on SKA plates as described [46]. In brief, *P. aeruginosa* was grown overnight in LB at 37 °C with shaking. The following day, 7 ml of *P. aeruginosa* culture were transferred to the center of 3 cm SKA plates, kept at room temperature for 3 hours and subsequently incubated for 24 hours at 37 °C and thereafter for 24 hours at 20 °C. For development assays, eggs harvested from hypochlorite treatment of 1-day old adults were transferred onto assay plates and thereafter incubated at 25 °C (*Pseudomonas* plates) or 20 °C (tunicamycin assays) for 72 hours; (N > 200 eggs for each strain and treatment). Animal development was scored using the following scoring classes: older than L4, L3/L4, younger than L3. For survival assays, 30-40 L4 animals were picked onto 3-4 *Pseudomonas* plates each and subsequently incubated at 25 °C. Worm survival was scored at least every 24 hours until the last animal had died. Animals were considered dead if repetitive (10x) poking with a platinum loop did not result in any visible body movement. Worms that died by exploding through the vulva or desiccating on the side of plates were censored. Data was processed using PRISM software and statistical significance was tested using a Gehan-Breslow-Wilcoxon test.

Longevity assays
L4 animals were transferred onto fresh NGM plates and subsequently picked onto new NGM plates at two-day intervals for 14 days. Animals were considered dead if repetitive poking (10x) with a platinum loop did not result in any visible body movement. Worms that died by exploding through the vulva or desiccating on the side of plates were censored.

Protein purification
Purification of recombinant $HypE_{aa187\text{-}437}$ E234G and $FIC\text{-}1_{258\text{-}508}$ for *in vitro* AMPylation assays as well as HSP-1, HSP-1 T342A, HSP-1 T496A, HSP-3, HSP-3 S370A/T371A, HSP-3 T523A was performed following methods described in [13]. eEF-1A.2, $eEF1A.2_{244\text{-}463}$, $eEF1A.2_{244\text{-}463}$ T261A $eEF1A.2_{244\text{-}463}$ T432A as well as Histone $H3_{allTtoA}$, $H3_{24\text{-}145}$ and $H3_{noSTmotif}$ were purified under denaturing conditions as described in [13].

For crystallization, FIC-$1_{134-508}$ was expressed in *E. coli* LOBSTR-RIL(DE3) (Kerafast) [47]. Transformed cells were grown at 37 °C to an OD_{600} of 0.6, the temperature was shifted to 18 °C and expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside to a final concentration of 0.2 mM for 16 hours. Cells were harvested by centrifugation at 6000 g, resuspended in lysis buffer (50 mM potassium phosphate, pH 8.0, 400 mM NaCl, 40 mM imidazole, 2 mM $MgCl_2$, 5 mM DTT, and 1 mM PMSF) and lysed using a cell disruptor (Constant Systems). The lysate was cleared by centrifugation at 10000 g for 25 minutes. The soluble fraction was incubated with Ni-Sepharose 6 Fast Flow beads (GE Healthcare) for 30 min at 4 °C. The beads were washed with lysis buffer, and the protein was eluted (250 mM imidazole, 20 mM Tris pH 8.0, 150 mM NaCl, 2 mM $MgCl_2$ and 5 mM DTT) and concentrated for further purification. Samples were purified to homogeneity via size-exclusion chromatography on a Superdex S200 26/60 column (GE Healthcare) equilibrated in running buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 2 mM $MgCl_2$, and 5 mM DTT). Finally, proteins were concentrated to 30 mg/ml, flash frozen in liquid nitrogen and stored at -80 °C until usage. FIC-$1_{134-508}$ E274G/T476A, FIC-$1_{134-508}$ E274G/T352A, FIC-$1_{134-508}$ E274G/V292D, FIC-$1_{134-508}$ E274G/I298D were purified accordingly.

Crystallization

Both wild type FIC-1 and the E274G constructs crystallized in the same condition. The initial crystal hit was obtained at a concentration of 1.8 mg/ml in Protein Complex Suite (Qiagen) screen condition F9. Larger crystals were obtained via hanging drop vapor diffusion and grew over two weeks at 18 °C in 0.1 M MES pH 6.5 and 1.1 M ammonium sulfate with a 1:1 ratio with mother liquor. Crystals were harvested and cryo-protected in mother liquor with 16% (v/v) glycerol, in the presence of 5 mM $MgCl_2$ and 5 mM AMP-PNP, although no ligand was detected in the structure.

Data collection and structure determination

Data was collected at beamline 24 at the Advanced Photon Source at Argonne National Laboratories. All data processing was done using programs provided by SBgrid [48]. Data reduction was performed with HKL2000, molecular replacement was done with PHASER, using a monomer from the human Fic-domain containing protein, HYPE (PDB code 4U07) as a search model [49]. Two non-symmetry related molecules were readily found in the asymmetric unit. The structures were manually built using Coot and refined with phenix.refine [50]. Data collection and refinement statistics are summarized in table S2.

In vitro AMPylation assays

In vitro AMPylation were performed essentially as described in [13]. The reaction was allowed to proceed for 1 hour at room temperature, followed by the addition of 1-5 μg target protein (Histones, Hsp-1, Hsp-3, eEF-1A).

For AMPylation site mapping, analogue reactions were performed using cold ATP (1 mM final concentration). Peptides were eluted using standard reverse-phase gradients. The effluent from the column was analyzed using an Orbitrap Elite (ThermoFisher) mass spectrometer (nanospray configuration) operated in a data dependent manner. The resulting fragmentation spectra were correlated against custom databases with Mascot (Matrix Science) 2.5.1 and PEAKS (Bioinformatics Solutions) 7.5 [51] using an AMP adduct mass of 329.0525 Da.

Acknowledgments
We are grateful to members of the Ploegh lab, H. Robert Horvitz and his lab and Collin Ewald for critical discussions and suggestions. Eric Spooner and the Whitehead proteomics core facility are acknowledged for assistance with mass spectrometry. We thank the international *C. elegans* Gene Knockout Consortium and the Caenorhabditis Genetics Center for strains. M.C.T. is the recipient of an Advanced Postdoc Mobility fellowship from the Swiss National Science Foundation (SNSF). C.E. is a member of the Horvitz lab. Part of this work is based upon research conducted at the Northeastern Collaborative Access Team synchrotron beamlines, which are funded by the National Institute of General Medical Sciences from the National Institutes of Health (P41 GM103403).

