Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie der
Ludwig-Maximilians-Universität München

Structural and biochemical characterization of interactions centered on RNA decay factors: MTR4 and SMG1

Mahesh Lingaraju
aus
Guntur, Indien

2019

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Frau Prof. Dr. Elena Conti betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 29.11.2019.

.....

Mahesh Lingaraju

Dissertation eingereicht am 29.11.2019

Erstgutachter: Prof. Dr. Elena Conti

Zweitgutachter: Prof. Dr. Andreas Ladurner

Mündliche Prüfung am 21.01.2020

SUMMARY

The nuclear exosome is the central 3'-5' RNA degradation machinery that performs a myriad roles critical for the health of a cell. The exosome associates with the MTR4 helicase, which binds and unwinds RNA substrates that are threaded through the exosome barrel for degradation. In several cases, MTR4 is targeted to specific RNA substrates via its association with adaptor proteins. Since MTR4 is a component of several exosome adaptor complexes, it was hypothesized that it might be recognizing the adaptor proteins via a common motif. The first results section of this thesis presents a study in which I identified and characterized the interactions of the MTR4 helicase with a pre-ribosome processing adaptor, NVL and the scaffolding and MTR4 activating component of the nuclear exosome targeting complex, ZCCHC8. I identified that the N-terminal regions of NVL and ZCCHC8 contain conserved sequences resembling the arch interacting motif (AIM) of the yeast rRNA processing factors. The structural and biochemical analysis indicate that these AIM-like motifs bind the MTR4 arch domain in a manner similar to that of the AIMs described earlier in the literature. Overall, the results suggest that nuclear exosome adaptors have evolved canonical and non-canonical AIM sequences to bind to human MTR4 and demonstrate the versatility and specificity with which the MTR4 arch domain can recruit a repertoire of different RNA-binding proteins.

Recognizing RNA substrates for degradation is not only important in the nucleus but also in the cytoplasm. Nonsense mediated decay (NMD) is a cytoplasmic RNA decay mechanism which recognizes and degrades aberrant mRNA containing premature stop codons. It has also been shown to function in the regulation of physiological gene expression. SMG1, a 410 kDa PI3K related kinase, plays a crucial role in metazoan NMD by phosphorylating the UPF1 helicase. The phosphorylation of UPF1 was shown to be essential for the execution of NMD and represents the committed step of the NMD pathway. Although earlier low-resolution electron microscopic structures of human SMG1 along with some of its interacting partners were useful in gaining insight into the domain architecture of SMG1, the mechanism and regulation of SMG1 phosphorylation activity by SMG8-SMG9 remain poorly understood and are a subject of current research. The second results section presents a study, where I contributed to the characterization the *C. elegans* SMG8-SMG9 structurally and biochemically in an attempt to gain insights into the architecture of the complex and its possible biochemical role in NMD. The structure of the SMG8-SMG9 complex revealed that

the complex exists as G-domain heterodimer with nucleotide binding capabilities. In a later study, presented as the third part of the results section, I contributed to understanding of the architecture of the SMG1-SMG8-SMG9 complex. The results not only recapitulate the findings of the SMG8-SMG9 complex but also provide structural basis for the SMG8-SMG9 interaction with SMG1. The structure also revealed that inositol-6-phosphate is a constitutive component of SMG1 and seems to play a role as a critical structural co-factor. The high-resolution structure of SMG1-SMG8-SMG9 provides a basis for several follow-up structural and biochemical studies centered on the early steps of NMD.

Table of Contents

LIST OF FIGURES

INTRODUCTION

RESULTS

EXTENDED DISCUSSION

LIST OF TABLES

LIST OF ABBREVIATIONS

AIM	Arch interacting motif
ARE	AU-rich element
ATM	Ataxia-telangiectasia mutated
ATP	Adenosine triphosphate
C-AIM	Cysteine centered arch interacting motif
Cryo-EM	Cryogenic Electron Microscopy
DECID	Decay inducing
DTT	Dithiothreitol
EJC	Exon Junction Complex
eRF	Eukaryotic release factor
FAM	Fluorescein amidite
FAT	FRAP, ATM, TRRAP
FATC	FRAP, ATM, TRRAP C-terminal
FRAP	FKBP-12-rapamycin associated protein
FRB	FKBP12 Rapamycin binding domain
GST	Glutathione-S-transferase
GTP	Guanosine triphosphate
HEAT	Huntingtin, elongation factor 3, protein phosphatase 2A, and the yeast kinase TOR1
I-AIM	Isoleucine centered arch interacting motif
iCLIP	Individual nucleotide resolution crosslinking and immuno-precipitation
IP	Immuno-precipitation
IP_6	Inositol hexaphosphate
ITC	Isothermal titration calorimetry
ITS	Internal transcribed spacer
KOW	Kyprides, Ouzounis, Woese
LAP	Localization and affinity purification
MESG	2-amino-6-mercapto-7-methylpurine ribonucleoside
mRNA	Messenger ribonucleic acid
MST	Microscale thermophoresis
mTOR	Mammalian target of rapamycin
MTR4	mRNA transport regulator 4
NEXT	Nuclear exosome targeting

NGD	No-Go Decay
NMD	Nonsense mediated decay
NMR	Nuclear Magnetic Resonance
NSD	Non-Stop Decay
NVL	Nuclear VCP like
PI3K	Phosphoinositide 3-kinase
PIKK	Phosphatidylinositol 3-kinase-related kinase
PROMPTS	Promoter upstream transcripts
PTC	Premature Termination Codon
PTM	Post Transcriptional Modification
RISC	RNA induced silencing complex
SLiM	Short linear motif
SMG	Suppressor of Morphogenesis in genitalia
snoRNA	Small nucleolar ribonucleic acid
SURF	SMG1-UPF1-Release Factor
TCEP	Tris(2-carboxyethyl)phosphine
TEV	Tobacco etch virus
TPR	Tetratricopeptide repeat
TRAMP	Trf4-Air2-MTR4 polyadenylation
TRRAP	Transformation/transcription domain-associated protein
UPF	Up frame shift
W-AIM	Tryptophan centered arch interacting motif
XRN	Exoribonuclease

1.0 PREFACE

RNA surveillance, degradation and turnover are ubiquitous in all forms of life. The entire process involves a myriad of protein factors, each performing a specific yet significant role to identify and target RNA substrates at the right time. In spite of the decades of progress made in the field, several mysteries remain to be unraveled due to the sheer complexity of the RNA decay and turnover processes. The work summarized in this dissertation represents a small contribution to the enormous and challenging task of understanding RNA decay.

This dissertation is written in a cumulative style. Chapter two begins with a general introduction to the field of eukaryotic RNA surveillance and decay with a focused description of the pathways centered on the eukaryotic exosome. Interactions involving the nuclear exosome helicase MTR4 and the NMD kinase SMG1 are highlighted, as the questions addressed in the thesis are formed and developed around these two RNA decay factors. Chapter three includes the results culminated from the work[1,2,3] performed since June 2014, in the form of three original manuscripts, each including its own introduction, detailed materials and methods, and results. Chapter four and five feature an extended discussion and a brief outlook based on the results described in chapter three and other relevant publications in the field.

1. Lingaraju.M., Johnsen.D., Schlundt.A., Langer.L.M., Basquin.J., Sattler.M., Jensen.T.H., Falk.S., Conti.E. (2019). The MTR4 helicase recruits nuclear adaptors of the human RNA exosome using distinct arch-interacting motifs. Nat Comm 10, 3393.

2. Liang.L., Lingaraju.M., Basquin.C., Basquin.J., Conti.E. (2017). Structure of a SMG8-SMG9 complex identifies a G-domain heterodimer in the NMD effector proteins. RNA 23, 1028-1034.

3. Gat.Y., Schuller.J.M., Lingaraju.M., Weyher.E., Bonneau.F., Strauss.M., Murray.P.J., Conti.E. InsP6 binding to PIKKs revealed by the cryo-EM structure of a SMG1-SMG8-SMG9 complex. Nat. Struc. Mol. Biol 12. 1089-1093.

2.0 INTRODUCTION

RNA biogenesis and maintenance of steady state RNA levels is a very complex process. Carefully orchestrated processing steps often follow transcription before the RNA can be functional. The processes of transcription and post-transcriptional processing are prone to errors, creating a need for several surveillance mechanisms to identify and degrade spurious RNA, which might otherwise negatively impact cellular function. Finally, functional RNAs need to be targeted for timely degradation to facilitate efficient cellular function. Despite the seeming complexity, many RNA degradation and processing pathways share substantial similarities in the mechanism of substrate recognition and decay. At their core, all decay mechanisms involve RNA degrading enzymes (RNases), belonging to three classes, namely the 5'-3' exonucleases, the endonucleases, and the 3'-5' exonucleases. Exonucleases and endonucleases often function hand-in-hand with several other RNA decay factors which confer substrate selectivity (refer to Table 1 for a summarized list of eukaryotic RNAses and their functions).

2.1 5'-3' EXORIBONUCLEASES AND ASSOCIATED FACTORS

Of all the ribonucleases known, very few of the enzymes possess 5'-3' exoribonuclease activity. The majority of known 5'-3' exoribonucleases belong to the XRN superfamily of enzymes and play a crucial role in various cellular activities, reviewed in (Jones et al., 2012; Krzyszton et al., 2012; Nagarajan et al., 2013). XRNs are evolutionarily conserved Mg^{2+} dependent enzymes (Stevens, 1980) present in the nucleus (XRN2 and nuclear XRN1) and in the cytoplasm (XRN1). Non-XRN 5'-3' exoribonucleases like Rrp17p (Oeffinger et al., 2009) and enzymes possessing 5'-3' exonuclease and endonuclease activity like CPSF-73 (X. C. Yang et al., 2009) still function in conjunction with the XRNs (Eaton et al., 2018).

Nuclear XRNs (XRN1 and XRN2) are involved in ribosomal processing, specifically in the internal transcribed spacer (ITS)-1 trimming steps (Sloan et al., 2013). The nuclear XRN1 has been implicated in DNA recombination and chromosome stability (Z. Liu et al., 1995). XRN2 is involved in the processing of small nucleolar RNAs (snoRNAs) (C. Y. Lee et al., 2003), maturation of microRNAs (Zhang et al., 2017), and degradation of unspliced messenger RNAs (mRNAs) (Danin-Kreiselman et al., 2003) and telomeric repeat containing RNAs (Luke et al., 2008). Perhaps the most intensely studied function of XRN2 is its role in

transcription termination (Fong et al., 2015; West et al., 2004). However, the mechanism by which XRN2 functions to bring about the termination is not clear. Nuclear XRN1 and XRN2 activities are functionally interchangeable in several cases (A. W. Johnson, 1997) except in the case of transcription termination (Dengl & Cramer, 2009).

To elicit these functions, nuclear XRNs interact with associated co-factors that modulate their activity. In yeast, XRN2 is known to associate with Rai1, which moderately stimulates its activity (Xiang et al., 2009). Although higher eukaryotes possess respective Rai1 homologs (Xue et al., 2000), their role in modulating XRN2 activity remains poorly understood. XRN2 is also a component of the ternary complex TXT (Twi1-XRN2-Tan1), which plays a role in tRNA degradation and rRNA processing (Couvillion et al., 2012). In addition, some co-factors contain a conserved XRN2 binding domain, which possibly allows them to stabilize XRN2-substrate interactions, thus modulating XRN2 activity via a generic mechanism (Miki et al., 2014; Richter et al., 2016).

In the cytoplasm, XRN1 is found localized to P-bodies (Kulkarni et al., 2010; Sheth & Parker, 2003) where it is involved in all the major cytoplasmic mRNA decay mechanisms (detailed in section 2.5). There, it associates with the decapping machinery facilitating mRNA decay once decapping takes place (Coller & Parker, 2004). Cytoplasmic XRN1 also plays a crucial role in antiviral response where it is believed to interact with viral gag protein to target viral RNA for decay (Rowley et al., 2016) and prevent viral replication (Y. Li et al., 2015).

2.2 ENDONUCLEASES

Most cellular RNAs are modified at the 5' and 3' ends to protect them against exonucleases thereby extending their half-life. In many such cases, endonucleases are required to target and perform initial cleavage reactions in order to free the RNA to be processed by the exonucleases. Endonucleolytic activity is also critical for several RNA processing steps. In the nucleus, the endonucleolytic activity of Rnt1 is required for efficient termination of PolI and PolII transcription (El Hage et al., 2008). Rnt1 is also involved in the processing of snoRNAs (Chanfreau et al., 1998) and is specifically implicated in initiation of the 5' end processing of C/D box snoRNA (Grzechnik et al., 2018). The PIN domain endonucleases, UTP23, UTP24 and NOB1 are involved in rRNA processing, where their activities are

required for 18S RNA maturation (An et al., 2018; Lamanna & Karbstein, 2009; Wells et al., 2017). Endonucleolytic activity of the cleavage and polyadenylation complex (CPSF) is required for site-specific cleavage of pre-mRNA prior to addition of the polyA tail (Ryan et al., 2004).

Endonucleolytic activity is central to protein complexes involved in microRNA pathways. The endonuclease activity of RNA-induced silencing complex (RISC) is believed to be predominantly cytoplasmic (Karginov et al., 2010), although endonucleases involved in the microRNA pathways also have roles in the nucleus (Gagnon et al., 2014). Human PMR1 (Gu et al., 2012) acts in the microRNA pathway (Gu et al., 2016) upstream of Dicer processing and is activated in an estrogen-dependent manner. APE1 endonuclease, involved in *c-myc* mRNA regulation (W. C. Kim et al., 2010), is also believed to regulate Dicer activity.

Although there is substantial indirect evidence of endonucleolytic activity in various mRNA quality control pathways in the cytoplasm (Arribere & Fire, 2018; Doma & Parker, 2006), very few bonafide endonucleases have been characterized. SMG6 is a bonafide endonuclease of the NMD pathway (Eberle et al., 2009; Gatfield et al., 2003; Huntzinger et al., 2008) (detailed in section 2.5.2). Here it displays targeted endonucleolytic activity by the way of a 14-3-3-like domain to exert selectivity and a PIN domain for catalytic activity (Glavan et al., 2006). Recently, Cue2, another endonuclease, was shown to be recruited to stalled ribosomes and promote No-Go decay (NGD) in yeast (D'Orazio et al., 2019; Glover et al., 2019).

IRE1 endonuclease (K. P. Lee et al., 2008) and the isoforms of the Zc3h12 zinc finger endonuclease, (Matsushita et al., 2009) play a role in the ER stress response and immune response, respectively, by regulating specific mRNA levels to elicit various signaling pathways.

Table 2.1: Summary of the major eukaryotic RNA degradation factors & their known functions

Enzymes	Associated co-factors	Functions
5'-3' decay		
XRN1		Ribosomal processing
		DNA recombination and chromosome stability
		Cytoplasmic RNA decay
XRN2	Rai1	Small nucleolar RNA degradation
	Twi1-Tan1	tRNA degradation and rRNA processing
		Transcription termination
Rrp17		Ribosomal processing
CPSF-73	Component of cleavage and polyadenylation complex	Histone pre-mRNA processing (putative)
Endonucleases		
Rnt1		Transcription termination
UTP23, UTP24, NOB1		rRNA processing
CPSF-73	Component of cleavage and polyadenylation complex	mRNA processing
RISC, PMR1		Micro RNA processing
APE1	Dicer	c-myc mRNA regulation
SMG6	UPF1	Nonsense mediated decay
Cue2		No-go and Non-stop decay pathways
IRE1, Zc3h12		ER stress response, immune response
3'-5' decay		
Dss1	Component of mitochondrial degradosome	Mitochondrial RNA degradation
Pan2	Pan3	Deadenylation
CCR4, Caf1	Components of the CCR4-NOT complex	Deadenylation
Rrp44	Component of the exosome	General 3'-5' decay functions

2.3 3'-5' EXORIBONUCLEASES AND ASSOCIATED FACTORS

3'-5' exoribonucleases often function together with the 5'-3' exoribonucleases and endonucleases, thereby completing the arsenal of eukaryotic RNA decay enzymes. Even though their roles appear to be redundant, they are indispensable and form a key aspect of the RNA metabolism (Ibrahim et al., 2008). There are primarily three distinct classes of 3'-5' exoribonucleases. The RNAse II enzymes are non-specific and have a preference for single-stranded substrates. They are represented in eukaryotes by the Rrp44/DIS3 subunit of the exosome (Lorentzen et al., 2008; Robinson et al., 2015). The functional aspects of the exosome and its components will be detailed in later sections. Dss1 is another RNAse II superfamily enzyme present as a component of the mitochondrial degradosome (Razew et al., 2018). The other two classes of 3'-5' exoribonucleases, which are beyond the scope of this thesis, are comprised of the DEDD and EEP superfamilies (Goldstrohm & Wickens, 2008). These enzymes are predominantly deadenylases and carry out most of their functional roles in the cytoplasm.

2.4 THE EXOSOME – ARCHITECTURE

The RNA exosome is an RNA decay complex contributing to the majority of the 3'-5' exoribonuclease activity targeting a variety of RNA substrates (Pefanis et al., 2014; Schneider et al., 2012). It is the most extensively characterized 3'-5' RNA decay factor since its discovery in yeast (Mitchell et al., 1997). The RNA exosome is ubiquitously present in the nucleus and the cytoplasm with varying subunit compositions. A scaffold of nine catalytically inactive subunits (Exo9) along with a 3'-5' exonuclease form the common core of the exosome. The Exo9 is comprised of six RNase PH-like domain-containing proteins (Rrp41, Rrp42, Rrp43, Rrp45, Rrp46, and Mtr3), which form the barrel of the exosome (Q. Liu et al., 2006; Makino & Conti, 2013). Three S1/KH RNA-binding domain-containing proteins (Rrp4, Rrp40, and Csl4), cap the barrel from the top. The tenth subunit, Rrp44/Dis3 is a Mg^{2+}-dependent 3'-5' exoribonuclease, which binds at the base of the Exo9 barrel and contributes to the central catalytic activity of the exosome (Dziembowski et al., 2007) (Fig. 2.1a & 2.1b). Additionally, Rrp44 possesses an endonucleolytic activity provided by its PIN domain (Schaeffer et al., 2009; Schneider et al., 2009). Notably in humans, there are two homologs of the yeast Rrp44 – nuclear DIS3 and cytoplasmic DIS3L, resulting in compartment-specific variants of the exosome core (Tomecki et al., 2010).

In addition to the core components, the nuclear exosome interacts with three more factors, catalytic Rrp6, Rrp47 (C1D in humans), and Mpp6 (Makino et al., 2015; Schuller et al., 2018) (Fig. 2.1a). Rrp6 is a distributive 3'-5' exonuclease. It binds to the top of the exosome barrel and is believed to control substrate threading to Rrp44 (Wasmuth et al., 2014). While Rrp6 is an integral component of the nuclear exosome, it might have cytoplasmic functions apart from the exosome (Tomecki et al., 2010). Rrp6 interacts tightly with Rrp47 and together they help to recruit the helicase Mtr4 to the exosome (Schuch et al., 2014). Mpp6 is recruited to the exosome via Rrp40 (Falk, Bonneau, et al., 2017) and is also believed to stabilize the interaction of Mtr4 to the exosome core (Gerlach et al., 2018; Schuller et al., 2018; Zinder & Lima, 2017). Mpp6 has been reported to be required for Rrp44-dependent degradation (K. Kim et al., 2016).