Author contributions
HLS supervised the project. MTC, HLS and CE planed and designed the all experiments. MTC performed all experiments. VEC, XG and TUS solved the crystal structures. MCT, HLS, VEC, CE and TUS wrote the manuscript.

Conflict of interest
The authors declare that they have no conflict of interest

Figures

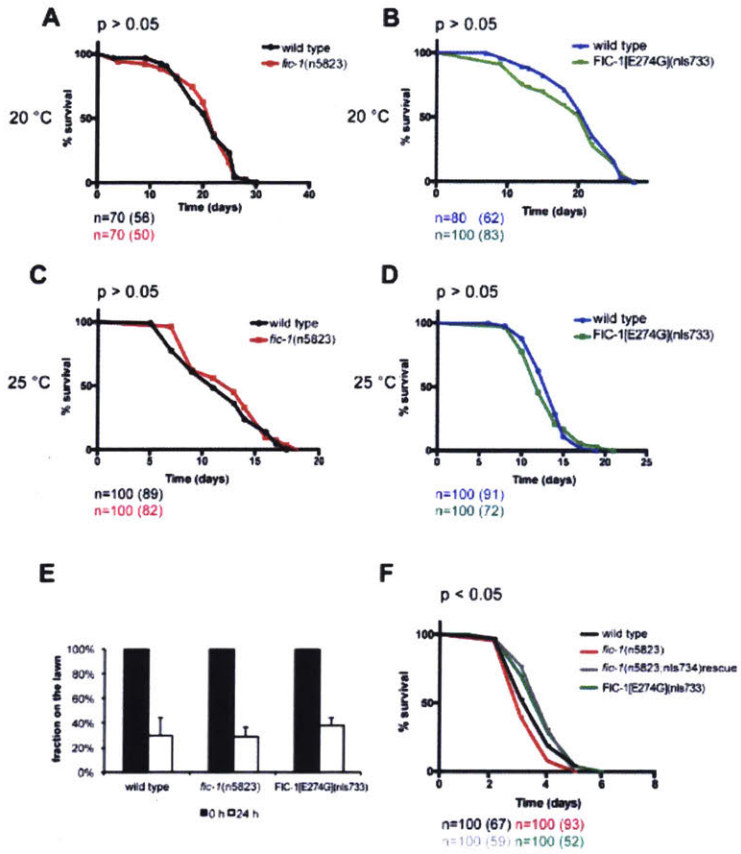


Figure 1: AMPylation plays a role in susceptibility to *P. aeruginosa* infections. AMPylation levels have no influence on aging: wild type, *fic-1*(n5823) and FIC-1[E274G](nIs733) animals were kept at either 20 C (A and B) or 25 (C and D) and survival was scored every other day. Depicted n refers to number of animals at experiment initiation; number in brackets represents total counted dead events. (E) AMPylation has no consequences on pathogen avoidance: L4 nematodes were placed in the center of a *P. aeruginosa* loan and animal localization was scored after 24 hours. (F) FIC-1[E274G](nIs733) increases while *fic-1*(n5823) decreases pathogen tolerance: L4 animals were place in the center of a *P. aeruginosa* loan and nematode survival was scored once per day until last animal vanished. Depicted n refers to number of animals at experiment initiation; number in brackets represents total counted dead events. Representative replica shown. P-values (Gehan-Breslow-Wilcoxon test) as compared to N2 wild type control: N2. vs *fic-1*(n5823): 0.046; N2 vs. *fic-1*(n5823, nIs734) rescue: 0.009; N2 vs. FIC-1[E274G](nIs733): 0.042; *fic-1*(n5823) vs. *fic-1*(n5823, nIs734) rescue or FIC-1[E274G](nIs733): <0.0001. Additional independent replica are depicted in Figure S7.

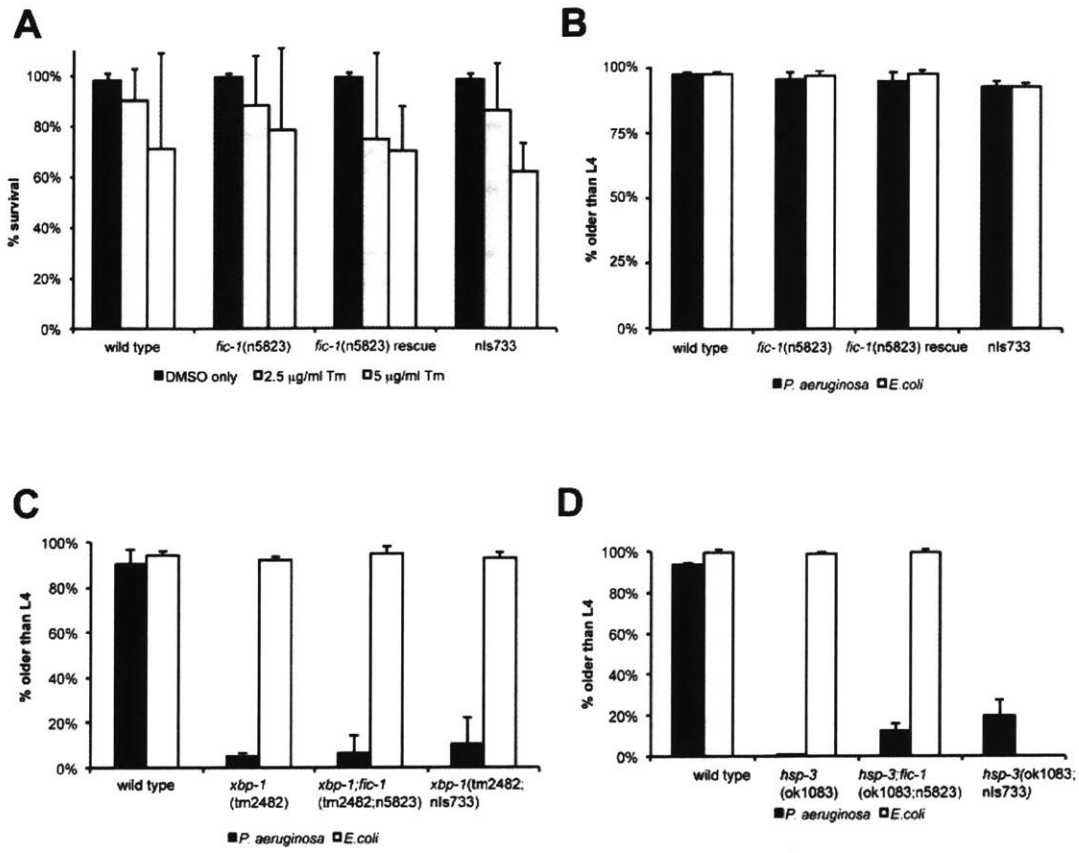


Figure 2: Hyper- or hypo-AMPylation has no apparent consequences on nematode viability and response to acute or chronic ER stress. (A) AMPylation has no influence on development under acute ER stress: eggs were transferred to OP50 plates containing different concentrations of tunicamycin to induce acute ER stress. Embryo development was scored. Average of three independent experiments shown here. (B) AMPylation has no influence on development under chronic ER stress: eggs were transferred to *P. aeruginosa* plates to induce chronic ER stress. Embryo development was scored. Average of three independent experiments shown here. (C) and (D) development assay under chronic ER stress: eggs of indicated lines were transferred to *P. aeruginosa* plates to induce chronic ER stress. Embryo development was scored. Average of three independent experiments shown here.

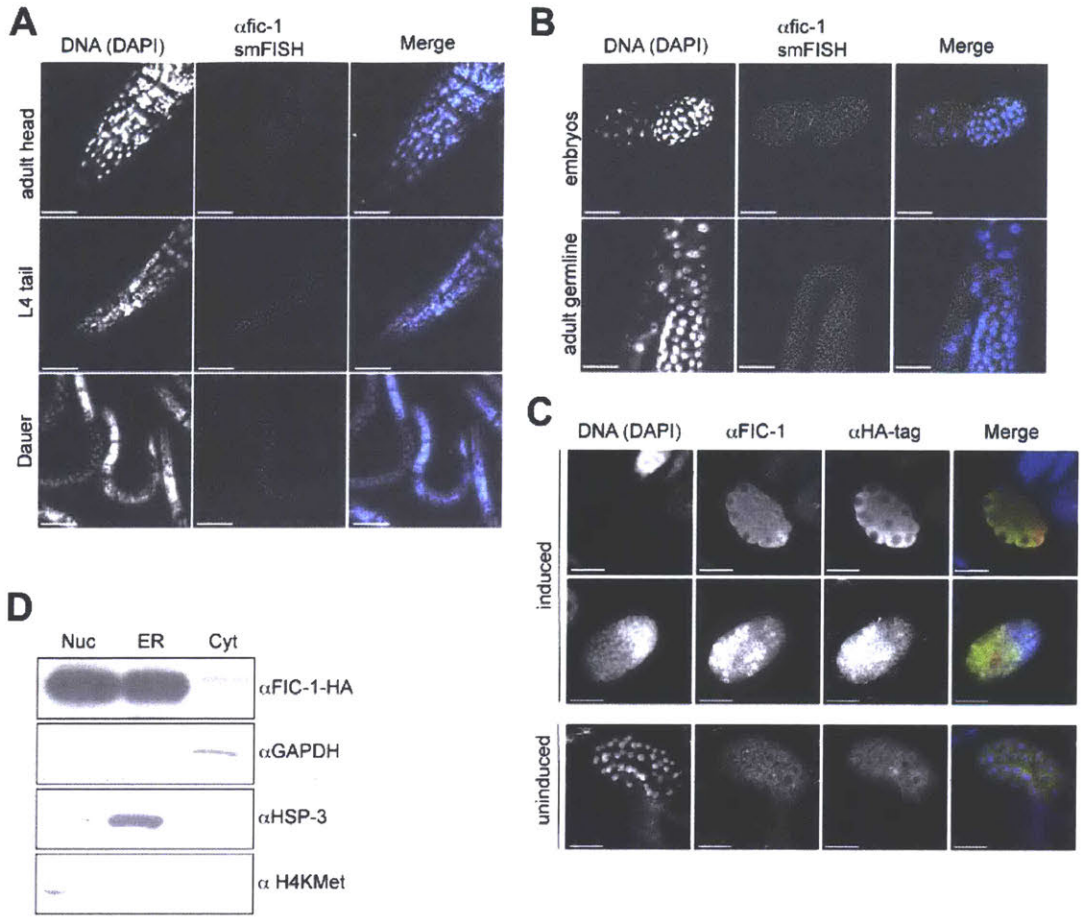


Figure 3: FIC-1 is enriched in the adult germline as well as nematode embryos and localizes to the nuclear membrane. (A) FIC-1 is expressed ubiquitously albeit at low levels: FIC-1 expression pattern as detected by smFISH analysis; samples stained with DAPI (left panel) and smFISH probe (middle panel). Right panel shows merged image where nuclei are represented in blue and smFISH signal in yellow. Distinct representative developmental stages shown here (B) FIC-1 is enriched in the adult germline and embryos: FIC-1 expression pattern as detected by smFISH analysis; samples stained with DAPI (left panel) and smFISH probe (middle panel). Right panel shows merged image where nuclei are represented in blue and smFISH signal in yellow. (C) FIC-1 preferentially localizes to the nuclear envelope/ER: embryos over-expressing HA-tagged FIC-1 were stained with indicated antibodies and dyes and FIC-1 localization was analyzed by confocal microscopy. (D) FIC-1 preferentially localizes to the nuclear envelope/ ER: sub-cellular fraction of *C. elegans* embryos. Individual fractions probed with indicated antibodies for enrichment of tested proteins. Nuc: nuclear fraction; ER: ER fraction; Cyt: cytosolic fraction.