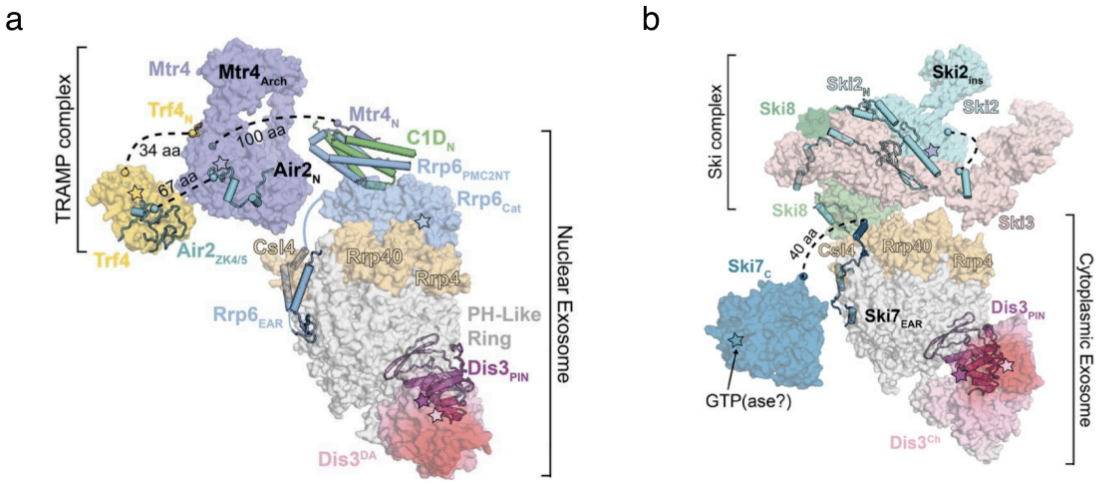


Figure 2.1: Representative architecture of the yeast nuclear and cytoplasmic exosomes. *a) A possible model of the exosome barrel together with C1D (Rrp47 in yeast), MTR4 helicase and associated cofactors shown to represent the nuclear exosome. b) A speculatory model of the core exosome barrel together with the Ski-complex shown to represent the functional cytoplasmic assembly of the exosome. The figure has been adapted from (Zinder & Lima, 2017)*

2.4.1 HELICASES IN EXOSOME-DEPENDENT RNA DECAY

RNA substrates of the exosome often have complex secondary and tertiary structures requiring the RNA to be linearized before it can be threaded through the narrow entry pore of the exosome barrel. While bacterial degradosomes possess helicase activity (RNAseR, RhlB, RhlE) to deal with the complication of structured RNA (Khemici & Linder, 2018), the eukaryotic degradosomes recruit helicases that belong to the Ski2-like helicase branch of the SF2 superfamily (Jankowsky et al., 2011). The cytoplasmic exosome functions together with the Ski complex (Ski2-Ski3-Ski8) with Ski2 providing the helicase activity (Fig. 2.1b) (Anderson & Parker, 1998; Halbach et al., 2013) whereas the nuclear exosome functions together with the Mtr4 helicase (Fig. 2.1a) (Schuch et al., 2014; Schuller et al., 2018).

Ski2 and Mtr4 are DExH helicases with 3'-5' unwinding activity (Fig. 2.2) and share a very similar architecture (Halbach et al., 2013; Halbach et al., 2012; Weir et al., 2010). They both harbor two highly conserved and essential RecA-like domains with sequence motifs involved in interactions with ATP and nucleic acid. In addition, both Ski2 and Mtr4 contain helical domains called the winged helix and ratchet domain that pack against the RecA domains in a manner reminiscent of Hel308 (Buttner et al., 2007). These additional helical domains in the exosome helicases seem to be functionally similar to their analogs in Hel308. In yeast Mtr4, the ratchet domain aids in RNA binding, and its deletion compromises the helicase activity (Holub et al., 2012; Taylor et al., 2014). The nucleic acid binding properties of the conserved residues in the ratchet and helical domain region are believed to enforce directionality of unwinding (Buttner et al., 2007). In addition to the helical core, both Ski2 and Mtr4 harbor a large insertion domain comprised of a helical stalk and a globular β-barrel domain (Fig. 2.2a & 2.2b). Although the sequence of the insertion domain is poorly conserved between the two helicases, its architecture is very similar. The insertion domain of the helicases has been shown to be involved in RNA binding and aid in substrate loading (Halbach et al., 2012; Weir et al., 2010).

The unwinding mechanism of Ski2 and Mtr4 is believed to be similar to that of Hel308, owing to highly similar structural features. A proposed model implicates that the helicases unwind the RNA in steps similar to the inchworm model of unwinding activity (Buttner et al., 2007; Tanner & Linder, 2001). ATP hydrolysis induces the conformational changes required for the translocation. Consistent with this model, single molecule studies

conducted with yeast Mtr4 have shown that the helicase unwinds RNA duplexes in steps of six base pairs and can perform only a single step of unwinding in the presence of a non-hydrolysable ATP analog (Patrick et al., 2017). However, the exact mechanism of strand splitting in Ski2 and Mtr4 still remains to be unraveled.

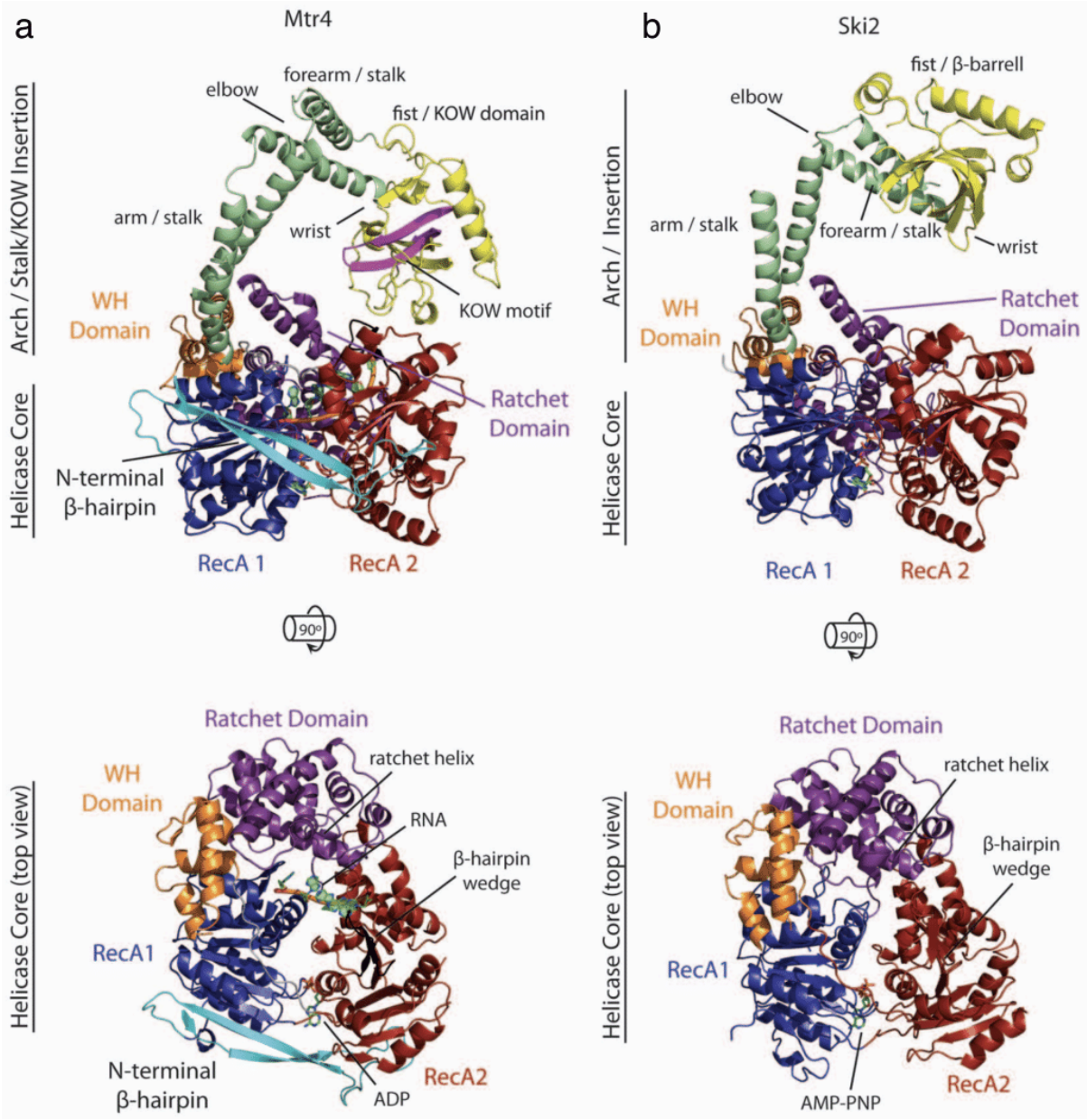


Figure 2.2: A comparison of the domain architecture of Mtr4 and Ski2 helicases

a) A representation of the crystal structure of the Mtr4 helicase (PDB 2XGJ) highlighting the individual domains of the helicase. b) A representation of the crystal structure of Ski2 helicase (PDB 4A4Z) highlighting the individual domains of the helicase. The figure has been adapted from (S. J. Johnson & Jackson, 2013).

Ski2 and Mtr4 interact with the exosome and provide a single-stranded 3' end of the RNA to be threaded through the barrel. Ski2 is constituent of the Ski complex (Halbach et al., 2013) and interacts with the exosome via Ski7 (Kowalinski et al., 2016). In the nucleus, Mtr4 is recruited to the exosome via the nuclear specific components, Rrp6-Rrp47 (Schuch et al., 2014) and Mpp6 (Falk, Bonneau, et al., 2017). Recent structural and biochemical studies have shown that the interactions of Mtr4 with the exosome lead to the formation of a continuous substrate-channeling path from the helicase through the barrel of the exosome (Gerlach et al., 2018; Schuller et al., 2018; Weick et al., 2018).

2.4.2 FUNCTIONS OF THE NUCLEAR EXOSOME

The nuclear exosome performs a variety of functions in the nucleus owing to a diversity of RNA substrates that need to be processed (Fig. 2.3). This section presents a brief summary of the various nuclear exosome targets. The first known function of the nuclear exosome is its role in pre-ribosomal RNA processing (Mitchell et al., 1997). In yeast, the activity of the nuclear exosome is required for the degradation of the 5' external transcribed spacer (ETS) in the process of maturation of 18S ribosomal RNA (de la Cruz et al., 1998). It is also implicated in the trimming of 21S pre-rRNA (Preti et al., 2013; Sloan et al., 2013). Finally, the nuclear exosome is also involved in the maturation of the pre-60S particle where it is responsible for the 7S to 5.8S rRNA processing step (Briggs et al., 1998; Schuller et al., 2018). Apart from rRNA, the nuclear exosome aids in the processing of other stable RNAs such as small nuclear RNAs, small nucleolar RNAs, and tRNAs (Allmang et al., 1999).

Another important function of the exosome is the degradation of RNAs produced by cryptic transcription (Szczepinska et al., 2015; Wyers et al., 2005). The exosome has been shown to target cryptic unstable transcripts (CUTs), promoter upstream transcripts (PROMPTs), and enhancer RNAs (Preker et al., 2008; Szczepinska et al., 2015). In general, the exosome is also involved in the turnover of non-coding RNAs.

The nuclear exosome is involved in the quality control of mRNAs at various levels of mRNA processing. It degrades transcripts that terminated aberrantly and pre-mRNAs with retained introns (Bousquet-Antonelli et al., 2000; Schneider et al., 2012). It can as well target aberrant mRNPs that accumulate due to RNA packaging errors, as observed in yeast with mRNA export defects (Rougemaille et al., 2007). Apart from targeting aberrant mRNA, the

nuclear exosome is implicated in the regulation of gene expression. For instance, in *S. pombe*, the nuclear exosome along with associated specificity factors targets meiotic RNAs that are generated during mitotic growth (Harigaya et al., 2006). Furthermore, the exosome plays a role in spliceosome-mediated decay, degrading unstable products emerging from splicing of intron-less transcripts (Volanakis et al., 2013).

Apart from direct regulation of RNA levels, the nuclear exosome is believed to play an indirect role in the DNA damage response. The activity of the nuclear exosome is downregulated in response to agents that cause DNA damage thereby allowing for the stabilization of certain transcripts leading to the DNA damage response (X. Wang et al., 2008). Moreover, the exosome is involved in maintaining genomic integrity by targeting R-loops that make the genome vulnerable to double strand breaks (X. Li & Manley, 2006).

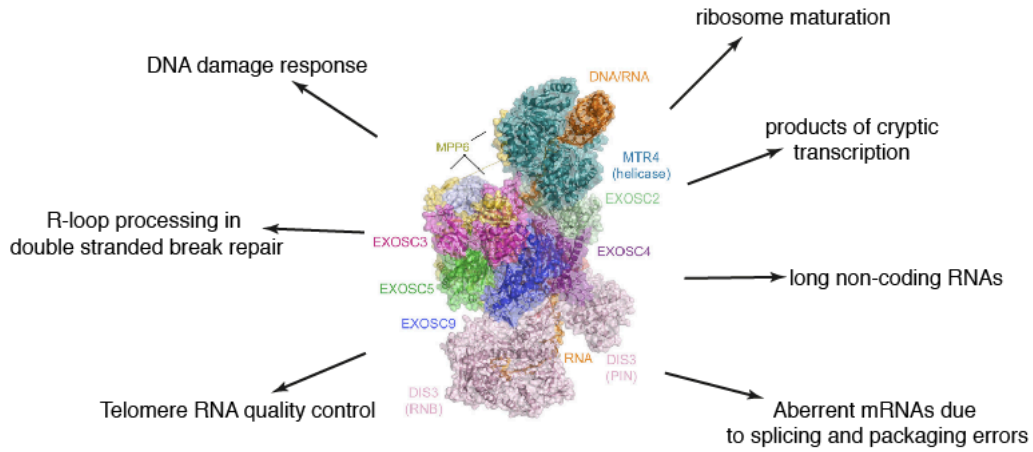


Figure 2.3: A summary of the functions of the eukaryotic nuclear exosome.
The individual components of the nuclear exosome are highlighted and labeled in the center of the figure as identified in (Weick et al., 2018). Several functions of the nuclear exosome are highlighted.

2.4.3 SUBSTRATE RECOGNITION – ROLE OF EXOSOME ADAPTORS

The diversity of the substrates that are processed or degraded by the exosome, especially in the nucleus, leads to an interesting conundrum. How can the exosome selectively recognize and recruit its substrate at the right time? Since the nuclear exosome relies on the helicase Mtr4 to unwind and 'feed' the substrates into the exosome barrel, Mtr4 has evolved to interact with several adaptor proteins that in turn target substrates to Mtr4 for unwinding (Fig. 2.4).

The Trf4-Air2-Mtr4 polyadenylation (TRAMP) complex is among the first discovered and characterized exosome adaptors (LaCava et al., 2005; Vanacova et al., 2005). In *S. cerevisiae*, the complex is also present as TRAMP5 where Mtr4 associates with Trf5 and Air1. Air1/2 are zinc finger proteins likely aid in substrate binding while Trf4/5 function as polymerases which add short poly(A) stretches to the 3' ends of RNA substrates (Holub & Vanacova, 2012; Schmidt & Butler, 2013). It is believed that these overhangs help load the helicase Mtr4 on to the structured RNA substrate to trigger unwinding and subsequent exosome threading (Jia et al., 2012). The TRAMP complex offers selectivity towards aberrant tRNAs (Kadaba et al., 2006), small nuclear and nucleolar RNA (Carneiro et al., 2007; Grzechnik & Kufel, 2008) and aberrant RNA polymerase II products (Tudek et al., 2014; Vasiljeva & Buratowski, 2006) via the Nrd1-Nab3-Sen1 complex through an interaction mediated via Trf4. The TRAMP complex in budding yeast has a similar architecture to that of fission yeast, albeit with a seemingly more specialized function (Keller et al., 2010; Larochelle et al., 2012). A TRAMP-like complex is also present in metazoans and is composed of MTR4, PAPD5 (polymerase) and ZCCHC7 (Zinc finger protein) (Lubas et al., 2011). However, the human TRAMP complex remains poorly characterized structurally and functionally.

In addition to the TRAMP complex, Mtr4 interacts with the ribosomal biogenesis factors Nop53 and Utp18, targeting the exosome to the pre-ribosome as indicated by studies in *S. cerevisiae* (Thoms et al., 2015). Interaction of Mtr4 with Nop53 facilitates the trimming of the 5.8S rRNA extension, which is required for maturation of the pre-60S particle (Falk, Tants, et al., 2017; Schuller et al., 2018). The Mtr4-Utp18 interaction, on the other hand, is believed to be involved in the degradation of the 5' ETS (Thoms et al., 2015). In addition to Nop53 and Utp18, human MTR4 was shown to interact with early ribosome biogenesis

factors, WDR74 and NVL2 that take part in ITS1 processing (Hiraishi et al., 2018; Hiraishi et al., 2015). However, the mechanisms of the role of the exosome in 21S pre-rRNA and ITS1 processing remain to be elucidated (Sloan et al., 2013).

Metazoan MTR4 complexes, specifically human MTR4, are much more diverse. The nuclear exosome-targeting complex (NEXT) is one of the central factors targeting various RNAs like PROMPTs, snRNA, and snoRNA to the exosome. This metazoan-specific complex is comprised of an RNA-binding protein, RBM7 (Hrossova et al., 2015), linked to MTR4 by the Zn-finger scaffold protein, ZCCHC8 (Falk et al., 2016; Hrossova et al., 2015; Lubas et al., 2011). iCLIP analysis revealed that RBM7 interacts with newly synthesized RNA indicating that NEXT functions in conjunction with several other RNA processing events (Hrossova et al., 2015; Lubas et al., 2015). Apart from providing RNA targeting abilities to NEXT, RBM7 and ZCCHC8 also promote helicase activity of MTR4 (Puno & Lima, 2018). Phosphorylation of the NEXT complex regulates its RNA binding capabilities (Tiedje et al., 2015) allowing for the regulation of NEXT-dependent exosomal activity.

MTR4-ZFC3H1 represents yet another important exosome adaptor complex that possibly targets the exosome to transcripts containing poly(A) tails via the nuclear polyA binding protein (PABPN1) (Meola & Jensen, 2017; Ogami et al., 2017). A homologous interaction is also observed in *S. pombe* where Mtl1, an MTR4-like helicase and Red1, a multi domain zinc finger protein, along with several other factors form a large complex called MTREC (Zhou et al., 2015). Many factors of the MTREC complex have homologs in metazoans suggesting that the interaction space of ZFC3H1 could quite possibly be very complex. However, the validity of these potential complexes remains to be verified. Interestingly, MTR4-ZFC3H1 seems to function via a similar targeting mechanism as that of the NEXT complex (Silla et al., 2018). Through this complex, the exosome seems to be targeted towards pre-mRNAs and functionally competes with the export pathway to degrade retained transcripts. Taken together, NEXT and MTR4-ZFC3H1 seem to operate as if exosome targeting and degradation is the default fate of transcription, allowing them to target a wide variety of substrates, while still allowing for some degree of regulation (Bresson & Tollervey, 2018).

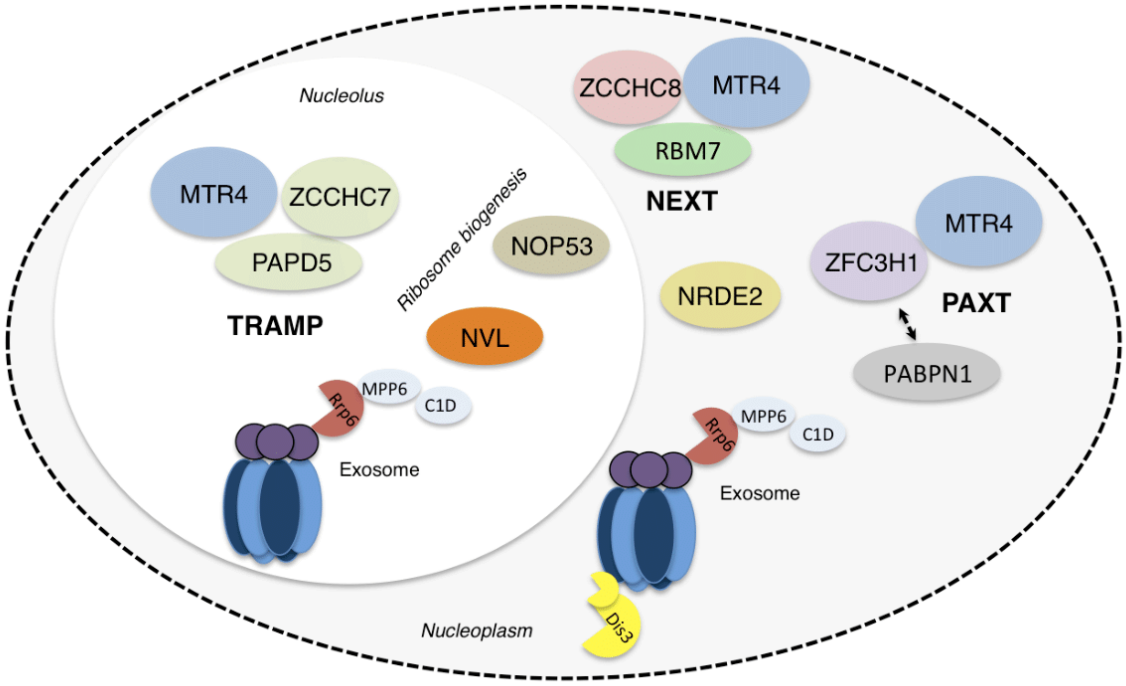


Figure 2.4: Human nuclear exosome adaptors and their localization.