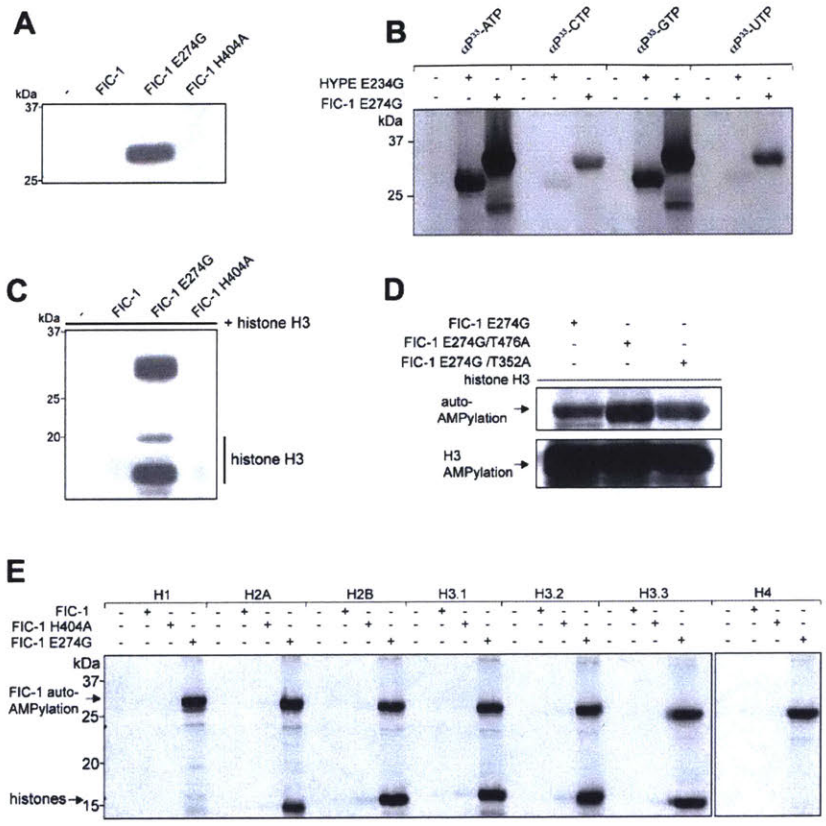


Figure 4: FIC-1 is an AMPylase. (A) FIC-1 exhibits auto-AMPylation activity: Recombinant FIC-1, FIC-1 E274G or FIC-1 H404A was incubated with α ^{33}P-ATP for an hour and incorporation of label was assessed by SDS-PAGE and autoradiography. (B) FIC-1 accepts different nucleotide substrates: FIC-1 E274G or HYPE E234G were incubated with respective α ^{33}P-labeled nucleotides for one hour at room temperature and sample autoradiography was assessed. (C) FIC-1 E274G AMPylates histone H3: Recombinant FIC-1, FIC-1 E274G or FIC-1 H404A was incubated with α ^{33}P-ATP for an hour at which point histone H3 was added and the mixture was incubated for an additional hour. Incorporation of label was assessed by SDS-PAGE and autoradiography. (D) FIC-1 E274G/T476A and FIC-1 E274G/T352A are fully active: Recombinant FIC-1 E274G, FIC-1 E274G/T476A and FIC-1 E274G/T352A was incubated with α ^{33}P-ATP for an hour at which point histone H3 was added and the mixture was incubated for an additional hour. Incorporation of label was assessed by SDS-PAGE and autoradiography. (E) FIC-1 AMPylates core histones H2 and H3 but not H4: Recombinant FIC-1, FIC-1 E274G or FIC-1 H404A was incubated with α ^{33}P-ATP for an hour at which point purified histone substrates were added and the mixture was incubated for an additional hour. Incorporation of label was assessed by SDS-PAGE and autoradiography.

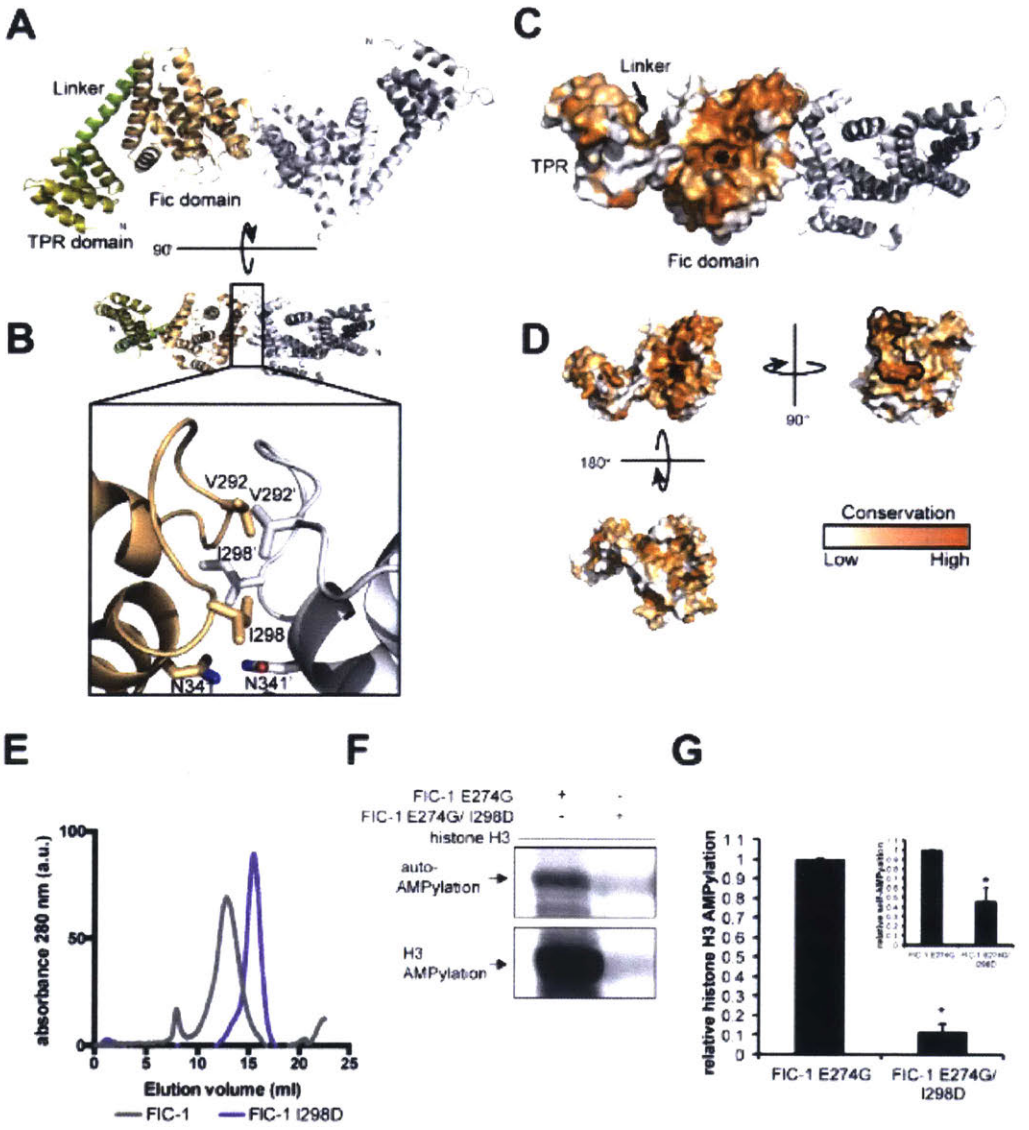


Figure 5: FIC-1 structure, domains and dimer interface. (A) Ribbon representation of FIC-1 dimer, with individual domains colored in a single monomer. (B) Cartoon representation of FIC-1; the dimer interface is highlighted in the inset where key side chain and backbone contacts are shown. (C) Surface representation of FIC-1 monomer and ribbon representation of a second monomer, the ATP binding site is highlighted with a white asterisk. (D) Surface representation of FIC-1 monomer; coloring is based on conservation from an alignment between Fic-domain containing proteins. Dimerization interface is outlined in black. (E) Size exclusion chromatogram showing elution profiles of FIC-1 wildtype (wt), and FIC-1 I298D. Elution volumes of standards are highlighted with arrows. (F) Monomeric FIC-1 E274G/I298D exposes reduced AMPylation activity: FIC-1 E274G or FIC-1 E274G/I298D were pre-incubated with

α ^{33}P-ATP for an hour before histone H3 was added and the mixture was incubated for another hour. Incorporation of label was assessed by SDS-PAGE and autoradiography. (G) Quantification of histone H3 and self-AMPylation (inlet). Data shown represents the average of two independent replica. * = p-value < 0.01 (t-test).

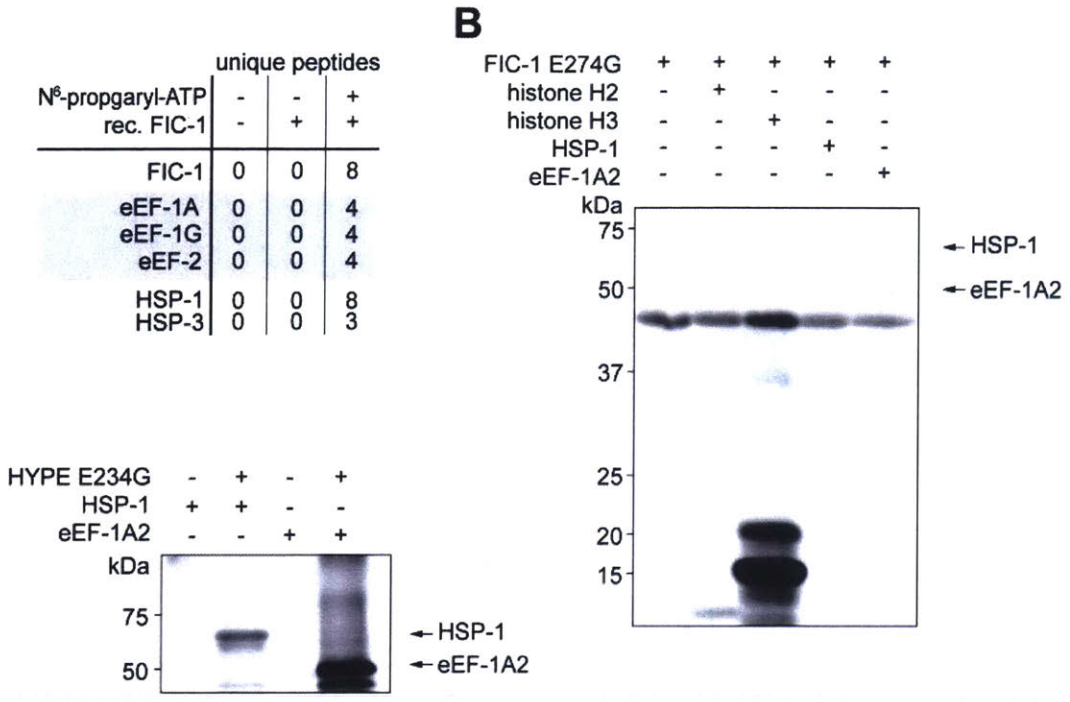


Figure 6: FIC-1 AMPylates conserved heat shock 70 family proteins and translation elongation factors. (A) Identification of new FIC-1 targets by mass spectrometry. (B) Validation of novel FIC-1 targets: Recombinant FIC-1 E274G was incubated with α ^{33}P-ATP for an hour at which point substrates (histone H3, HSP-1 or eEF-1A2) were added and the mixture was incubated for an additional hour. Sample autoradiography was assessed. (C) Novel FIC-1 targets are modified by HYPE: Recombinant HYPE E234G was incubated with α ^{33}P-ATP for an hour at which point substrates (HSP-1 or eEF-1A) were added and the mixture was incubated for an additional hour. Sample autoradiography was assessed.