The figure represents a graphical visualization of the subnuclear localization of exosome components, MTR4 and associated complexes. The TRAMP complex and MTR4 interacting ribosome biogenesis factors are localized to the nucleolus whereas NEXT, PAXT and exosome regulator NRDE2 are localized to the nucleoplasm.

2.4.4 EXOSOME ADAPTOR INTERACTION HOTSPOTS ON THE MTR4 HELICASE

While a complete picture of how Mtr4 could recruit several different factors is lacking, a wealth of structural and biochemical work has been performed in an effort to address this question. Examination of the existing literature suggests that Mtr4 is recruited to the various complexes through the catalytically active DExH core aided by the arch domain (Falk, Tants, et al., 2017; Falk et al., 2014; Schuller et al., 2018; Thoms et al., 2015).

In yeast, Trf4 and Air2 interact with Mtr4 at the DExH core (Falk et al., 2014). Interestingly, NRDE2, a novel metazoan MTR4 interaction partner has also been shown to interact at the same site as Trf4 (J. Wang et al., 2019). Furthermore, the C-terminal domain of ZCCHC8 interacts at the same site as MPP6 indicating the presence of yet another hotspot on the DExH core of MTR4 (Puno & Lima, 2018; Weick et al., 2018).

The arch domain of MTR4 has also been demonstrated to directly recruit ribosome biogenesis factors, Nop53 and Utp18 via a conserved arch interacting motif (AIM) (Thoms et al., 2015). Biochemical and crystallographic evidence showed that a conserved arginine in the Mtr4 arch is crucial for the interaction with the AIM-containing Nop53 (Falk, Tants, et al., 2017). The presence of the AIM in multiple unrelated Mtr4-interacting proteins, namely ribosome biogenesis factors Nop53 and Utp18, a zinc finger protein which is a component of the TRAMP complex, Air2, and a G-patch domain protein, Sqs1, suggests that it could function as a versatile motif to recruit several other MTR4-exosome adaptors (Losh & van Hoof, 2015; Thoms et al., 2015). Chapter 3.1 of this thesis provides further evidence in support of this hypothesis.

2.4.5 FUNCTIONS OF THE CYTOPLASMIC EXOSOME

The cytoplasmic exosome functions together with the Ski complex and plays a redundant role in cytoplasmic mRNA turnover. While XRN1 elicits the 5'-3' decay pathway following decapping, the exosome elicits 3'-5' decay following deadenylation. Studies in yeast have shown that at least one of the decay pathways needs to be viable for survival. Quantification of decay rates in yeast indicate that exosome-dependent decay is much slower than decapping-dependent decay. This is thought to be because of the rate limiting nature of deadenylation compared to decapping, reviewed in (Labno et al., 2016; Schaeffer et al., 2011). In addition to its role in canonical mRNA turnover, the cytoplasmic exosome plays a key role in antiviral defense. Typical eukaryotic mRNAs have protective features like the poly(A) tail and the cap structure whereas many viral RNAs lack these features. Many host defense mechanisms rely on these differences to efficiently target and clear the invading RNAs. Similarly, exosomal targeting of viral RNAs could be imagined as an innate immune response. In fact, the first described function of the Ski complex is related to antiviral defense (Widner & Wickner, 1993). Recently, it was shown that the TRAMP-exosome, which is canonically a nuclear complex, migrates to the cytoplasm in response to the viral infection (Molleston et al., 2016).

In ways reminiscent of its nuclear function, the exosome plays a crucial role in mRNA surveillance in the cytoplasm. The exosome and XRN1 rely on specific factors that recognize and target aberrant mRNA for degradation. The degradation of these faulty mRNA is much more rapid and allows for “easier” exosome-based degradation because the rate

limiting step of deadenylation is bypassed by endonucleases. There are three major mRNA surveillance pathways in the cytoplasm, all of which culminate in degradation by the 5'-3 'and 3'-5' decay pathways. They are termed non-stop decay (NSD), no-go decay (NGD) and nonsense-mediated decay (NMD) based on aberrations in the mRNA caused by the lack of a stop codon, a block to translation or a spurious stop codon, respectively. Exosomal decay of mRNA is also the final step of several gene regulation pathways like ARE-mediated decay (Haile et al., 2003), micro-RNA mediated decay (Valencia-Sanchez et al., 2006) and Staufen-mediated decay (Park & Maquat, 2013).

2.5 mRNA SURVEILLENCE PATHWAYS – PREAMBLES TO EXOSOME/XRN1-MEDIATED DECAY

2.5.1 NO-GO AND NON-STOP DECAY PATHWAYS

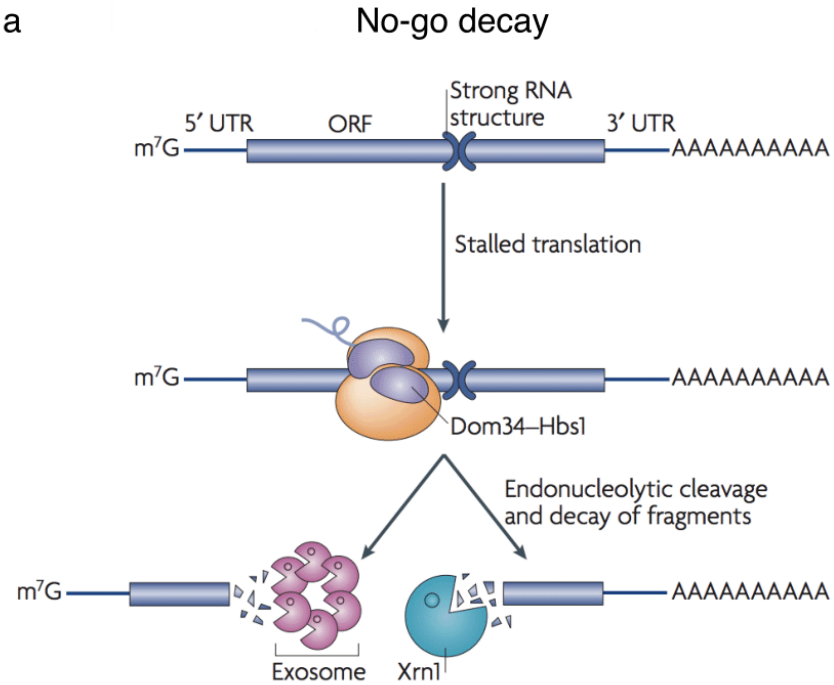
Both NGD and NSD pathways seem to be initiated in response to ribosome stalling. The NGD pathway targets mRNA substrates with features that would cause the ribosome to stall before reaching the end of the message (Doma & Parker, 2006) (Fig. 2.5a). NSD targets mRNAs lacking a stop codon where a ribosome might stall on a truncated mRNA codon or on a message without an in-frame stop codon causing the ribosome to translate the poly(A) tail and thus stall (Vasudevan et al., 2002) (Fig. 2.5b). Despite the substrates being seemingly different, NGD and NSD have several unifying features. Furthermore, an NGD substrate would become an NSD-like substrate after an endonucleolytic cleavage. The secondary stall formed in the upstream translating ribosome would result in an NSD-like scenario where the ribosome is stalled at the end of the message, reviewed in (Graille & Seraphin, 2012; Lykke-Andersen & Bennett, 2014; Shoemaker & Green, 2012; Simms et al., 2017). Finally, both NGD and NSD would require similar resolution mechanisms to recognize and rescue stalled ribosomes.

Two central factors involved in NGD and NSD related ribosomal rescue mechanisms are Dom34 (or Pelota in metazoans) and Hbs1 (Saito et al., 2013; Shoemaker et al., 2010). They are structurally related to the canonical termination factors eRF1 and eRF3 and therefore interact directly with the empty aminoacyl site (A-site) of the ribosome (Hilal et al., 2016). However, unlike the canonical release factors, Dom34-Hbs1 lack the hydrolytic activity required to release the nascent chain (Pisareva et al., 2011). Structural evidence has shown that Hbs1 possibly prevents Dom34 from binding actively translating ribosomes

thereby acting as a sensor of translational status (Hilal et al., 2016). while Dom34, once bound to the empty A-site promotes ribosomal splitting by Rli1/ABCE1 (Shoemaker & Green, 2011). Cryo-electron microscopic (Cryo-EM) analysis revealed that Dom34-Hbs1 bind to NGD and NSD stalled ribosomes in a similar manner suggesting a general mechanism of ribosomal rescue (Becker et al., 2011; Hilal et al., 2016). However, the exact role of Dom34 and Hbs1 in NSD remains to be elucidated. Another unifying feature of the two quality control pathways is the necessity for an endonucleolytic cleavage event to trigger the pathway. Recently, Cue2 endonuclease was shown as the competent endonuclease required for NGD via Xrn1-mediated decay and possibly NSD substrates (D'Orazio et al., 2019).

Even though the exosome possibly has a role in the clearance of both NGD and NSD substrates, the exosome-associated Ski complex has been shown to play a key role in recognizing NSD substrates (van Hoof et al., 2002). While all components of the Ski complex are required for promoting the degradation of both normal mRNAs and NSD substrates, the C-terminal domain of Ski7 seems to be specifically involved in recognizing NSD substrates (van Hoof et al., 2002). Given that Ski7 shares structural similarity to Hbs1, it is possible that the role of Dom34: Hbs1 in NSD is much more complex than previously envisioned. Furthermore, endonucleolytic activity of the exosome in yeast seems to play a role in NSD (Schaeffer & van Hoof, 2011) indicating a more central role for the cytoplasmic exosome in clearing NSD substrates. However, further research is required to clearly understand the role of endonucleolytic activity of the exosome in NSD.

a


No-go decay
Strong RNA structure
5′ UTR
ORF
3′ UTR
m7G
AAAAAAAAAA
Stalled translation
Dom34–Hbs1
Endonucleolytic cleavage and decay of fragments
Exosome
Xrn1

b

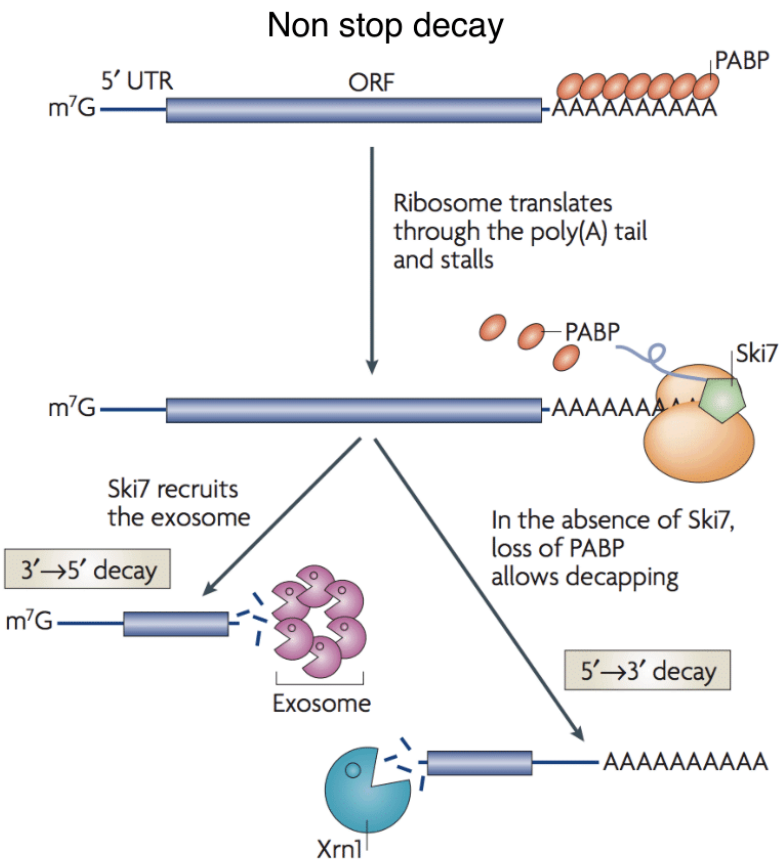

Non stop decay
5′ UTR
ORF
PABP
m7G
AAAAAAAAAA
Ribosome translates through the poly(A) tail and stalls
PABP
Ski7
Ski7 recruits the exosome
3′→5′ decay
In the absence of Ski7, loss of PABP allows decapping
5′→3′ decay
Exosome
Xrn1

Figure 2.5: Cartoon representation of the No-go and Non-stop decay mechanisms
a) A cartoon representation of the no-go decay pathway where a strong mRNA secondary structure results in stalling of the ribosome. The ribosome stall triggers the recruitment of the Dom34-Hbs1 complex resulting in the resolution of the stall and subsequent clearance of the aberrant mRNA by the exosome and XRN1. b) A cartoon representation of the non-stop RNA decay pathway highlighting a model where Ski7 mimics the tRNA and binds to the A-site triggering the recruitment of the cytoplasmic exosome. The figure has been adapted from (Garneau et al., 2007). The graphics are not to scale and do not represent the true architecture of the protein complexes.

2.5.2 NONSENSE-MEDIATED DECAY

Nonsense-mediated decay (NMD) is a conserved quality control pathway in eukaryotes, recognizing and degrading faulty mRNAs that contain premature termination codons. Although several NMD factors and aspects of NMD were first described based on work carried out in yeast and nematodes, only the current mechanistic view of EJC-dependent mammalian NMD is discussed here (Fig. 2.6).

A mammalian splicing-dependent NMD substrate is defined based on the location of the stop codon with respect to the exon junction complex (EJC). The EJC is deposited upon splicing in the nucleus, about 20-24 nucleotides upstream of splice junctions (Le Hir et al., 2000). The EJC forms a binding platform for the NMD trans-acting factors in the cytoplasm and appears to coordinate with the terminating ribosome. NMD is believed to be elicited if a ribosome terminates at a premature termination codon (PTC) that is at least 50-55 nucleotides upstream of a splice junction, i.e. at least ~20 nucleotides upstream of an EJC. It is possible that this minimal distance reflects the physical space taken up by protein-protein interactions between the terminating ribosome and the EJC, but the nature of these interactions is not well understood.

The initiation of NMD is a complex process and involves several transacting factors organized around the EJC. The core NMD factors are the up-frame shift proteins UPF1, UPF2 and UPF3 (so named from studies in yeast) (Karam & Wilkinson, 2012). The metazoan-specific NMD trans-acting factors that were originally identified from screens in *C. elegans* include the proteins SMG1, SMG5, SMG6, and SMG7 (Hodgkin et al., 1989; Pulak

& Anderson, 1993). More recently, the additional factors SMG8 and SMG9 (Yamashita et al., 2009) and DHX34 and NAG/NBAS (Longman et al., 2007) have been identified. Finally, eukaryotic release factors (eRF1 and eRF3) that are found on terminating ribosomes as well as the poly(A) binding protein (PABPC1) are also known to be involved in the NMD pathway.

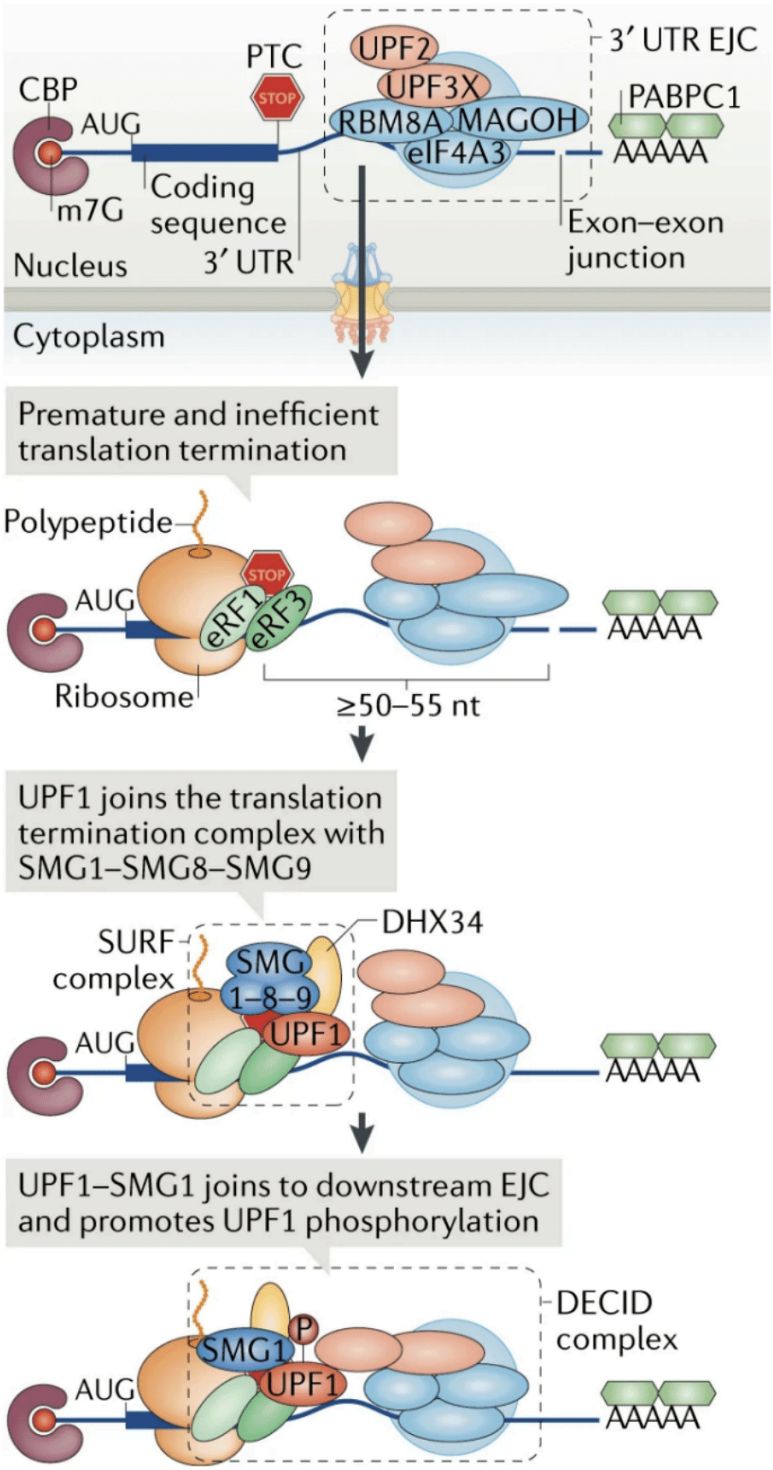


Figure 2.6: Initial steps of the EJC-dependent NMD pathway.

The first panel shows a model of a typical NMD substrate, which contains a premature termination codon upstream of the exon-junction complex. The second panel highlights the requirement of a minimal distance of 50-55 nucleotides between the stalled ribosome and the exon-junction complex, which triggers the recruitment of UPF1 and SMG1 via an unknown mechanism. The conversion of the surveillance complex (SURF) to the decay-inducing complex (DECID) is represented in panels 3 and 4. The figure has been adapted from (Kurosaki et al., 2019). The graphics are not to scale and do not represent the true architecture of the protein complexes.

Although several aspects of NMD are debated, a consensus has emerged from studies over the years. The key NMD factor UPF1 is an RNA helicase. The ATPase activity of UPF1 is essential for NMD and is thought to be required to remodel the messenger ribonucleoprotein particles (mRNP). mRNP remodeling would make the RNA accessible to the action of exoribonucleases (Franks et al., 2010). Degradation and recycling of the mRNA and associated proteins are the final steps of NMD and are the result of a coordinated series of events.

The function of UPF1 is regulated by protein-protein interactions and post-translational modifications. The ATPase activity of UPF1 is switched on by the binding of UPF2 (Chakrabarti et al., 2011; Clerici et al., 2009). UPF2 interacts with UPF3 (Kadlec et al., 2004), which in turn binds the EJC (Buchwald et al., 2010). In metazoans, the helicase domain of UPF1 is flanked by N- and C-terminal unstructured regions, which are the sites of phosphorylation by SMG1 (Okada-Katsuhata et al., 2012). UPF1 phosphorylation is an important signal to recruit SMG6 and the SMG5-SMG7 heterodimer. These are multidomain proteins with phospho-serine binding 14-3-3 like domains (Fukuhara et al., 2005; Jonas et al., 2013). SMG6 endonucleolytically cleaves the NMD target (which is a committed step towards its degradation) via its PIN domain (Eberle et al., 2009; Glavan et al., 2006). SMG5 and SMG7 bring the transcript to P bodies, which are the site of decapping and 5'-3' degradation, and recruit the phosphatase that dephosphorylates UPF1 (Chiu et al., 2003; Unterholzner & Izaurralde, 2004).

The phosphorylation and dephosphorylation cycle of UPF1 is essential for metazoan NMD (Chang et al., 2007). However, how and when it occurs is unclear. Unphosphorylated

UPF1 is found together with eRF1, eRF3 and SMG1 as part of the so-called SURF complex as identified by co-immunoprecipitation studies (Kashima et al., 2010). Although the validity of the model remains uncertain, it suggests that SURF-bound UPF1 interacts with the downstream UPF2-UPF3-bound EJC to form the so-called decay-inducing (DECID) complex. In this context, UPF1 is believed to be phosphorylated; inducing the decay of the mRNA it is bound to. The two possible regulators of SMG1 (SMG8-SMG9) are part of the SURF complex (Yamashita et al., 2009), and are likely involved in controlling timely activation of the kinase activity. Similarly, SMG1-UPF1 has been recently shown also to interact with UPF2 (Melero et al., 2014) and with the ATPase DHX34 (Hug & Caceres, 2014). These early steps in NMD centered on SMG1 and the SURF complex are poorly understood at the mechanistic and structural level.

2.5.2.1 SMG1 KINASE – THE GATE KEEPER OF NONSENSE-MEDIATED DECAY

The committed step of metazoan NMD is phosphorylation of UPF1 by SMG1 kinase, which triggers the recruitment of the endonuclease SMG6 (Chakrabarti et al., 2014; Eberle et al., 2009). As such, understanding how SMG1 is recruited to an NMD event and how its activity is regulated is critical for understanding NMD.

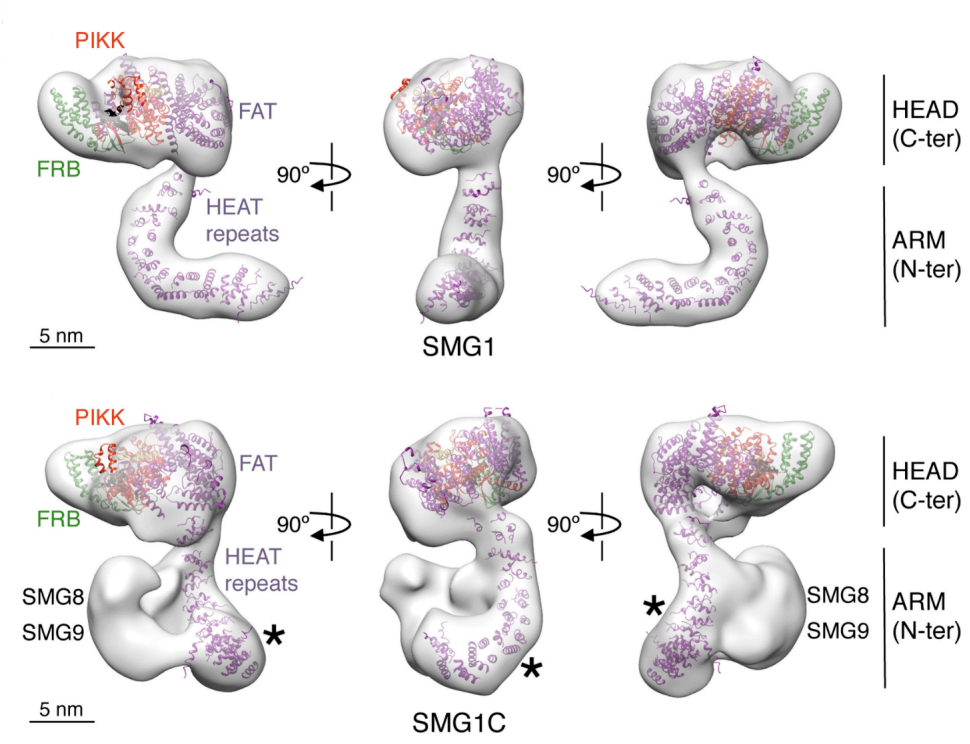


Figure 2.7 Low resolution architecture of SMG1 and SMG1-SMG8-SMG9 (SMG1C)
Architecture of SMG1 showing the C-terminal head domain comprised of the FAT, FRB and the kinase domain and the N-terminal HEAT repeat arm. The HEAT repeat density is fitted with the homology model and the head density is fitted with the structure of mTOR by the authors. The lower panel shows how the HEAT repeat region undergoes a conformational change upon SMG8-SMG9 binding. The figure has been adapted from (Melero et al., 2014).

SMG1 is a large multi domain kinase of about 410 kDa in humans. It belongs to the phosphatidylinositol (PI) 3-kinase-related kinase (PIKK) family and shares a similar architecture with other mammalian PIKKs (Grimson et al., 2004). Low-resolution EM analyses have revealed that SMG1 forms a two-lobed structure (Fig. 2.7). The larger lobe (the so-called 'head') appears to include the C-terminal part of the polypeptide amounting to about two thirds of the molecule. This C-terminal region is expected to contain a FAT domain, a PIKK-like kinase domain, a poorly understood insertion domain that possibly plays a role in substrate recruitment and a small C-terminal FATC domain (Arias-Palomo et al., 2011; Deniaud et al., 2015; Melero et al., 2014). Attached to the 'head' is an extended 'arm' that is likely to be formed by the N-terminal HEAT-repeat region amounting to about one third of the polypeptide (Deniaud et al., 2015; Melero et al., 2014).

Almost all the domains of SMG1 were directly, or indirectly shown to have a role in regulating its kinase activity (Morita et al., 2007). UPF1, the physiological substrate of SMG1 is recruited to the head region as shown by the low-resolution EM studies. Biochemical analysis revealed that the phosphorylation sites on UPF1 are located at both the N-terminus and C-terminus of the helicase. The 'head' region of SMG1 was also shown to interact with the helicase DHX34 (Melero et al., 2016). Although, the exact role of DHX34 in NMD is unclear, it has been hypothesized that DHX34 interacts preferentially with unphosphorylated UPF1 and promotes the interaction of SMG1 with other NMD factors (Melero et al., 2016). Another NMD factor that is believed to affect SMG1 catalytic activity via an interaction at the 'head' region is UPF2. Initially, it was believed that UPF2 modulates SMG1 activity in an allosteric manner (Melero et al., 2014). However, a later study found that UPF2 functions possibly in substrate release and sequestering phosphorylated UPF1 away from SMG1 (Deniaud et al., 2015). Biochemical analysis of SMG1 identified key mutations in the FATC domain that affected its activity suggesting a structural rearrangement of the kinase domain with respect to the FATC domain upon

substrate binding (Morita et al., 2007). Indeed, low-resolution EM studies comparing SMG1 and a UPF1-bound state of SMG1 indicate a rearrangement near the 'head' region of SMG1 (Melero et al., 2014). The N-terminal heat repeat region is known to recruit the G-domain containing proteins, SMG8 and SMG9. Structural studies have shown that the N-terminal region undergoes a significant conformational change upon binding SMG8. In addition, SMG8-SMG9 mutants were shown to effect the stability of PTC-containing transcripts (Yamashita et al., 2009), and deletion of the N-terminal region of SMG1 resulted in a loss of kinase activity indicating a regulatory role of SMG8-SMG9 via the N-terminal region of SMG1 (Morita et al., 2007). However, the exact mechanism of this regulation is poorly understood with *in vitro* data showing that SMG8 suppresses the kinase activity of SMG1 while *in vivo* data shows that SMG8-SMG9 are required for productive NMD.

Despite the wealth of biochemical and low-resolution structural data, lack of a high-resolution structure of SMG1 and its complexes have strongly hindered understanding the mechanistic basis of the initial steps of NMD and rationalization of the biochemical data.

2.6 AIM AND SCOPE OF THE THESIS

One of the most interesting questions in RNA decay is how substrates are accurately recognized. As described in the section 2.4.3, the Mtr4 helicase seems to play a crucial role in directing the activity of the nuclear exosome by participating in various complexes that localize in distinct compartments in the nucleus. The work performed in this thesis builds on the hypothesis that certain exosome adaptors are recruited to MTR4 via a conserved motif known as the 'arch interacting motif' (Thoms et al., 2015). The research questions were designed to explore whether the arch interacting motif identified in ribosome biogenesis factors could be a more general motif, thus establishing recruitment of exosome adaptors via the Mtr4 KOW domain as a general interaction mechanism. This thesis question is addressed by structural and biochemical characterization of the interactions of MTR4 with an early ribosomal biogenesis factor, NVL and with ZCCHC8, the scaffolding and MTR4 stimulating factor of the NEXT complex.

Additional work that contributed to the thesis explores the architecture of the SMG1-SMG8-SMG9 complex. Despite a wealth of low-resolution structural and biochemical work, many of the findings have yet to be rationalized at a structural level. Structural basis for the activity and function of the SMG1 kinase could potentially lead to new testable hypotheses regarding the early steps of NMD. Work performed as part of the thesis lead to the establishment of a robust expression system for the SMG1-SMG8-SMG9 complex that culminated in productive structural studies.

3.0 RESULTS

3.1 CHARACTERIZATION OF MTR4-EXOSOME ADAPTOR INTERACTIONS

Lingaraju.M., Johnsen.D., Schlundt.A., Langer.L.M., Basquin.J., Sattler.M., Jensen.T.H., Falk.S., Conti.E. (2019). The MTR4 helicase recruits nuclear adaptors of the human RNA exosome using distinct arch-interacting motifs. Nat Comm 10, 3393.

This study characterizes the interactions of MTR4 with ZCCHC8 and NVL. The results demonstrate that both ZCCHC8 and NVL harbor an arch interacting motif extending the repertoire of the motif to recognize and bind the MTR4 KOW domain.

The work was performed under the supervision of Dr. Falk S. and Prof. Conti E. Lingaraju M. participated in the design of the project, crystallography and biochemical analysis. The collaborators contributed to *in cellulo* experiments and NMR analysis. Detailed author contributions are included in the attached article.

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-019-11339-x OPEN

The MTR4 helicase recruits nuclear adaptors of the human RNA exosome using distinct arch-interacting motifs

Mahesh Lingaraju[1], Dennis Johnsen[2,7], Andreas Schlundt[3,4,5,7], Lukas M. Langer[1], Jérôme Basquin[1], Michael Sattler[3,4], Torben Heick Jensen[2], Sebastian Falk[1,6] & Elena Conti[1]

The nuclear exosome and its essential co-factor, the RNA helicase MTR4, play crucial roles in several RNA degradation pathways. Besides unwinding RNA substrates for exosome-mediated degradation, MTR4 associates with RNA-binding proteins that function as adaptors in different RNA processing and decay pathways. Here, we identify and characterize the interactions of human MTR4 with a ribosome processing adaptor, NVL, and with ZCCHC8, an adaptor involved in the decay of small nuclear RNAs. We show that the unstructured regions of NVL and ZCCHC8 contain short linear motifs that bind the MTR4 arch domain in a mutually exclusive manner. These short sequences diverged from the arch-interacting motif (AIM) of yeast rRNA processing factors. Our results suggest that nuclear exosome adaptors have evolved canonical and non-canonical AIM sequences to target human MTR4 and demonstrate the versatility and specificity with which the MTR4 arch domain can recruit a repertoire of different RNA-binding proteins.

[1] Department of Structural Cell Biology, Max-Planck-Institute of Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany. [2] Department of Molecular Biology and Genetics, Aarhus University, C.F. Møllers Alle 3, 8000 Aarhus C, Denmark. [3] Center for Integrated Protein Science Munich (CIPSM) at Department of Chemistry, Technical University of Munich (TUM), 85747 Garching, Germany. [4] Institute of Structural Biology, Helmholtz-Zentrum München, 85764 Neuherberg, Germany. [5] Present address: Institute for Molecular Biosciences and Center for Biomolecular Magnetic Resonance (BMRZ) at Johann Wolfgang Goethe-University, Frankfurt am Main 60438, Germany. [6] Present address: Max F. Perutz Laboratories, Department of Structural and Computational Biology, University of Vienna, Campus Vienna Biocenter 5, 1030 Vienna, Austria. [7] These authors contributed equally: Dennis Johnsen, Andreas Schlundt. Correspondence and requests for materials should be addressed to S.F. (email: sebastian.falk@univie.ac.at) or to E.C. (email: conti@biochem.mpg.de)

Eukaryotic cells generate a multitude of RNA species that require timely maturation and decay to maintain a healthy transcriptome. A central machinery in nuclear RNA processing, quality control and decay pathways is a conserved 3′–5′ exoribonuclease complex known as the RNA exosome (reviewed in[1,2]). Most mechanistic studies to date have analyzed the RNA exosome from *S. cerevisiae*, the species in which it was originally identified 20 years ago[3]. The yeast exosome consists of a 10-subunit core complex (Exo_{10})[4], the activity of which depends on a single processive exoribonuclease (Rrp44, also known as Dis3)[5,6]. The Exo_{10} core is present in both nuclear and cytoplasmic compartments, but its cofactors and regulators have distinct subcellular localizations (reviewed in[1,2]). In the nucleus, Exo_{10} is bound to the distributive ribonuclease Rrp6 and its associated protein Rrp47 as well as to the Mpp6 protein[7–9]. Together, Rrp6-Rrp47 and Mpp6 recruit the RNA helicase Mtr4 to the exosome[10–12]. Orthologues of all these 14 proteins exist in human cells, and engage in similar interactions to form the corresponding human nuclear exosome complex[12,13].

In both yeast and humans, the nuclear helicase Mtr4/MTR4 is central to exosome function[2,14,15]. First, it functions as an enzyme to remodel ribonucleoprotein (RNP) substrates with its 3′–5′ unwinding activity, and to present the unwound RNA substrate to the exosome core[12,16]. Furthermore, it functions as a binding platform for RNA-binding adaptors, providing the primary interactions to transcripts subjected to exosomal degradation in both RNA processing and decay pathways[17,18]. For example, *S. cerevisiae* Mtr4 binds Nop53, a ribosome biogenesis factor that recruits the exosome for a late step in rRNA processing, namely the trimming of ITS2 (Internal Transcribed Spacer 2). Two other factors, Trf4 and Air2, bind Mtr4 to form the so-called TRAMP complex[19–21], which allows the exosome to target aberrant tRNAs[22], rRNAs and small nuclear and nucleolar RNAs (sn/snoRNAs) for decay[23]. Higher eukaryotes not only have orthologues of Nop53 and TRAMP, but also have an increased number of nuclear exosome adaptors. In human cells, MTR4 has been reported to interact with the early ribosome biogenesis factors WDR74 and NVL, which take part in the processing of ITS1 (Internal Transcribed Spacer 1)[24,25]. Human MTR4 also binds to two large Zinc-finger proteins, ZCCHC8 and ZFC3H1. ZCCHC8 interacts with MTR4 and the RNA-binding protein RBM7 to form the trimeric Nuclear EXosome Targeting (NEXT) complex[17,26], which targets enhancer RNAs (eRNAs), promoter upstream transcripts (PROMPTs) and intronic RNAs for exosome-mediated decay[27]. ZFC3H1 instead directs MTR4 and the nuclear exosome to polyadenylated nuclear RNAs by connecting to the nuclear poly(A) binding protein PABN1[28–30]. Furthermore, MTR4 binds NRDE-2, a negative regulator that prevents the nuclear exosome from targeting RNAs that should be exported to the cytoplasm[31].

How does the nuclear exosome helicase mediate binding to such a diverse and functionally distinct set of proteins? Structural studies have shown that yeast Mtr4 contains an unstructured N-terminal region, a DExH helicase core and an 'arch' domain with a globular Kyprides, Ouzounis, Woese (KOW) domain[14,32]. All Mtr4 domains are involved in protein-protein interactions: the N-terminal region binds Rrp6-Rrp47[11], the helicase core binds Mpp6 as well as Trf4-Air2[33,34] and the KOW domain binds a short sequence known as 'arch-interacting motif' that is present in Nop53, Utp18, and Air2[35,36]. In the case of human MTR4, both the arch domain and the DExH helicase core bind NRDE-2[31]. Structural data have also elucidated how the helicase core of human MTR4 binds a region of ZCCHC8[37]. This protein is however expected to harbor another MTR4-binding region[37], which remains to be identified. Also unclear is how other human MTR4-binding proteins are recognized. With the exception of the expected AIM sequence in the human NOP53 orthologue (also known as GLTSCR2) and in NRDE-2[31], there is no consensus motif that can be identified with confidence at the sequence level in other MTR4-binding factors. Here, we used a combination of biochemical studies, X-ray crystallography and nuclear magnetic resonance (NMR) experiments to obtain mechanistic insights into how human MTR4 interacts with two metazoan RNA exosome adaptors: the RNA processing factor NVL and the RNA decay factor ZCCHC8.

Results

The unstructured region of human NVL interacts with MTR4. The nuclear VCP-like (NVL) protein, also known as NVL2, is a ribosome biogenesis factor of the AAA-ATPase family that has been reported to interact with human MTR4[38]. NVL is a multi-domain protein characterized by an N-terminal nucleolin-binding domain (residues 10–74)[39] and a linker region (residues 76–266) followed by two globular domains, characteristic of AAA-ATPases that are responsible for catalytic activity (Fig. 1a, Supplementary Fig. 1a). Bioinformatic analyses suggested that the linker region is mostly unstructured (residues 76–239) (Supplementary Fig. 1b). In other AAA-ATPases, unstructured regions upstream of the catalytic domains often mediate protein-protein interactions[40–42]. To test if the portion upstream of the NVL catalytic domains mediates the interaction with MTR4, we expressed and purified the N-terminal region of human NVL tagged with thioredoxin (Trx) as the prey protein ($Trx\text{-}NVL^{1–266}$) and the structured portion of MTR4 tagged with glutathione-S-transferase (GST) as the bait protein (GST-MTR4-ΔN). In these assays, $Trx\text{-}NVL^{1–266}$ specifically co-precipitated with GST-MTR4-ΔN (Fig. 1b, lane 11, compare with GST control in lane 9). Interestingly, pull-down assays with the recombinant yeast orthologues showed that the N-terminal unstructured region of Rix7 does not interact with yeast Mtr4 (Supplementary Fig. 1c), indicating that the interaction between human NVL and MTR4 is not conserved in *S. cerevisiae*.