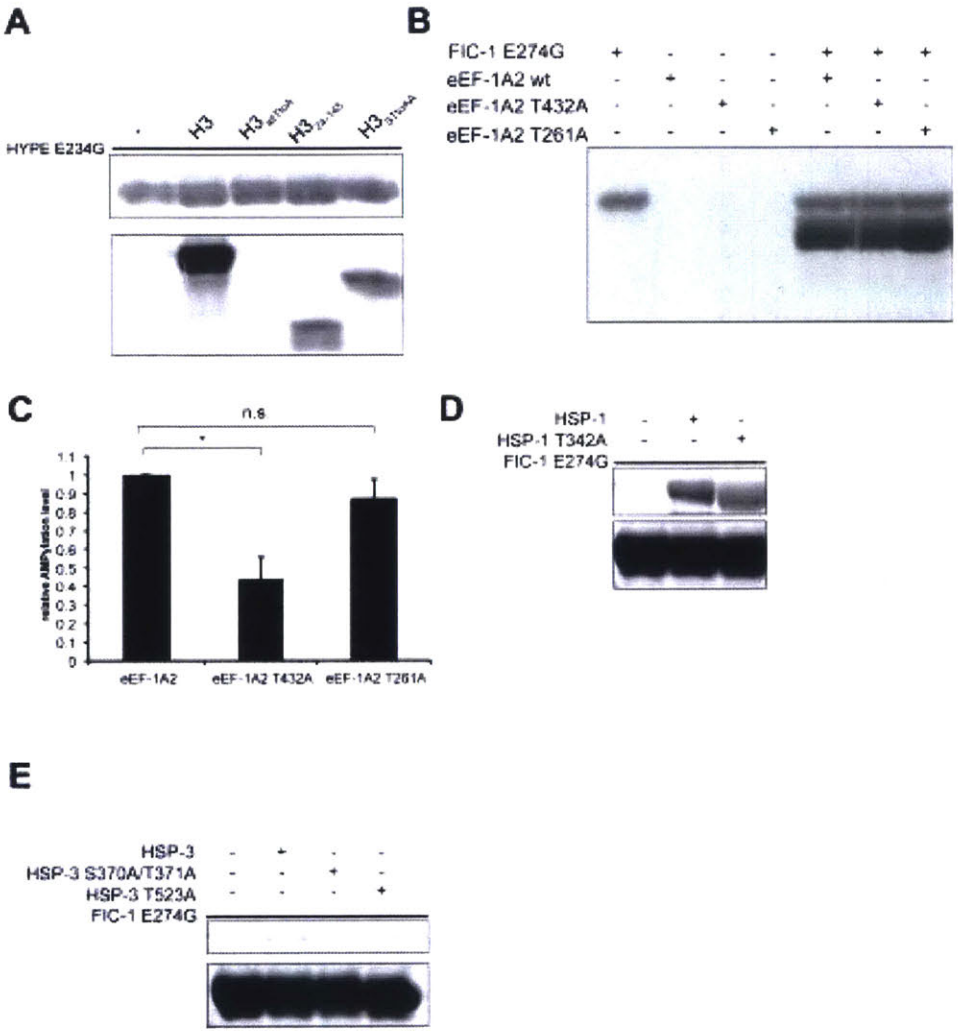


Figure 7: FIC-1 and Hype AMPylate *C. elegans* targets on multiple sites. (A) HYPE AMPylates threonines on histone H3: Recombinant HYPE E234G was incubated with α ^{33}P-ATP for an hour at which point substrates (histone H3 wild type and mutants) were added and the mixture was incubated for an additional hour. Sample autoradiography was assessed. (B-C) FIC-1 modifies eEF-1A2 on T432: Recombinant FIC-1 E274G was incubated with α ^{33}P-ATP for an hour at which point substrates (eEF-1A2$_{244\text{-}463}$ wild type and mutants) were added and the mixture was incubated for an additional hour. Sample autoradiography was assessed qualitatively (B) and quantitatively (C); data shown here represents the average of two independent replicas. (D-E) FIC-1 modifies HSP-1 and HSP-3 on distinct sites from human BiP: Recombinant FIC-1 E274G was incubated with α ^{33}P-ATP for an hour at which point substrates (HSP-1, HSP-3 and respective mutants) were added and the mixture was incubated for an additional hour. Sample autoradiography was assessed.

A

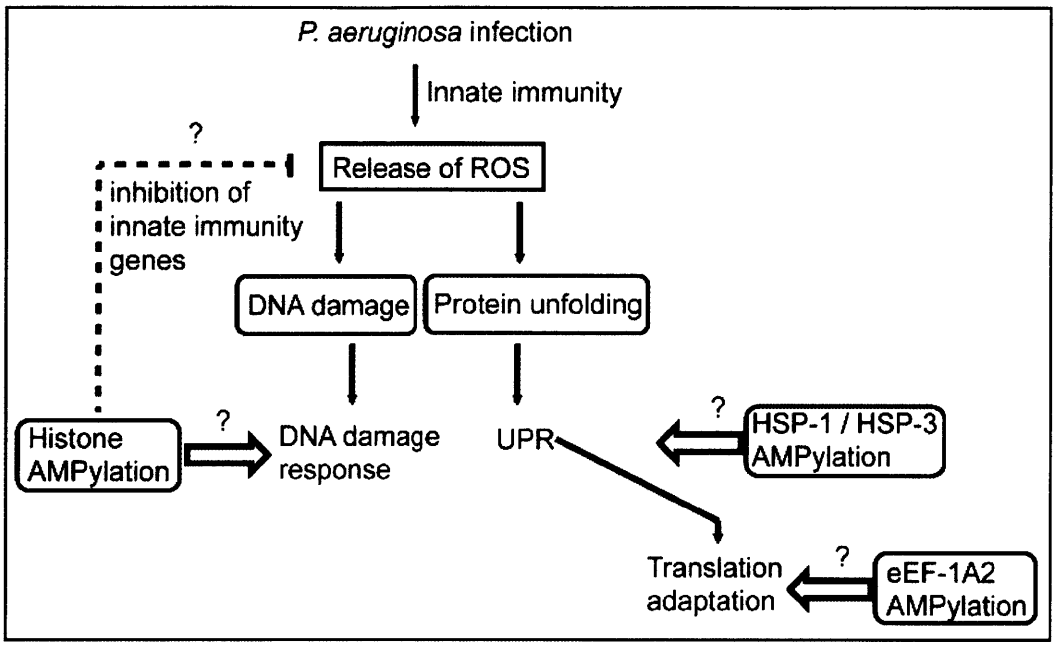


Fig 8. Proposed model.

Schematic representation of how FIC-1 might be involved in controlling antimicrobial responses in *C. elegans*.