We narrowed down the MTR4-interacting region of human NVL based on bioinformatic analysis. Sequence alignments showed that the N-terminal unstructured region of NVL contains an insertion that is present in the human protein and other chordates but not in the yeast orthologue (Supplementary Fig. 1a). Upon testing whether this insertion is responsible for the interaction with human MTR4, we found that a construct encompassing the conserved portion of the human NVL insertion spanning residues 167–216 (Supplementary Fig. 1a) was indeed able to co-precipitate with GST-MTR4-ΔN in pull-down assays (Fig. 1b, lane 14, $Trx\text{-}NVL^{167–216}$). Thus, the MTR4-binding determinants of NVL reside in an unstructured region that is present in the human orthologue and more generally in chordate NVL proteins, but not in lower eukaryotes.

NVL targets the MTR4 KOW domain. To identify where the MTR4-binding determinants for NVL reside, we performed GST pull down assays with MTR4 constructs harboring the DExH core of the helicase (GST-MTR4ΔNΔarch) and the KOW domain (GST-MTR4 KOW). Neither $Trx\text{-}NVL^{1–266}$ nor $Trx\text{-}NVL^{167–216}$ co-precipitated with GST-MTR4-ΔNΔarch (Fig. 1b, lanes 12 and 15). In contrast, the KOW domain characteristic of the MTR4 arch (GST-MTR4-KOW) was able to co-precipitate both $Trx\text{-}NVL^{1–266}$ and $Trx\text{-}NVL^{167–216}$ (Fig. 1b, lanes 13 and 16). We quantified the strength of the interaction by biophysical approaches. Using microscale thermophoresis (MST), we measured a dissociation constant of ~1 μM between a YFP-tagged version of $NVL^{167–216}$ and MTR4-KOW (Fig. 1c). A similar dissociation constant was obtained when testing the interaction of $NVL^{167–216}$

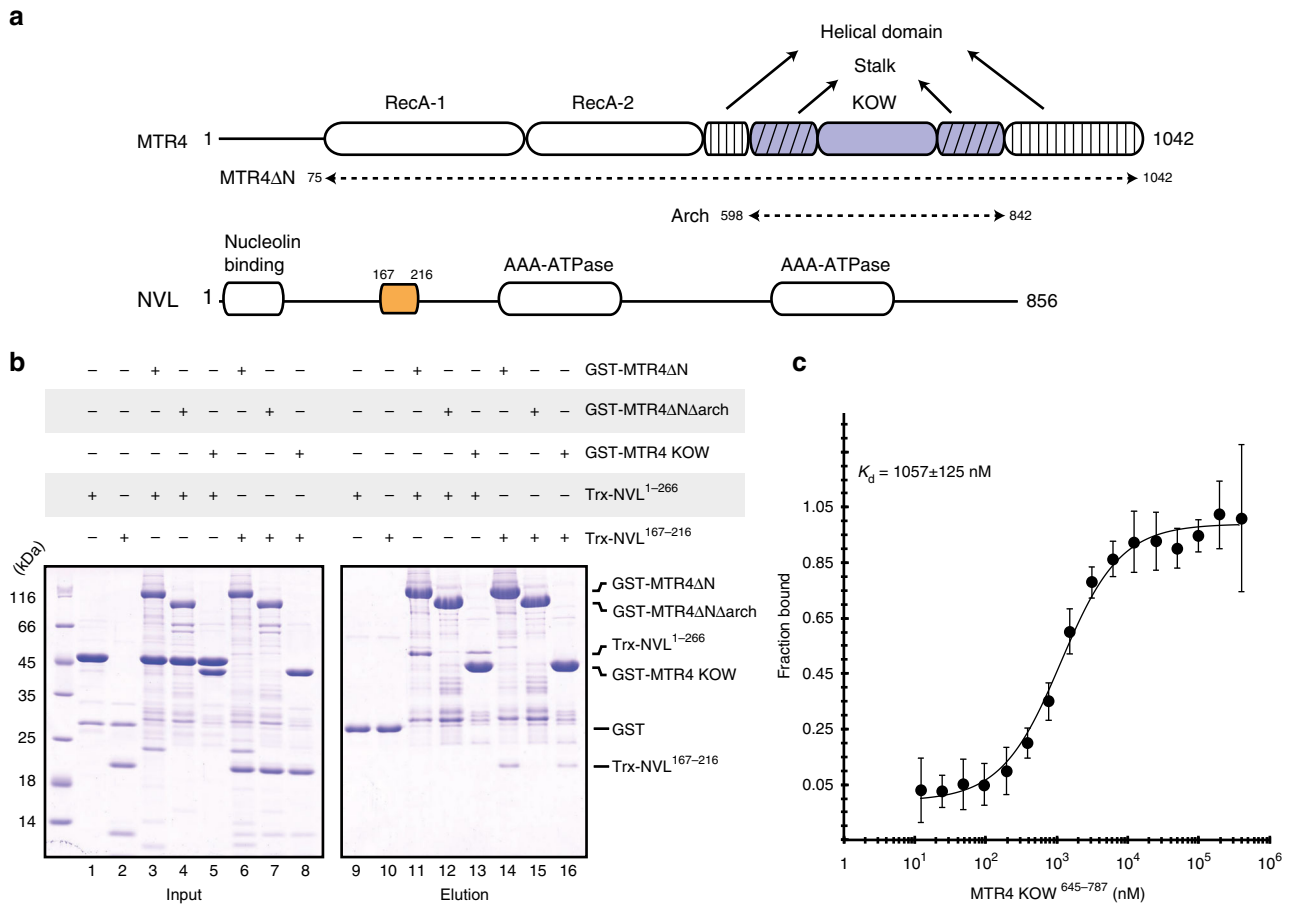


Fig. 1 The N-terminal unstructured region of NVL interacts with the MTR4 KOW. **a** Schematic representation of the domain organization of NVL and MTR4. Domain boundaries are obtained from previous studies[31,37] and the crystal structure in the current report. The arrows indicate nomenclature for MTR4 constructs used in the text. **b** Protein co-precipitation assays. GST-tagged MTR4ΔN, MTR4ΔNΔarch, and MTR4 KOW were incubated with either Trx-NVL$^{1-266}$ or Trx-NVL$^{167-216}$ before co-precipitation with glutathione sepharose beads. A total of 3% of the input (lanes 1–8) and 30% of the eluates (lanes 9–16) were analyzed on 15% SDS-PAGE gels stained with coomassie brilliant blue. **c** Microscale thermophoresis experiment with MTR4 KOW and Trx-NVL$^{167-216}$-$(GS)_3$-eYFP. In all, 50 nM of Trx-NVL$^{167-216}$-$(GS)_3$-eYFP was incubated with increasing concentrations of MTR4 KOW and thermophoresis was measured by tracking the fluorescence of the NVL-YFP fusion protein. The binding isotherm was calculated using MO software (Nanotemper technologies) and the dissociation constant (K_d) is given in the inset. The error bars represent the standard deviations of each data point calculated from three independent thermophoresis measurements

and MTR4-ΔN using isothermal titration calorimetry (ITC) (Supplementary Fig. 1d). The fact that the KOW domain alone binds NVL$^{167-216}$ as strongly as MTR4-ΔN indicated that the NVL$^{167-216}$ binding region resides within the KOW domain.

We proceeded to characterize the NVL$^{167-216}$-binding site of MTR4 with structural approaches. Using NMR spectroscopy, we confirmed that the MTR4-KOW domain has a sequential arrangement of secondary structure elements consistent with the fold observed in a recent crystal structure of an MTR4-NRDE-2 complex[31], namely a five-stranded β-sheet flanked by a long C-terminal helix and containing smaller helical segments within loops (Supplementary Fig. 2). We then carried out titration experiments using a ^{15}N-labelled MTR4-KOW protein and adding increasing amounts of unlabeled NVL$^{167-216}$. In line with a dissociation constant of ~1 µM, we observed an intermediate exchange regime for most of the peaks during the titration in HSQC (heteronuclear single quantum coherence) spectra. Chemical shift perturbations (CSP) measured upon NVL2$^{167-216}$ addition revealed significant effects (Fig. 2a–c) with clusters of strongly shifting peaks around residues 658, 695, 743, and 764. From the NMR analysis of the NVL$^{167-216}$ - MTR4-KOW interaction, the CSPs were very similar to those we had previously reported for the yeast Nop53-Mtr4 KOW interaction[36] (Supplementary Fig. 3a, 3b). Consistently, a reverse-charge substitution of Arg743 (R743E mutant) impaired the binding of both NVL$^{167-216}$ (pull-down assays in Fig. 2d, lane 8) and of the human Nop53 orthologue (Supplementary Fig. 3c).

Identification of W-AIM: a tryptophan arch-interacting motif of NVL. Given that NVL$^{167-216}$ binds the KOW domain of MTR4 similarly to Nop53, we expected the presence of a similar AIM motif. A stretch of amino acids within this segment (NVL residues 185–190) appeared to show significant sequence similarity to the Nop53 AIM sequence (Fig. 3a). Surprisingly in this context, the F186A and D189R mutations in the Nop53-like stretch did not affect NVL$^{167-216}$ binding to MTR4 in a pull-down assay with recombinant proteins (Fig. 3b, compare lanes 5 and 6), indicating that this segment is not a bona-fide AIM. To identify the arch-interacting motif in NVL$^{167-216}$, we analyzed the sequences from different species in terms of evolutionary conservation. We noticed that vertebrates feature a conserved hydrophobic segment GWFIDKTP (residues 172–179, Fig. 3a; Supplementary Fig. 1a). Mutations in this segment (W173A,

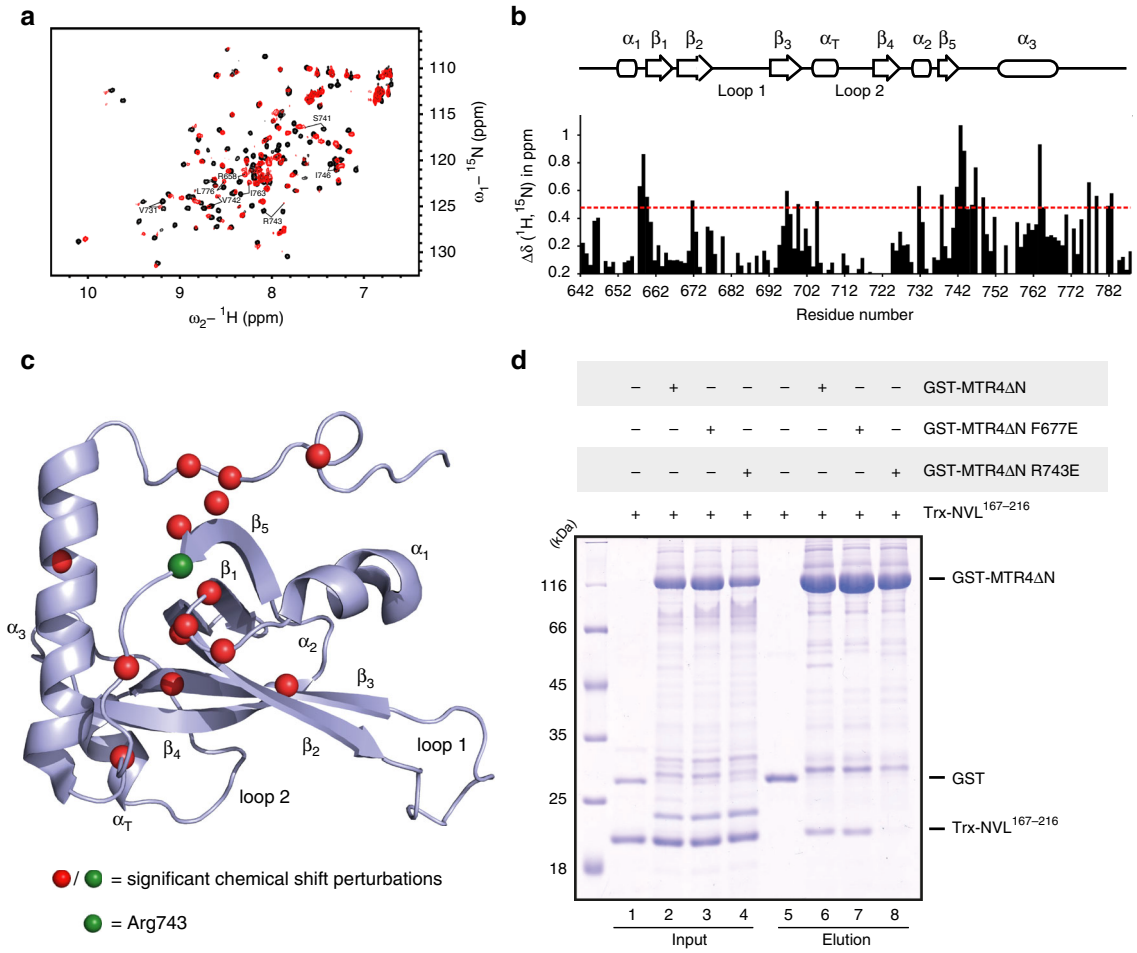


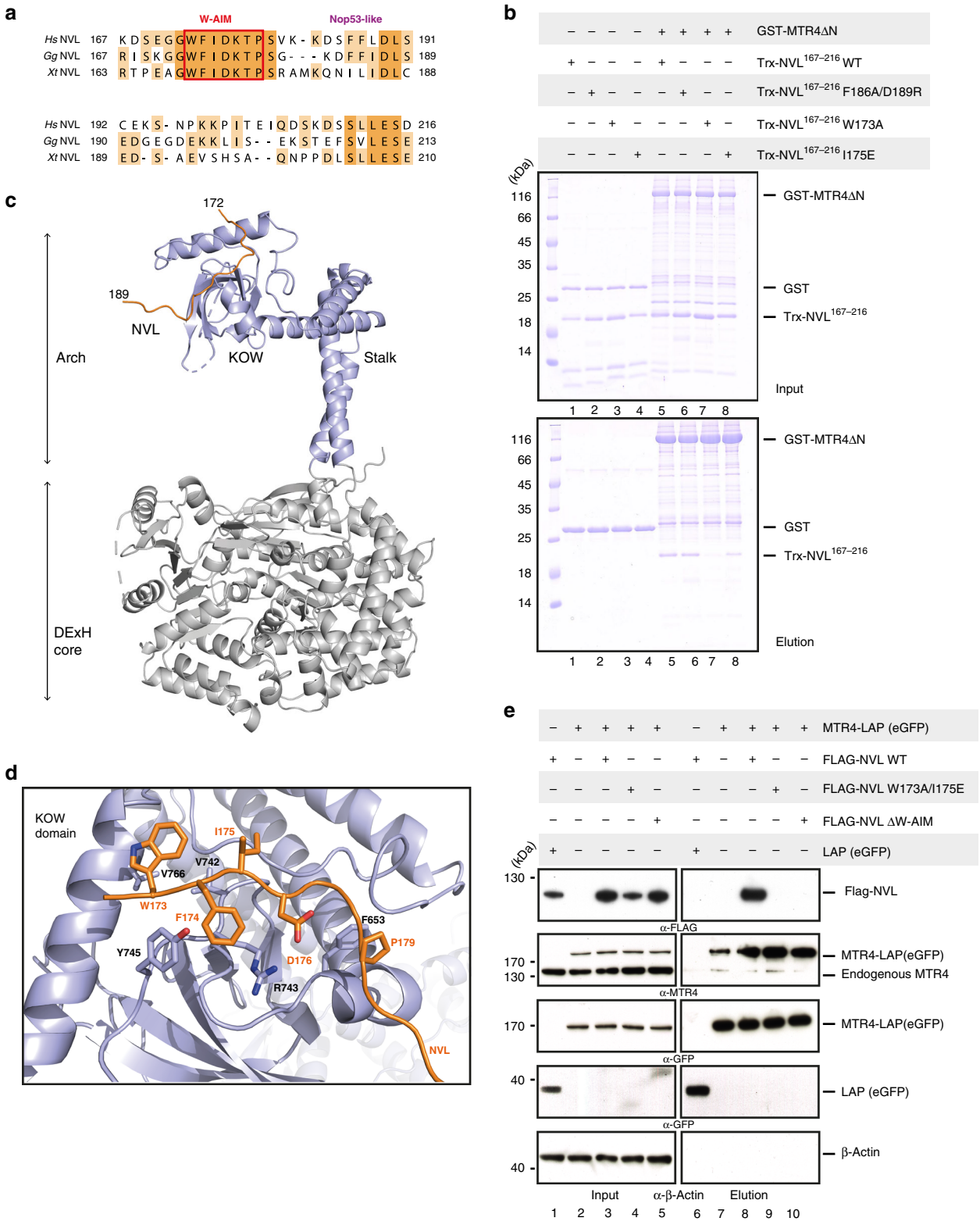
Fig. 2 Analysis of MTR4 KOW-NVL complex by NMR and site directed mutagenesis. **a** Overlay of ^{1}H-^{15}N-HSQCs of either MTR4 KOW alone (black) or in complex with a six-fold excess of NVL2 (red). Selected residues experiencing large chemical shift perturbations (CSPs) are labeled. **b** Plot of CSPs per residue of the MTR4 KOW sequence. The red line marks the threshold of significant CSPs, which are mapped on the model of MTR4 KOW (PDB 6IEH) in panel c. Gaps indicate either prolines or residues that could not be assigned. **c** MTR4 KOW domain (PDB 6IEH) with labeled secondary structural elements as assigned by NMR and residues showing significant CSPs (displayed as red/green spheres). **d** Protein co-precipitations by pull down assays. GST-tagged MTR4ΔN (WT or mutants) were incubated with Trx-NVL$^{167-216}$ before co-precipitation with glutathione sepharose beads. A total of 3% of the input (lanes 1-4) and 30% of the eluates (lanes 5-8) were analyzed on 15% SDS-PAGE gels and visualized by staining with coomassie brilliant blue

I175E) either abolished or impaired NVL$^{167-216}$ binding to MTR4ΔN in pull-down assays in vitro (Fig. 3b, lanes 7 and 8), suggesting that this stretch functions as an arch-interacting motif, which we coin 'tryptophan-centered arch - interacting motif' (W-AIM).

We co-crystallized the NVL W-AIM peptide in complex with a construct of human MTR4 encompassing both the DExH core and the arch domain (residues 70–1042) (Table 1). The overall structure of the helicase is similar to that of the human NRDE-2-MTR4 complex[31], but with a notably different conformation of the arch domain. In the NVL-bound crystal structure, the arch domain of MTR4 is an open conformation, with the KOW domain clearly separated from the helicase core (Fig. 3c and Supplementary Fig. 4a). The NVL W-AIM polypeptide chain binds in an extended conformation, similar to that reported for NRDE-2[31], Nop53[36] and Air2[34] (Supplementary Fig. 4b, 4c). The hydrophobic side chains of Trp173$_{NVL}$, Phe174$_{NVL}$, and Ile175$_{NVL}$ contact Val766$_{MTR4}$, Tyr745$_{MTR4}$, and Val742$_{MTR4}$ respectively, while Asp176$_{NVL}$ interacts with Arg743$_{MTR4}$ (Fig. 3d). The NVL polypeptide chain then forms a pronounced bend at Pro179$_{NVL}$ (which contacts Phe653$_{MTR4}$) and continues to form a crystal contact, docking at the Trf4-binding site of a symmetry related helicase core (Supplementary Fig. 5a). Consistent with the structural data, mutation of Asp176$_{NVL}$ (D176A mutant) impairs the binding of NVL$^{167-216}$ to MTR4ΔN in a GST pull down experiment (Supplementary Fig. 5b).

Finally, we tested the effect of the MTR4-interacting residues of NVL in human cells. To this end, we carried out co-immunoprecipitation (co-IP) assays in HeLA cells stably expressing MTR4 with an eGFP 'localization and affinity purification' (LAP) tag at the N-terminus and transiently expressing FLAG-tagged full-length NVL constructs (wild type, or with the W173A/I175A mutation or with deletion of the entire hydrophobic segment). Western blotting analysis confirmed that wild type NVL bound MTR4 and that disruption of the 172–180 segment, either by mutation or deletion, impaired the NVL-MTR4 interaction (Fig. 3e, compare lanes 8, 9, 10). Taken together, we conclude that NVL interacts with the KOW domain of MTR4 using a short linear W-AIM sequence. The NVL W-AIM is more hydrophobic than the Nop53 AIM[35], rationalizing why it binds MTR4 with an order of magnitude higher affinity than the yeast Nop53-Mtr4[36] and human Nop53-MTR4 interactions (Supplementary Fig. 3d).

The unstructured region of ZCCHC8 interacts with the MTR4 KOW domain. Identification of the W-AIM sequence in NVL



motivated us to examine whether other known MTR4 interactors also contain a similar tryptophan-centered motif. One such interactor is ZCCHC8, the scaffolding subunit of the NEXT complex[27,37,43]. This modular protein contains a predicted N-terminal coiled–coil domain (residues 40–80), a Zinc-knuckle domain (residues 222–246), a proline-rich domain (residues 287–334) that interacts with RBM7[43] and a C-terminal domain (CTD) (residues 659–707) (Fig. 4a) that interacts with the DExH core of MTR4 and activates it[37]. The N-terminal portion of ZCCHC8 is expected to contain an additional MTR4-binding site[37]. Within the N-terminal portion, we focused our attention on the linker between the coiled-coil and the Zinc-knuckle domains, as it appeared to contain an NVL-like tryptophan-containing sequence (Supplementary Fig. 6a). We expressed a large portion of this linker (residues 91–211) tagged to maltose binding protein (MBP-ZCCHC8$^{91–211}$) and purified the resulting

Fig. 3 The vertebrate specific W-AIM in NVL is crucial for binding to MTR4 KOW. **a** Sequence alignment of vertebrate specific insertion regions of representative NVL sequences, *Homo sapiens* (*Hs*), *Gallus gallus* (*Gg*), *Xenopus tropicalis* (*Xt*), highlighting the W-AIM (GWFIDKTP) (red box), and the Nop53-like region (LFXϕD) (purple). The sequences were obtained from the UniProt database and aligned using the T-coffee server[59]. **b** Protein co-precipitations by pull down assays. GST-tagged MTR4ΔN was incubated with either wild type Trx-NVL$^{167-216}$ or its indicated mutant before co-precipitation with glutathione sepharose beads. A total of 3% of the input (top) and 30% of the eluates (bottom) were analyzed on 15% SDS-PAGE gels and visualized by staining with coomassie brilliant blue. **c** Overall structure of MTR4ΔN$_2^{70-1042}$ - NVL$^{167-216}$ complex, with the DExH core of MTR4 colored in gray and arch colored in light blue. NVL is colored in orange. **d** Zoom-in view of the interactions between MTR4 KOW (light blue) domain and NVL (orange). Domains are colored as in Fig. 3a and viewed are 90° rotation along the horizontal and vertical axes with respect to the view in Fig. 3a. Residues discussed in the text are highlighted and labeled. **e** Cellular co-IP assay. FLAG-tagged wild type NVL construct, or its indicated mutant variant, were transiently expressed in cells stably expressing MTR4-LAP. After precipitation of MTR4-LAP, a total of 0.5% of the input (left) and 8.0% of the eluates (right) were analyzed on 4–12% SDS-PAGE gel followed by western blotting analysis. The primary antibody used is indicated below the panel

Table 1 Data collection and refinement statistics

	Mtr4ΔN$^{70-1042}$ - NVL$^{167-216}$
Data collection	
Space group	$P6_1$
Cell dimensions	
a, *b*, *c* (Å)	184.37, 184.37, 90.53
α, β, γ (°)	90.0, 90.0, 120.0
Resolution (Å)	92.18–3.07
R_{sym} or R_{merge}	15.7 (527.6)[a]
I / σI	12.6 (0.9)
Completeness (%)	99.9 (99.5)
Redundancy	16.5
$CC_{1/2}$	99.9 (62.7)
Refinement	
Resolution (Å)	60.35–3.07 (3.22–3.07)
No. reflections	32,919
R_{work}/R_{free}	22.3 / 25.7
No. atoms	
Protein	7074
Ligand/ion	58
Water	6
B-factors	
Protein	123.57
Ligand/ion	138.03
Water	92.34
R.m.s. deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.478

[a]values in parentheses are for highest-resolution shell

construct for pull-down assays with the versions of GST-tagged MTR4 described above (Fig. 1a). MBP-ZCCHC8$^{91-211}$ co-precipitated with GST-MTR4-ΔN but not with GST-MTR4-ΔNΔarch (Fig. 4b, lanes 6 and 7). Similar to the results we had obtained for NVL$^{167-216}$, the pull-down assays showed that MBP-ZCCHC8$^{91-211}$ interacts with GST-MTR4-KOW (Fig. 4b, lane 8). Next, we determined the affinity of the MTR4-KOW-ZCCHC8$^{91-211}$ interaction using microscale thermophoresis (MST) with a fluorescent-tagged version of ZCCHC8$^{91-211}$ (that we had engineered by fusing a YFP to the C-terminus). In this quantitative assay, we measured a dissociation constant (K_D) of ~0.3 μM (Fig. 4c), indicating a higher affinity than that of the NVL$^{167-216}$ fragment. To corroborate these results, we pre-formed a GST-MTR4-ΔN - Trx-NVL$^{167-216}$ complex and incubated it with increasing amounts of MBP-ZCCHC8$^{91-211}$ before subjecting the mixtures to GST pull-down assays. The competition assay showed that ZCCHC8$^{91-211}$ could displace NVL$^{167-216}$ from the pre-formed complex (Fig. 4d, lanes 5–8), suggesting that they interact with the exosome helicase in a mutually exclusive manner.

Identification of C-AIM: a cysteine arch-interacting motif of ZCCHC8. We proceeded to identify the MTR4-binding site within the N-terminal region of ZCCHC8. We first interrogated the NVL-like tryptophan-containing patch (GWEIPK, residues 197–202, Fig. 5a). Surprisingly, mutation of Trp198 and Lys202 (W198A/K202E mutant) did not alter the interaction of ZCCHC8$^{91-211}$ with MTR4-ΔN in GST pull down assays (Fig. 5b, lane 7), indicating that this patch of ZCCHC8 is not a bona-fide W-AIM sequence.

To identify the arch-interacting motif in ZCCHC8$^{91-211}$, we took a similar bioinformatics approach as described above for NVL. When analyzing the evolutionary conservation of ZCCHC8$^{91-211}$, we identified a conserved patch centered at the FCLDKLG segment (residues 178–184, Fig. 5a). Mutation of Phe178 and Asp181 (F178A/D181R mutant) in this segment impaired the interaction with MTR4-ΔN (Fig. 5b, lane 6). This cysteine-centered arch-interacting motif (C-AIM) is loosely related to the Nop53 AIM and the NVL W-AIM segments, and is thus expected to bind to the same site of the KOW domain. Consistently, the Arg743 reverse-charge substitution (R743E) in MTR4 also impaired the binding of ZCCHC8$^{91-211}$ (Fig. 5c, lane 8).

ZCCHC8 also contains a non-canonical arch-interacting motif: I-AIM. While attempting to narrow down the KOW binding region of ZCCHC8 further, we observed that N-terminal truncation of ZCCHC8$^{91-211}$ resulted in a near loss of MTR4 binding (Supplementary Fig. 6b). Based on these results, we reasoned that ZCCHC8$^{91-211}$ might harbor an additional MTR4-binding motif that would be predicted to bind to an adjacent surface on the KOW domain (and thus show no resemblance to the previously identified arch-interacting motifs). Using bioinformatics approaches, we identified another conserved segment upstream of C-AIM (IEEF, residues 112–115). Indeed, mutation of Ile112 and Phe115 (I112R/F115R mutant) severely weakened the MTR4-ΔN - ZCCHC8$^{91-211}$ interaction in GST pull-down assays (Fig. 5b, lane 8). Thus, this segment (which we refer to as an isoleucine-centered arch-interacting motif, or I-AIM) also contributes to MTR4 binding. Next, we mapped the possible MTR4-binding site of I-AIM. The β-barrel of the MTR4 KOW domain is structurally related to that of Tudor domains (Supplementary Fig. 7a), small globular folds that generally present a hydrophobic pocket for binding methylated arginines and lysines[44]. Although there is little overall sequence similarity, MTR4-KOW features hydrophobic residues at the equivalent position of the substrate–binding residues of Tudor domains (Supplementary Fig. 7b)[44,45]. In particular, MTR4 Phe677 is evolutionarily conserved (Supplementary Fig. 7b) and is located on a surface adjacent to the Arg743 site where the Nop53 like AIM, W-AIM and C-AIM sequences are recognized (Supplementary Fig. 7c). In line with this site being used for the additional I-AIM sequence of ZCCHC8$^{91-211}$, mutation of Phe677 (F677E) disrupted binding of ZCCHC8$^{91-211}$ (Fig. 5c, lane 7) and did not affect the binding

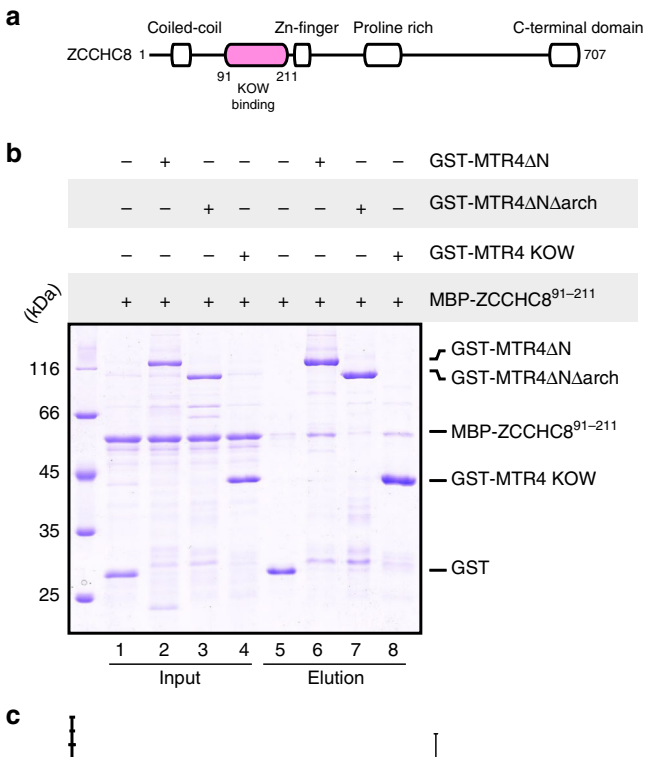


Fig. 4 The N-terminus of ZCCHC8 interacts with the MTR4 KOW domain. **a** Schematic representation of the domain organization of ZCCHC8. Domain boundaries are obtained from previous studies[37,43] and computational sequence analysis. **b** Protein co-precipitations by pull down assays. GST-tagged MTR4ΔN, MTR4ΔNΔarch, and MTR4 KOW were incubated with MBP-ZCCHC8 $^{91–211}$ before co-precipitation with glutathione sepharose beads. A total of 3% of the input (lanes 1–4) and 30% of the eluates (lanes 5–8) were analyzed on 12% SDS-PAGE gels and visualized by staining with coomassie brilliant blue. **c** Microscale thermophoresis experiment with MTR4 KOW and MBP-ZCCHC8 $^{91–211}$-$(GS)_3$-eYFP. In all, 1 μM MBP-ZCCHC8 $^{91–211}$-$(GS)_3$-eYFP was incubated with increasing concentrations of MTR4 KOW and thermophoresis was measured by tracking the fluorescence of the ZCCHC8-YFP fusion protein. The binding curve was calculated using MO software (Nanotemper technologies) and the dissociation constant (K_d) is given in the inset. The error bars represent the standard deviations of each data point calculated from three independent thermophoresis measurements. **d** Competition experiment using GST pull down assay technology. A preformed GST-MTR4ΔN:Trx-NVL$^{167–216}$ complex was incubated with increasing concentrations of MBP-ZCCHC8$^{(91–211)}$ before precipitation with glutathione sepharose beads. A total of 3% of the input (lanes, 1–4) and 30% of the eluates (5–8) were analyzed on 15% SDS-PAGE gels and visualized by staining with coomassie brilliant blue

of NVL$^{167–216}$ (Fig. 2d, lane 7) or human Nop53 (Supplementary Fig. 3c, lane 7).

We tested the importance of the I-AIM segment in human cells. HeLa LAP-MTR4 cells were transfected with FLAG-tagged full-length ZCCHC8 constructs (wild type, the I-AIM I112R/F115R double mutant or a deletion construct lacking the ZCCHC8 CTD). Western blotting analysis of the resulting co-IPs revealed that disrupting the I-motif by mutation is sufficient to impair the ZCCHC8-MTR4 interaction, whereas deletion of the CTD did not significantly affect complex formation (Fig. 5d, lanes 8, 9, 10). These mutations also did not affect the ZCCHC8-RBM7 interaction (Supplementary Fig. 6c). The CTD of ZCCHC8 has been shown to bind the DExH core of MTR4 and to enhance helicase activity[37]. In contrast, we found that neither the ZCCHC8 nor the NVL AIM motifs have a significant effect on the catalytic properties of MTR4, as judged by assaying both ATP hydrolysis and RNA helicase activities (Supplementary Fig. 8). These results suggest a division of labor of the N-terminal and C-terminal ZCCHC8 regions: while the C-terminal domain regulates the activity of the helicase[37], the N-terminal region plays has a central scaffolding role in incorporating MTR4 into the NEXT complex.

Discussion

In this study, we show that the human nuclear exosome adaptors NVL and ZCCHC8 bind the MTR4 KOW domain on a surface that is also employed by Nop53[36] and NRDE-2[31] using distinct arch-interacting motifs[35]. The AIMs of Nop53/Utp18/Air2 ($LFx\phi D(x)_{1–2}G/P$), NVL (GWFIDKTP), ZCCHC8 (NFCLDKLG), and NRDE-2 (SFRTDKKP) can best be considered as subfamilies of canonical AIMs. With insight, sequences in both NVL and ZCCHC8 that at first glance appeared to resemble known AIMs instead contain individual amino acids that are likely to prevent MTR4 binding. The consensus sequence of canonical bona-fide AIMs can thus be re-defined as $x\omega xxD(x)_{1/2}G/P$, with a C-terminal glycine or proline residue that allows the polypeptide chain to bend as it extends away from its binding site. The aromatic (ω) and polar/non-polar amino acids might vary, but tend to be conserved within individual subfamilies of canonical AIMs. Despite the variability, all canonical AIM sequences are recognized, in a mutually exclusive manner, at a surface pocket of the MTR4 KOW domain that is defined by the presence of Arg743 (yeast Mtr4 Arg774). Mutation of this surface pocket may thus be a useful tool to probe new MTR4-interacting

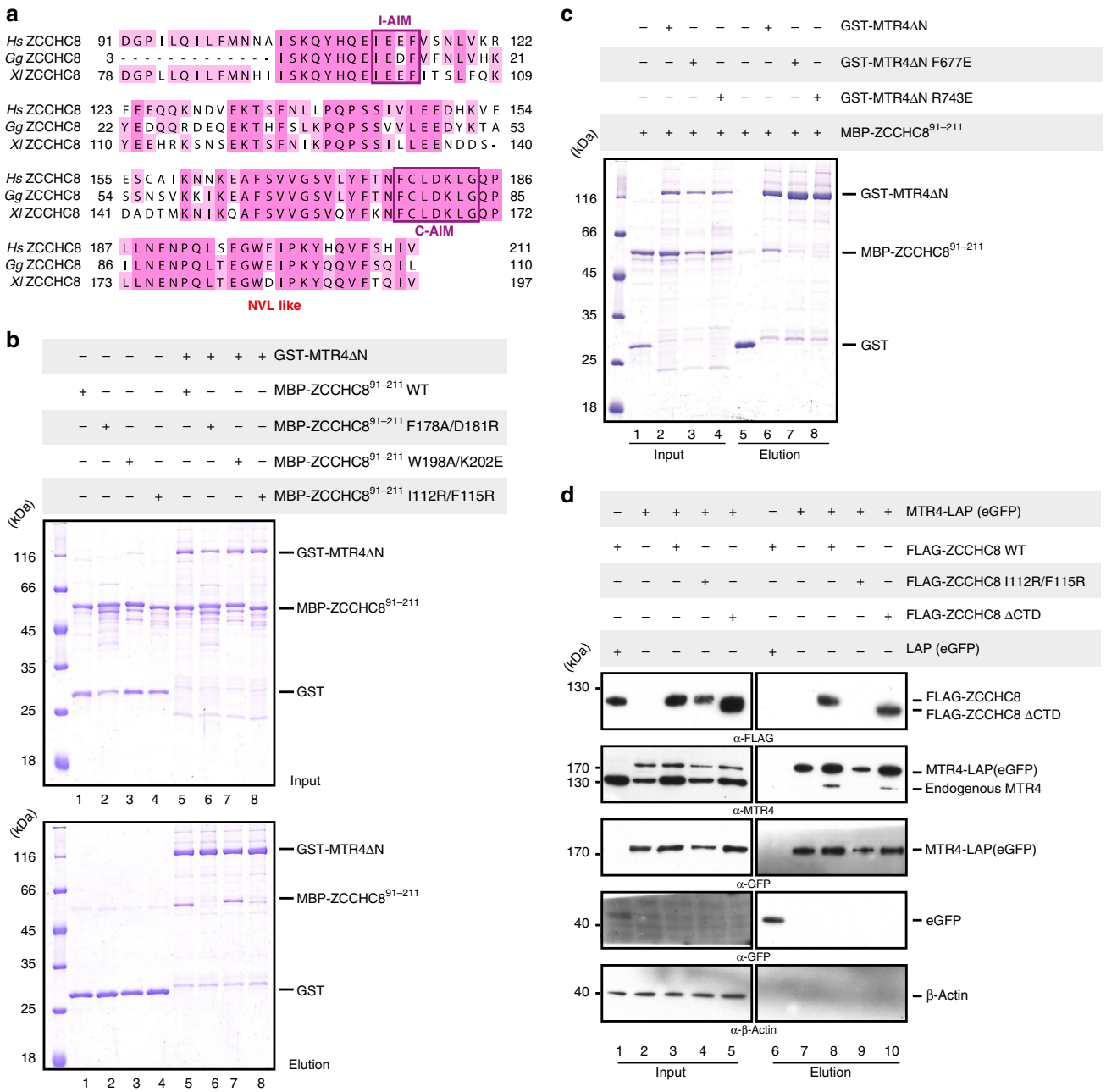


Fig. 5 Analysis of the ZCCHC8-MTR4 KOW complex by site directed mutagenesis. **a** Sequence alignment of the region between the predicted coiled coil domain and the zinc finger of ZCCHC8 from representative metazoan species, *Homo sapiens* (*Hs*), *Gallus gallus* (*Gg*), *Xenopus laevis* (*Xl*), highlighting NVL-like region (red), Nop53-like AIM (C-AIM) and ZCCHC8 specific I-AIM (purple boxes). The sequences were obtained from the UniProt database and aligned using the T-coffee server[59]. **b** Protein co-precipitations by pull down assays testing ZCCHC8 $^{91–211}$ mutants for MTR4ΔN binding ability. GST-tagged MTR4ΔN was incubated with either ZCCHC8$^{91–211}$ WT or mutants before co-precipitation with glutathione sepharose beads. A total of 3% of the input (top) and 30% of the eluates (bottom) were analyzed on 12% SDS-PAGE gels and visualized by staining with coomassie brilliant blue. **c** Protein co-precipitations by pull down assays testing MTR4ΔN mutants for their ZCCHC8 $^{91–211}$ binding ability. GST-tagged MTR4ΔN (WT or mutant variants) were incubated with MBP-ZCCHC8 $^{91–211}$ before co-precipitation with glutathione sepharose beads. A total of 3% of the input (lanes, 1–4) and 30% of the eluates (5–8) were analyzed on 12% SDS-PAGE gels and visualized by staining with coomassie brilliant blue. **d** Cellular co-IP assay. FLAG-tagged ZCCHC8 constructs (WT/IF mutant/CTD deletion) were transiently expressed in cells stably expressing MTR4-LAP. After precipitation of MTR4 taking advantage of the LAP tag, a total of 0.5% of the input (left) and 8.0% of the eluates (right) were analyzed on 4–12% SDS-PAGE gel followed by western blotting analysis. The primary antibody used is indicated below the panel

proteins that contain arch-interacting motifs, as these motifs are difficult to identify due to the degeneracy of their consensus sequence. Furthermore, we found that ZCCHC8 harbors an additional arch-interacting motif that does not conform to the canonical AIM consensus sequences. These findings show how different exosome adaptor proteins have evolved similar mechanisms to recognize MTR4 in a specific and mutually exclusive manner, but can also modulate the affinity and thus selectivity with which they are ultimately recruited to the nuclear exosome.