References

1. Brown MS, Segal A, Stadtman ER. Modulation of glutamine synthetase adenylylation and deadenylylation is mediated by metabolic transformation of the P II -regulatory protein. Proc Natl Acad Sci U S A. 1971;68(12):2949-53. PubMed PMID: 4399832; PubMed Central PMCID: PMCPMC389567.
2. O'Shea JJ, Holland SM, Staudt LM. JAKs and STATs in immunity, immunodeficiency, and cancer. N Engl J Med. 2013;368(2):161-70. doi: 10.1056/NEJMra1202117. PubMed PMID: 23301733.
3. Woolery AR, Luong P, Broberg CA, Orth K. AMPylation: Something Old is New Again. Front Microbiol. 2010;1:113. doi: 10.3389/fmicb.2010.00113. PubMed PMID: 21607083; PubMed Central PMCID: PMCPMC3095399.
4. Itzen A, Blankenfeldt W, Goody RS. Adenylylation: renaissance of a forgotten post-translational modification. Trends Biochem Sci. 2011;36(4):221-8. doi: 10.1016/j.tibs.2010.12.004. PubMed PMID: 21256032.
5. Engel P, Goepfert A, Stanger FV, Harms A, Schmidt A, Schirmer T, et al. Adenylylation control by intra- or intermolecular active-site obstruction in Fic proteins. Nature. 2012;482(7383):107-10. doi: 10.1038/nature10729. PubMed PMID: 22266942.
6. Worby CA, Mattoo S, Kruger RP, Corbeil LB, Koller A, Mendez JC, et al. The fic domain: regulation of cell signaling by adenylylation. Mol Cell. 2009;34(1):93-103. doi: 10.1016/j.molcel.2009.03.008. PubMed PMID: 19362538; PubMed Central PMCID: PMCPMC2820730.
7. Feng F, Yang F, Rong W, Wu X, Zhang J, Chen S, et al. A Xanthomonas uridine 5'-monophosphate transferase inhibits plant immune kinases. Nature. 2012;485(7396):114-8. doi: 10.1038/nature10962. PubMed PMID: 22504181.
8. Cruz JW, Rothenbacher FP, Maehigashi T, Lane WS, Dunham CM, Woychik NA. Doc toxin is a kinase that inactivates elongation factor Tu. J Biol Chem. 2014;289(28):19276. doi: 10.1074/jbc.A113.544429. PubMed PMID: 25016031.
9. Castro-Roa D, Garcia-Pino A, De Gieter S, van Nuland NA, Loris R, Zenkin N. The Fic protein Doc uses an inverted substrate to phosphorylate and inactivate EF-Tu. Nat Chem Biol. 2013;9(12):811-7. doi: 10.1038/nchembio.1364. PubMed PMID: 24141193; PubMed Central PMCID: PMCPMC3836179.
10. Harms A, Stanger FV, Scheu PD, de Jong IG, Goepfert A, Glatter T, et al. Adenylylation of Gyrase and Topo IV by FicT Toxins Disrupts Bacterial DNA Topology. Cell Rep. 2015. doi: 10.1016/j.celrep.2015.07.056. PubMed PMID: 26299961.
11. Yarbrough ML, Li Y, Kinch LN, Grishin NV, Ball HL, Orth K. AMPylation of Rho GTPases by Vibrio VopS disrupts effector binding and downstream signaling. Science. 2009;323(5911):269-72. doi: 10.1126/science.1166382. PubMed PMID: 19039103.
12. Mattoo S, Durrant E, Chen MJ, Xiao J, Lazar CS, Manning G, et al. Comparative analysis of Histophilus somni immunoglobulin-binding protein A (IbpA) with other fic domain-containing enzymes reveals differences in substrate

and nucleotide specificities. J Biol Chem. 2011;286(37):32834-42. doi: 10.1074/jbc.M111.227603. PubMed PMID: 21795713; PubMed Central PMCID: PMCPMC3173180.
13. Truttmann MC, Wu Q, Stiegeler S, Duarte JN, Ingram J, Ploegh HL. HypE-specific Nanobodies as Tools to modulate HypE-mediated Target AMPylation. J Biol Chem. 2015. doi: 10.1074/jbc.M114.634287. PubMed PMID: 25678711.
14. Sanyal A, Chen AJ, Nakayasu ES, Lazar CS, Zbornik EA, Worby CA, et al. A Novel Link Between Fic (Filamentation induced by cAMP)-mediated Adenylylation/AMPylation and the Unfolded Protein Response. J Biol Chem. 2015. doi: 10.1074/jbc.M114.618348. PubMed PMID: 25601083.
15. Ham H, Woolery AR, Tracy C, Stenesen D, Krämer H, Orth K. Unfolded protein response-regulated dFic reversibly AMPylates BiP during endoplasmic reticulum homeostasis. J Biol Chem. 2014. doi: 10.1074/jbc.M114.612515. PubMed PMID: 25395623.
16. Richardson CE, Kooistra T, Kim DH. An essential role for XBP-1 in host protection against immune activation in C. elegans. Nature. 2010;463(7284):1092-5. doi: 10.1038/nature08762. PubMed PMID: 20182512; PubMed Central PMCID: PMCPMC2834299.
17. Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. Imaging individual mRNA molecules using multiple singly labeled probes. Nat Methods. 2008;5(10):877-9. doi: 10.1038/nmeth.1253. PubMed PMID: 18806792; PubMed Central PMCID: PMCPMC3126653.
18. Bunney TD, Cole AR, Broncel M, Esposito D, Tate EW, Katan M. Crystal Structure of the Human, FIC-Domain Containing Protein HYPE and Implications for Its Functions. Structure. 2014. doi: 10.1016/j.str.2014.10.007. PubMed PMID: 25435325.
19. Garcia-Pino A, Zenkin N, Loris R. The many faces of Fic: structural and functional aspects of Fic enzymes. Trends Biochem Sci. 2014;39(3):121-9. doi: 10.1016/j.tibs.2014.01.001. PubMed PMID: 24507752.
20. Grammel M, Luong P, Orth K, Hang HC. A chemical reporter for protein AMPylation. J Am Chem Soc. 2011;133(43):17103-5. doi: 10.1021/ja205137d. PubMed PMID: 21942216; PubMed Central PMCID: PMCPMC3246509.
21. Kiang JG, Tsokos GC. Heat shock protein 70 kDa: molecular biology, biochemistry, and physiology. Pharmacol Ther. 1998;80(2):183-201. PubMed PMID: 9839771.
22. Haynes CM, Ron D. The mitochondrial UPR - protecting organelle protein homeostasis. J Cell Sci. 2010;123(Pt 22):3849-55. doi: 10.1242/jcs.075119. PubMed PMID: 21048161.
23. Nilsson J, Nissen P. Elongation factors on the ribosome. Curr Opin Struct Biol. 2005;15(3):349-54. doi: 10.1016/j.sbi.2005.05.004. PubMed PMID: 15922593.
24. Andersen GR, Nyborg J. Structural studies of eukaryotic elongation factors. Cold Spring Harb Symp Quant Biol. 2001;66:425-37. PubMed PMID: 12762045.