Methods

Protein expression and purification. Human MTR4 constructs (full-length, MTR4-ΔN (75–1042), MTR4-KOW (645–787) and MTR4-ΔNΔarch, where residues 598–842 were substituted by 2xGS linker) were expressed as 6xHis-GST-tagged fusion (cleavable with 3 C protease) proteins in BL21 star (DE3) *E. coli* cells, grown either in terrific broth (for biochemical studies) or minimal medium supplemented with ^{15}N labelled ammonium chloride and/or ^{13}C labeled Glucose (for NMR studies). The proteins were purified using a Ni-nitrilotriacetate (NTA) affinity column and a heparin column (GE healthcare) for ion exchange chromatography. When appropriate, the His-GST tag was cleaved at this point by incubation with 3 C protease, followed by removal of the tag with an additional Ni-NTA affinity step. Finally, the protein was subjected to size exclusion chromatography on a Superdex 200 column (GE healthcare) in 50 mM Hepes/NaOH pH 7.5, 150 mM NaCl, 5% (v/v) glycerol, 2 mM DTT.

The human MTR4 construct used for crystallization (71–1042) was cloned as GST fusion construct with a TEV cleavage site immediately preceding the MTR4 sequence which when cleaved would leave a glycine (the natural 70th residue in MTR4) yielding MTR4 protein with residues 70–1042 (referred to as $MTR4\Delta N_2$ in the text). The construct was purified as described above.

The primers used for cloning all the constructs described are listed in Supplementary table 1.

All NVL and ZCCHC8 constructs and truncation mutants described in the text were expressed as 6xHis-Trx and 6xHis-MBP tagged fusion (cleavable with 3 C protease) proteins, respectively in BL21 star (DE3) E. *coli* cells. The proteins were purified using a Ni-NTA affinity column. When appropriate, the tags were cleaved upon incubation with 3 C protease at this step, followed by an additional Ni-NTA affinity step for the removal of the tag. In case of ZCCHC8, the tag cleavage was performed only in the presence of MTR4 to prevent the protein from precipitating. In the final purification step, the proteins were subjected to size-exclusion chromatography on a Superdex 75 in 50 mM Hepes/NaOH pH 7.5, 150 mM NaCl, 2 mM DTT.

Biophysical assays. The microscale thermophoresis measurements were performed on a NanoTemper Monolith NT.115 machine. Before the measurements, all samples were dialyzed against a buffer containing in 50 mM Hepes/NaOH pH 7.5, 150 mM NaCl, 5% (v/v) glycerol, 0.5 mM TCEP. For NVL, 50 nM of Trx-$NVL^{167–217}$-$(GS)_3$-eYFP was incubated with increasing concentrations of unlabeled MTR4 KOW and thermophoresis was measured with an MST power of 20% and an LED power of 20%. For ZCCHC8, 1 μM of MBP-$ZCCHC8^{91–211}$ -$(GS)_3$-eYFP was incubated with increasing concentrations of unlabeled MTR4-KOW and thermophoresis was measured with an MST power of 70% and an LED power of 15%. In both cases, the MTR4 KOW concentration series was produced by serial dilution (1:1). Titrations were performed in triplicates and the data were analyzed using the Thermophoresis with T-Jump strategy option with the MO software (NanoTemper Technologies).

Isothermal calorimetry experiments were carried out using the ITC200 Isothermal titration calorimeter from Microcal. Before the measurement, all samples were dialysed against a buffer containing 50 mM Hepes/NaOH pH 7.5, 150 mM NaCl, and 0.5 mM TCEP. For NVL: MTR4-ΔN (the reactant) samples were concentrated to 40 μM and NVL (the injectant) to 500 μM. For human NOP53: MTR4-KOW (the reactant) samples were concentrated to 100 μM and NOP53 (the injectant) to 1 mM. Titrations were carried out at 25 °C with 2 μL of the injectant per injection added to 200 μL of reactant cell solution. All data were processed and curves fitted using Origin5.0.

Biochemical assays. For pull-down assays, appropriate protein mixtures were incubated in 50 mM Hepes (pH 7.5), 150 mM NaCl, 5% (v/v) glycerol, 0.05% (v/v) NP40, 1 mM DTT for 30 min at 4 °C. For ZCCHC8, 2 μM GST-MTR4ΔN was incubated with 8 μM MBP-ZCCHC8 constructs. For NVL, 10 μM GST-MTR4ΔN was incubated with 20 μM Trx-NVL constructs. For NOP53, 30 μM GST-MTR4ΔN was incubated with 60 μM Trx-$NOP53^{84–123}$ in a total volume of 30 μl. The protein mixtures were then incubated with Glutathione sepharose beads (GE healthcare) for 2 h. Post incubation, the beads were washed three times with 0.1 ml incubation buffer and the retained material was eluted with incubation buffer supplemented with 30 mM reduced glutathione. Input material (1–3%) and eluates (~30%) were analyzed by SDS-PAGE and Coomassie staining.

For ATPase assays, 150 pmol of MTR4ΔN or MTR4ΔN-containing complexes were incubated 40 nmol MESG (2-amino-6-mercapto-7-methylpurine ribonucleoside) and 0.5 U purine nucleoside phosphorylase (Enzchek Phosphate Assay kit, Invitrogen) in a buffer containing 50 mM MOPS pH 6.5, 50 mM Nacl, 2.5 mM $MgCl_2$, 5 mM β-mercaptoethanol and 5% (v/v) glycerol. For reactions containing RNA, the mixture was incubated with 2 μg poly-U RNA (Sigma). The reaction was initiated by addition of ATP to a final concentration of 1 mM. The generation of 2-amino-6-mercapto-7-methylpurine from MESG was monitored by measuring absorbance increase at 360 nm on a plate reader (Infinite M1000 Pro, Tecan) for 12 min at 60 s intervals. The data were normalized by subtracting the y-intercept from the raw data. The experiment was performed in triplicate. The mean ($n = 3$) and standard deviation (error bars) were plotted using Graphpad prism 8.

Helicase assays were performed essentially as described by Puno et al.[37]. A duplex RNA was formed by mixing a 5' Fluorescein amidite (FAM) labeled RNA (FAM-AGCACCGUAAAGACGC) with 1.5 molar excess of complementary RNA with a 25 A overhang (GCGUCUUUACGGUGCUAAAAAAAAAAAAAAAAAAAAAAAA) in a buffer containing 20 mM Tris-HCl pH 7.5 and 50 mM NaCl. To anneal the RNA duplex, the mixture was heated to 95 °C and allowed to cool down slowly to ambient temperature by turning off the heat block. 0.3 pmol of duplex RNA was incubated with 3.75, 7.5, 30 pmol of MTR4ΔN or MTR4ΔN-containing complexes in a buffer containing 50 mM MOPS pH 6.5, 50 mM Nacl, 0.5 mM $MgCl_2$, 5 mM β-mercaptoethanol and 5% (v/v) glycerol for 5 min at 30 °C. The reaction was then initiated by the addition of ATP, $MgCl_2$ and a trap DNA oligo complementary to the FAM labeled RNA (GCGTCTTTACGGTGCT) to a final concentration of 2 mM, 2 mM and 400 nM respectively. The reactions were quenched after 40 min by placing the tubes on ice and adding quenching buffer to a final concentration of 0.5% (w/v) SDS, 10 mM EDTA, 10% (v/v) glycerol and 0.005% (w/v) xylene cyanol. The sample were analyzed by electrophoresis on a 15% acrylamide–Tris base, boric acid, EDTA (TBE) gel. The fluorescence was imaged using a Typhoon FLA 7000. The oligonucleotides used in the assay were obtained from Ella Biotech GmBH.

NMR spectroscopy. NMR measurements of MTR4-KOW were performed in phosphate buffered saline (10 mM PO_4^{3-}, 137 mM NaCl, 2.7 mM KCl) mixed with 10% (v/v) D_2O. Backbone chemical shift assignments of the KOW region were obtained from two ^{13}C, ^{15}N-labelled samples with protein concentration of 700 μM and 500 μM, respectively. HNCA, HNCACB, HNcoCA, HNCO, HNcaCO, and 3D ^{15}N-edited NOESY spectra[46] were acquired at 298 K on Bruker Avance III spectrometers at field strengths corresponding to 600 and 800 proton Larmor frequency, equipped with TCI cryogenic probe heads. The ^{15}N steady-state heteronuclear $\{^1H\}$-^{15}N NOE experiment was performed at 170 μM and a field strength of 500 MHz as described previously[47]. Protein binding was measured from HSQC experiments containing water-flip-back/WATERGATE[48,49] sequences. Titrations with the NVL peptide were carried out at 298 K with a KOW concentration of 53 μM and in presence of 0.25, 0.5, 1, 2, and 6 stoichiometric molar equivalents of NVL. Spectra were recorded and processed with Topspin3.5 and analyzed with CCPNMR Analysis 2.4[50] and Sparky (http://www.cgl.ucsf.edu/home/sparky). The chemical shift perturbations were calculated as CSP (ppm) = $[6(\Delta H)^2 + (\Delta N)^2]^{0.5}$.

Crystallization and structure determination. $MTR4\Delta N_2$ (70–1042) was mixed with 2 molar excess of NVL (167–216) in a buffer containing 20 mM Tris pH 7.5, 150 mM Nacl, 1 mM $Mgcl_2$, 2 mM ADP and 1 mM TCEP. The crystallization trials were performed using a vapour diffusion setup. Initial crystals were obtained in the A4 (2 M ammonium sulfate, 0.1 M Tris pH 8.5) condition of SG1™ screen (Molecular dimensions). The best diffracting crystals were obtained at a concentration of 10 mg/ml at 277 K in 0.1 M Tris pH 8, 1.8 M ammonium sulfate. The crystals were cryo-protected with reservoir solution supplemented with 30% glycerol prior flash-freezing in liquid nitrogen. Data were collected at 100 K at PXIII beamline of the Swiss light Source (Villigen, Switzerland). Data processing and scaling was performed using Xia2/DIALS[51,52] within CCP4i2 software suite[53]. The crystals belong to the hexagonal spacegroup 169 ($P6_1$) containing one molecule ($MTR4\Delta N_2$ – $NVL^{167–216}$) in the asymmetric unit and diffract to 3.07 Å resolution. The structure was solved by molecular replacement with Phaser[54] within Phenix using the co-ordinates of the DExH core (98-593 and 847–1042) and KOW domain (645–787) of MTR4 (PDB 6IEH)[31] as search models. The model was manually completed in COOT[55] and refined using phenix.refine[56]. The optimal TLS groups for TLS refinement were determined using TLSMD server[57]. 96.1% of the protein backbone dihedral angles in the final model are in Ramachandran favored region. The figures of crystallographic models were prepared using pyMOL (Schrödinger, LLC).

Cell culture and co-immunoprecipitation assays. Human HeLa Kyoto LAP, MTR4-LAP, RBM7-LAP cell lines were generated as outlined by Poser et al.[58] and used for example by Lubas et al.[27] (RBM7-LAP) and Meola et al.[30] (MTR4-LAP). Briefly, HeLa Kyoto LAP MTR4-LAP, RBM7-LAP cells were cultured in DMEM + 10% FBS. Transfection was carried out using 10 μg of plasmid containing FLAG-tagged NVL or ZCCHC8 constructs, with Lipofectamine 2000 (Thermo Fisher) following the manufacturer's instructions. Forty-eight hours after transfection the cells were collected and resuspended in extraction buffer (150 mM NaCl, 20 mM HEPES pH 7.4, 0.5% (v/v) Triton X-100) containing Protease Inhibitors (Roche). The lysates were sonicated twice for 5 s at 20 W, and cell debris was removed by centrifugation at 10,000 g for 10 min. Lysates were incubated 1 h with Dynabeads M-270 Epoxy (Invitrogen) coupled to a polyclonal llama anti-GFP antibody. Beads were washed three times in extraction buffer, then incubated 20 min at 25 °C with 100 units of Benzonase (Sigma) and 2 mM $MgCl_2$. Beads were then washed twice in extraction buffer and proteins were eluted using NuPage LDS Sample Buffer (Invitrogen) at 70 °C for 10 min, and NuPage Sample Reducing Agent was added. The input material (0.5%) and the eluate (8.0%) were analyzed by SDS-PAGE on a NuPage Novex 4–12% Bis-Tris gel (Invitrogen). Western blotting analysis was performed following standard protocols. Following primary antibodies were used: anti-FLAG M2 (Dilution 1:50000; Sigma, F1804), anti-MTR4 (Dilution 1:4000; Abcam, ab70551), anti-GFP (Dilution 1:1000; Santa Cruz Biotechnology, SC-9996),

anti-β-actin (Dilution 1:100000; Sigma, A2228) and anti-RBM7 (Dilution 1:1000; human protein atlas, HPA013993). Anti-mouse and anti-rabbit secondary antibodies coupled to Horseradish Peroxidase (Dako) were used.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

A reporting summary for this Article is available as a Supplementary Information file. NMR backbone chemical shifts of the human MTR4 KOW domain were deposited at the BMRB under accession number 27831. The coordinates and the structure factors have been deposited in the Protein Data Bank with accession code PDB ID 6RO1. The source data are provided in Supplementary Fig. 10. All data is available from the corresponding author upon reasonable request.

Received: 14 January 2019 Accepted: 7 July 2019
Published online: 29 July 2019

References

1. Chlebowski, A., Lubas, M., Jensen, T. H. & Dziembowski, A. RNA decay machines: the exosome. *Biochim Biophys. Acta* **1829**, 552–560 (2013).
2. Zinder, J. C. & Lima, C. D. Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. *Genes Dev.* **31**, 88–100 (2017).
3. Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M. & Tollervey, D. The exosome: a conserved eukaryotic RNA processing complex containing multiple 3′–>5′ exoribonucleases. *Cell* **91**, 457–466 (1997).
4. Makino, D. L., Halbach, F. & Conti, E. The RNA exosome and proteasome: common principles of degradation control. *Nat. Rev. Mol. Cell Biol.* **14**, 654–660 (2013).
5. Dziembowski, A., Lorentzen, E., Conti, E. & Seraphin, B. A single subunit, Dis3, is essentially responsible for yeast exosome core activity. *Nat. Struct. Mol. Biol.* **14**, 15–22 (2007).
6. Liu, Q., Greimann, J. C. & Lima, C. D. Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell* **127**, 1223–1237 (2006).
7. Butler, J. S. & Mitchell, P. Rrp6, Rrp47 and cofactors of the nuclear exosome. *Adv. Exp. Med Biol.* **702**, 91–104 (2010).
8. Milligan, L. et al. A yeast exosome cofactor, Mpp6, functions in RNA surveillance and in the degradation of noncoding RNA transcripts. *Mol. Cell Biol.* **28**, 5446–5457 (2008).
9. Schilders, G., Raijmakers, R., Raats, J. M. & Pruijn, G. J. MPP6 is an exosome-associated RNA-binding protein involved in 5.8S rRNA maturation. *Nucleic Acids Res* **33**, 6795–6804 (2005).
10. Falk, S., Bonneau, F., Ebert, J., Kogel, A. & Conti, E. Mpp6 incorporation in the nuclear exosome contributes to RNA channeling through the Mtr4 helicase. *Cell Rep.* **20**, 2279–2286 (2017).
11. Schuch, B. et al. The exosome-binding factors Rrp6 and Rrp47 form a composite surface for recruiting the Mtr4 helicase. *EMBO J.* **33**, 2829–2846 (2014).
12. Weick, E. M. et al. Helicase-dependent RNA decay illuminated by a Cryo-EM structure of a human nuclear RNA exosome-MTR4 complex. *Cell* **173**, 1663–1677 e21 (2018).
13. Gerlach, P. et al. Distinct and evolutionary conserved structural features of the human nuclear exosome complex. *Elife* **7**, e38686 (2018).
14. Jackson, R. N. et al. The crystal structure of Mtr4 reveals a novel arch domain required for rRNA processing. *EMBO J.* **29**, 2205–2216 (2010).
15. Kilchert, C., Wittmann, S. & Vasiljeva, L. The regulation and functions of the nuclear RNA exosome complex. *Nat. Rev. Mol. Cell Biol.* **17**, 227–239 (2016).
16. Johnson, S. J. & Jackson, R. N. Ski2-like RNA helicase structures: common themes and complex assemblies. *RNA Biol.* **10**, 33–43 (2013).
17. Lubas, M. et al. Interaction profiling identifies the human nuclear exosome targeting complex. *Mol. Cell* **43**, 624–637 (2011).
18. Ogami, K., Chen, Y. & Manley, J. L. RNA surveillance by the nuclear RNA exosome: mechanisms and significance. *Noncoding RNA* **4**, 8 (2018).
19. Vanacova, S. et al. A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS Biol.* **3**, e189 (2005).
20. LaCava, J. et al. RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* **121**, 713–724 (2005).
21. Wyers, F. et al. Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* **121**, 725–737 (2005).
22. Kadaba, S., Wang, X. & Anderson, J. T. Nuclear RNA surveillance in Saccharomyces cerevisiae: Trf4p-dependent polyadenylation of nascent hypomethylated tRNA and an aberrant form of 5S rRNA. *RNA* **12**, 508–521 (2006).
23. Carneiro, T. et al. Depletion of the yeast nuclear exosome subunit Rrp6 results in accumulation of polyadenylated RNAs in a discrete domain within the nucleolus. *Mol. Cell Biol.* **27**, 4157–4165 (2007).
24. Hiraishi, N., Ishida, Y. & Nagahama, M. AAA-ATPase NVL2 acts on MTR4-exosome complex to dissociate the nucleolar protein WDR74. *Biochem Biophys. Res Commun.* **467**, 534–540 (2015).
25. Hiraishi, N., Ishida, Y. I., Sudo, H. & Nagahama, M. WDR74 participates in an early cleavage of the pre-rRNA processing pathway in cooperation with the nucleolar AAA-ATPase NVL2. *Biochem Biophys. Res Commun.* **495**, 116–123 (2018).
26. Gustafson, M. P., Welcker, M., Hwang, H. C. & Clurman, B. E. Zcchc8 is a glycogen synthase kinase-3 substrate that interacts with RNA-binding proteins. *Biochem Biophys. Res Commun.* **338**, 1359–1367 (2005).
27. Lubas, M. et al. The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. *Cell Rep.* **10**, 178–192 (2015).
28. Ogami, K. et al. An Mtr4/ZFC3H1 complex facilitates turnover of unstable nuclear RNAs to prevent their cytoplasmic transport and global translational repression. *Genes Dev.* **31**, 1257–1271 (2017).
29. Silla, T., Karadoulama, E., Makosa, D., Lubas, M. & Jensen, T. H. The RNA exosome adaptor ZFC3H1 functionally competes with nuclear export activity to retain target transcripts. *Cell Rep.* **23**, 2199–2210 (2018).
30. Meola, N. et al. Identification of a nuclear exosome decay pathway for processed transcripts. *Mol. Cell* **64**, 520–533 (2016).
31. Wang, J. et al. NRDE2 negatively regulates exosome functions by inhibiting MTR4 recruitment and exosome interaction. *Genes Dev* **33**, 536–549 (2019).
32. Weir, J. R., Bonneau, F., Hentschel, J. & Conti, E. Structural analysis reveals the characteristic features of Mtr4, a DExH helicase involved in nuclear RNA processing and surveillance. *Proc. Natl Acad. Sci. USA* **107**, 12139–12144 (2010).
33. Schuller, J. M., Falk, S., Fromm, L., Hurt, E. & Conti, E. Structure of the nuclear exosome captured on a maturing preribosome. *Science* **360**, 219–222 (2018).
34. Falk, S. et al. The molecular architecture of the TRAMP complex reveals the organization and interplay of its two catalytic activities. *Mol. Cell* **55**, 856–867 (2014).
35. Thoms, M. et al. The exosome is recruited to RNA substrates through specific adaptor proteins. *Cell* **162**, 1029–1038 (2015).
36. Falk, S. et al. Structural insights into the interaction of the nuclear exosome helicase Mtr4 with the preribosomal protein Nop53. *RNA* **23**, 1780–1787 (2017).
37. Puno, M. R. & Lima, C. D. Structural basis for MTR4-ZCCHC8 interactions that stimulate the MTR4 helicase in the nuclear exosome-targeting complex. *Proc. Natl Acad. Sci. USA* **115**, E5506–E5515 (2018).
38. Nagahama, M. et al. The AAA-ATPase NVL2 is a component of pre-ribosomal particles that interacts with the DExD/H-box RNA helicase DOB1. *Biochem. Biophys. Res Commun.* **346**, 1075–1082 (2006).
39. Fujiwara, Y. et al. Structure and function of the N-terminal nucleolin binding domain of nuclear valosin-containing protein-like 2 (NVL2) harboring a nucleolar localization signal. *J. Biol. Chem.* **286**, 21732–21741 (2011).
40. Meyer, H., Bug, M. & Bremer, S. Emerging functions of the VCP/p97 AAA-ATPase in the ubiquitin system. *Nat. Cell Biol.* **14**, 117–123 (2012).
41. Hanzelmann, P. & Schindelin, H. The structural and functional basis of the p97/valosin-containing protein (VCP)-interacting motif (VIM): mutually exclusive binding of cofactors to the N-terminal domain of p97. *J. Biol. Chem.* **286**, 38679–38690 (2011).
42. Buchberger, A., Schindelin, H. & Hanzelmann, P. Control of p97 function by cofactor binding. *FEBS Lett.* **589**, 2578–2589 (2015).
43. Falk, S. et al. Structure of the RBM7-ZCCHC8 core of the NEXT complex reveals connections to splicing factors. *Nat. Commun.* **7**, 13573 (2016).
44. Cote, J. & Richard, S. Tudor domains bind symmetrical dimethylated arginines. *J. Biol. Chem.* **280**, 28476–28483 (2005).
45. Friberg, A., Oddone, A., Klymenko, T., Muller, J. & Sattler, M. Structure of an atypical tudor domain in the drosophila polycomblike protein. *Protein Sci.* **19**, 1906–1916 (2010).
46. Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. NMR Spectrosc.* **34**, 93–158 (1999).
47. Farrow, N. A. et al. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry* **33**, 5984–6003 (1994).
48. Grzesiek, S. & Bax, A. The importance of not saturating H_2O in protein NMR - application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc.* **115**, 12593–12594 (1993).
49. Piotto, M., Saudek, V. & Sklenář, V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* **2**, 661 (1992).