25. Preissler S, Rato C, Chen R, Antrobus R, Ding S, Fearnley IM, et al. AMPylation matches BiP activity to client protein load in the endoplasmic reticulum. Elife. 2015;4. doi: 10.7554/eLife.12621. PubMed PMID: 26673894.
26. Chambers JE, Petrova K, Tomba G, Vendruscolo M, Ron D. ADP ribosylation adapts an ER chaperone response to short-term fluctuations in unfolded protein load. J Cell Biol. 2012;198(3):371-85. doi: 10.1083/jcb.201202005. PubMed PMID: 22869598; PubMed Central PMCID: PMCPMC3413365.
27. Gaut JR. In vivo threonine phosphorylation of immunoglobulin binding protein (BiP) maps to its protein binding domain. Cell Stress Chaperones. 1997;2(4):252-62. PubMed PMID: 9495282; PubMed Central PMCID: PMCPMC313004.
28. Singh V, Aballay A. Regulation of DAF-16-mediated Innate Immunity in Caenorhabditis elegans. J Biol Chem. 2009;284(51):35580-7. doi: 10.1074/jbc.M109.060905. PubMed PMID: 19858203; PubMed Central PMCID: PMCPMC2790988.
29. Broncel M, Serwa RA, Bunney TD, Katan M, Tate EW. Global profiling of HYPE mediated AMPylation through a chemical proteomic approach. Mol Cell Proteomics. 2015. doi: 10.1074/mcp.O115.054429. PubMed PMID: 26604261.
30. Vanfleteren JR, Van Bun SM, Van Beeumen JJ. The histones of Caenorhabditis elegans: no evidence of stage-specific isoforms. An overview. FEBS Lett. 1989;257(2):233-7. PubMed PMID: 2583267.
31. Kayne PS, Kim UJ, Han M, Mullen JR, Yoshizaki F, Grunstein M. Extremely conserved histone H4 N terminus is dispensable for growth but essential for repressing the silent mating loci in yeast. Cell. 1988;55(1):27-39. PubMed PMID: 3048701.
32. Bhasin M, Reinherz EL, Reche PA. Recognition and classification of histones using support vector machine. J Comput Biol. 2006;13(1):102-12. doi: 10.1089/cmb.2006.13.102. PubMed PMID: 16472024.
33. Rossetto D, Avvakumov N, Côté J. Histone phosphorylation: a chromatin modification involved in diverse nuclear events. Epigenetics. 2012;7(10):1098-108. doi: 10.4161/epi.21975. PubMed PMID: 22948226; PubMed Central PMCID: PMCPMC3469451.
34. Shimada M, Niida H, Zineldeen DH, Tagami H, Tanaka M, Saito H, et al. Chk1 is a histone H3 threonine 11 kinase that regulates DNA damage-induced transcriptional repression. Cell. 2008;132(2):221-32. doi: 10.1016/j.cell.2007.12.013. PubMed PMID: 18243098.
35. Blatch GL, Lässle M. The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. Bioessays. 1999;21(11):932-9. doi: 10.1002/(SICI)1521-1878(199911)21:11<932::AID-BIES5>3.0.CO;2-N. PubMed PMID: 10517866.
36. Troemel ER, Chu SW, Reinke V, Lee SS, Ausubel FM, Kim DH. p38 MAPK regulates expression of immune response genes and contributes to longevity in C. elegans. PLoS Genet. 2006;2(11):e183. doi:

10.1371/journal.pgen.0020183. PubMed PMID: 17096597; PubMed Central PMCID: PMCPMC1635533.
37. Cabiscol E, Piulats E, Echave P, Herrero E, Ros J. Oxidative stress promotes specific protein damage in Saccharomyces cerevisiae. J Biol Chem. 2000;275(35):27393-8. doi: 10.1074/jbc.M003140200. PubMed PMID: 10852912.
38. Cooke MS, Evans MD, Dizdaroglu M, Lunec J. Oxidative DNA damage: mechanisms, mutation, and disease. FASEB J. 2003;17(10):1195-214. doi: 10.1096/fj.02-0752rev. PubMed PMID: 12832285.
39. Harding HP, Zhang Y, Bertolotti A, Zeng H, Ron D. Perk is essential for translational regulation and cell survival during the unfolded protein response. Mol Cell. 2000;5(5):897-904. PubMed PMID: 10882126.
40. Brenner S. The genetics of Caenorhabditis elegans. Genetics. 1974;77(1):71-94. PubMed PMID: 4366476; PubMed Central PMCID: PMCPMC1213120.
41. Dickinson DJ, Ward JD, Reiner DJ, Goldstein B. Engineering the Caenorhabditis elegans genome using Cas9-triggered homologous recombination. Nat Methods. 2013;10(10):1028-34. doi: 10.1038/nmeth.2641. PubMed PMID: 23995389; PubMed Central PMCID: PMCPMC3905680.
42. Friedland AE, Tzur YB, Esvelt KM, Colaiácovo MP, Church GM, Calarco JA. Heritable genome editing in C. elegans via a CRISPR-Cas9 system. Nat Methods. 2013;10(8):741-3. doi: 10.1038/nmeth.2532. PubMed PMID: 23817069; PubMed Central PMCID: PMCPMC3822328.
43. Clark SG, Lu X, Horvitz HR. The Caenorhabditis elegans locus lin-15, a negative regulator of a tyrosine kinase signaling pathway, encodes two different proteins. Genetics. 1994;137(4):987-97. PubMed PMID: 7982579; PubMed Central PMCID: PMCPMC1206075.
44. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012;9(7):676-82. doi: 10.1038/nmeth.2019. PubMed PMID: 22743772; PubMed Central PMCID: PMCPMC3855844.
45. Korčeková D, Gombitová A, Raška I, Cmarko D, Lanctôt C. Nucleologenesis in the Caenorhabditis elegans embryo. PLoS One. 2012;7(7):e40290. doi: 10.1371/journal.pone.0040290. PubMed PMID: 22768349; PubMed Central PMCID: PMCPMC3388055.
46. Kim DH, Feinbaum R, Alloing G, Emerson FE, Garsin DA, Inoue H, et al. A conserved p38 MAP kinase pathway in Caenorhabditis elegans innate immunity. Science. 2002;297(5581):623-6. doi: 10.1126/science.1073759. PubMed PMID: 12142542.
47. Andersen KR, Leksa NC, Schwartz TU. Optimized E. coli expression strain LOBSTR eliminates common contaminants from His-tag purification. Proteins. 2013;81(11):1857-61. doi: 10.1002/prot.24364. PubMed PMID: 23852738; PubMed Central PMCID: PMCPMC4086167.
48. Morin A, Eisenbraun B, Key J, Sanschagrin PC, Timony MA, Ottaviano M, et al. Collaboration gets the most out of software. Elife. 2013;2:e01456. doi:

10.7554/eLife.01456. PubMed PMID: 24040512; PubMed Central PMCID: PMCPMC3771563.
49. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 2):213-21. doi: 10.1107/S0907444909052925. PubMed PMID: 20124702; PubMed Central PMCID: PMCPMC2815670.
50. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 4):486-501. doi: 10.1107/S0907444910007493. PubMed PMID: 20383002; PubMed Central PMCID: PMCPMC2852313.
51. Zhang J, Xin L, Shan B, Chen W, Xie M, Yuen D, et al. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. Mol Cell Proteomics. 2012;11(4):M111.010587. doi: 10.1074/mcp.M111.010587. PubMed PMID: 22186715; PubMed Central PMCID: PMCPMC3322562.