50. Vranken, W. F. et al. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696 (2005).
51. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190 (2010).
52. Winter, G. et al. DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr D. Struct. Biol.* **74**, 85–97 (2018).
53. Potterton, L. et al. CCP4i2: the new graphical user interface to the CCP4 program suite. *Acta Crystallogr. D. Struct. Biol.* **74**, 68–84 (2018).
54. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr* **40**, 658–674 (2007).
55. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr* **66**, 486–501 (2010).
56. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta. Crystallogr. D Biol. Crystallogr.* **66**, (213–221 (2010).
57. Painter, J. & Merritt, E. A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D. Biol. Crystallogr* **62**, 439–450 (2006).
58. Poser, I. et al. BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nat. Methods* **5**, 409–415 (2008).
59. Di Tommaso, P. et al. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res* **39**, W13–W17 (2011).

Acknowledgements

We would like to thank the Crystallization Facility of MPI Biochemistry and also Vincent D. Maciej for his contributions at the early stages of this project. This study was supported by the Max Planck Gesellschaft, the European Commission (ERC Advanced Investigator Grant 294371) and the Deutsche Forschungsgemeinschaft (SFB646, SFB1035, GRK1721, FOR1680 to E.C., and SFB1035 and GRK1721 to M.S. This work was supported by the Cluster of Excellence EXC114 (to E.C. and M.S.). T.H.J. was supported by the ERC (grant 339953), the Lundbeck- and the Novo Nordisk-Foundations.

Author contributions

S.F and E.C. initiated the project; M.L. performed the in vitro experiments; D.J. performed the co-IP experiments under the supervision of T.H.J.; M.L. and J.B. performed the crystallography experiments, M.L. and S.F. built, refined and analyzed the structure; A.S. and M.S. collected and analyzed NMR data; L.L. performed preliminary in vitro experiments with NVL under the supervision of S.F.; M.L., S.F., and E.C. prepared the paper.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-11339-x.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Peer review information: *Nature Communications* thanks David Tollervey, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

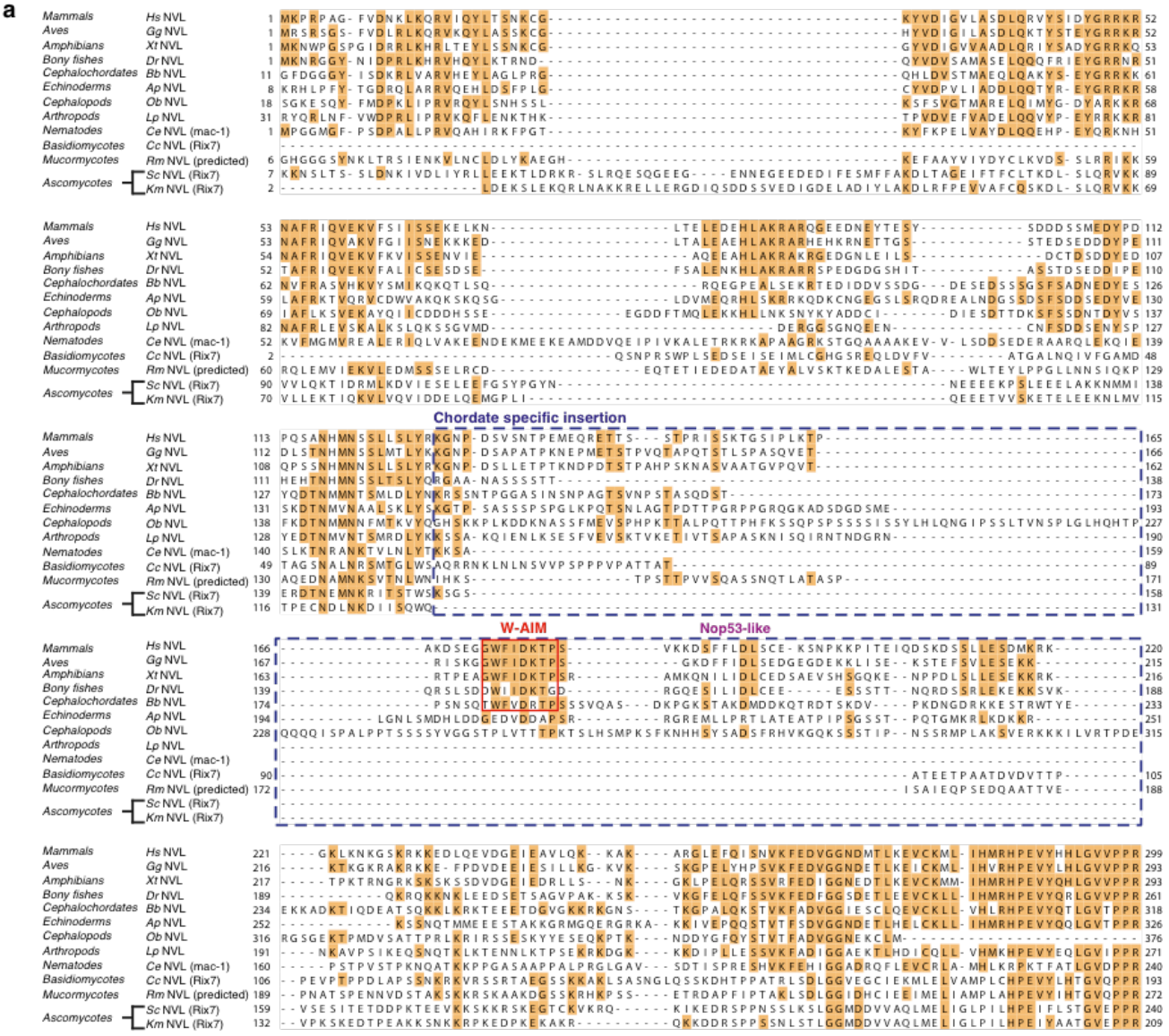
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

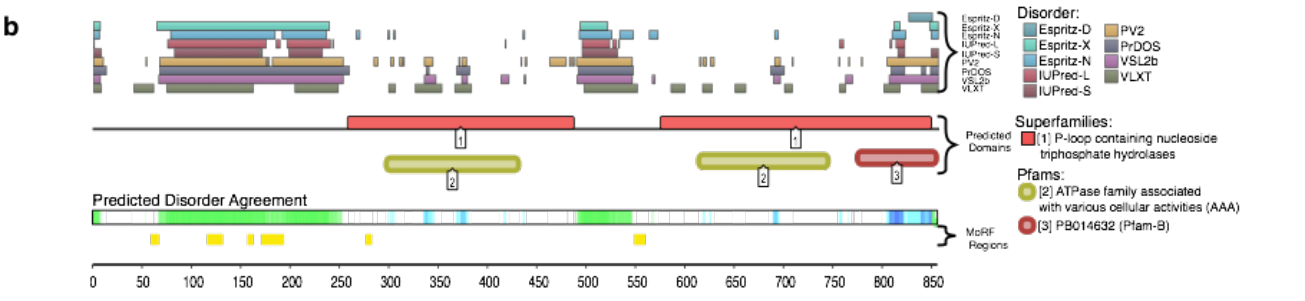
SUPPLEMENTARY INFORMATION

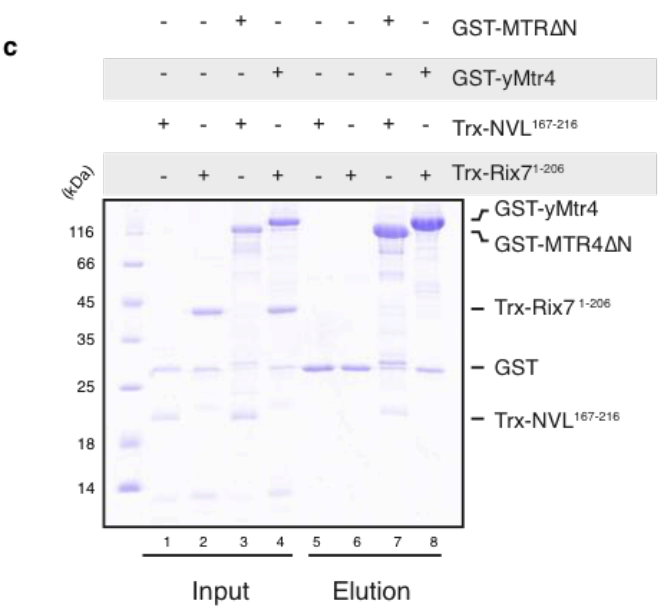
The MTR4 helicase recruits nuclear adaptors of the human RNA exosome using distinct arch-interacting motifs

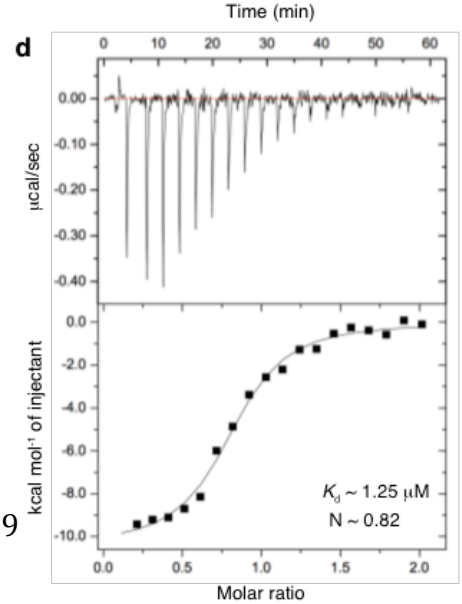
Lingaraju et al.

Supplementary Figure 1: Vertebrate specific N-terminal insertion in NVL interacts with MTR4



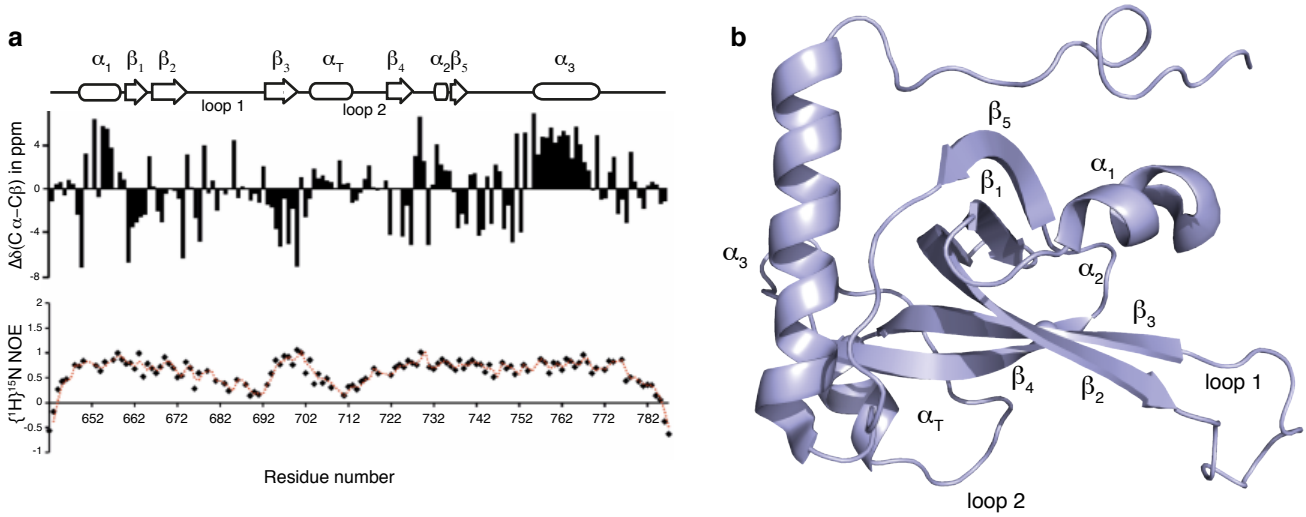






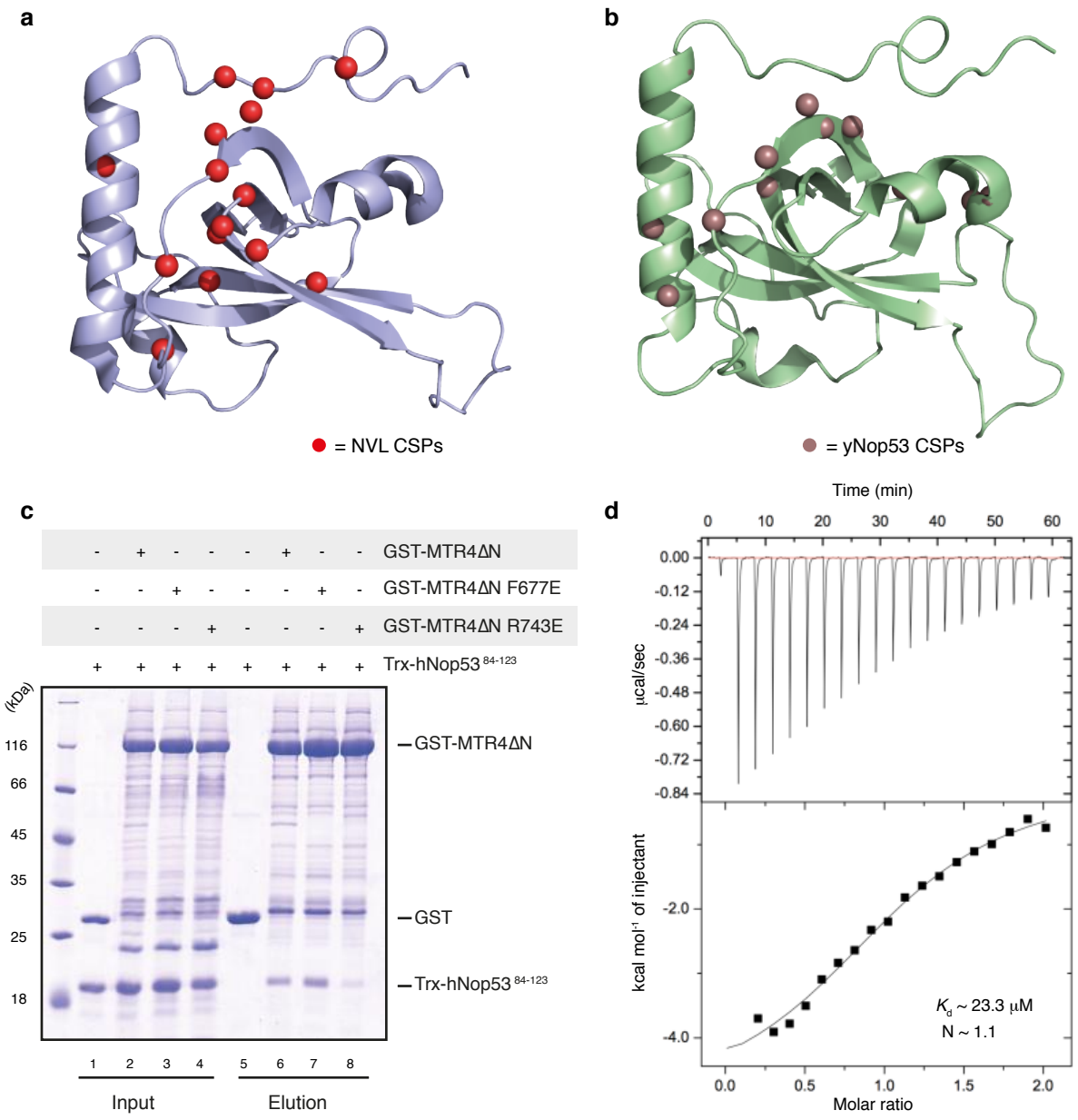
Supplementary Figure 1 (a) Sequence alignment of N-termini of representative vertebrate and fungal NVL sequences, highlighting the chordate specific insertion (blue dashed box), the W-AIM (red box) and a Nop53 (LFXϕD) like region (purple). The sequences were obtained from the NCBI database and aligned using the T-coffee server[1]. *Hs* stands for *Homo sapiens*, *Gg* for *Gallus gallus*, *Xt* for *Xenopus tropicalis*, *Dr* for *Danio rerio*, *Bb* for *Branchiostoma belcheri*, *Ap* for *Acanthaster planci*, *Ob* for *Octopus bimaculoides*, *Lp* for *Limulus polyphemus*, *Ce* for *Caenorhabditis elegans*, *Cc* for *Coprinopsis cineria*, Rm for *Rhizopus microsporus*, *Sc* for *Saccharomyces cerevisiae*, and *Km* for *Kluyveromyces marxianus*. (b) Disorder prediction of human NVL as obtained from the D^2P^2 database[2]. The region of NVL that is commonly predicted to be disordered by multiple algorithms is colored in bright green. (c) Protein co-precipitations by pull down assays. GST tagged MTR4ΔN or yMtr4 were incubated with the vertebrate specific NVL insertion, Trx-$NVL^{167-216}$ and the N-terminus of Rix7, Trx-$Rix7^{1-206,}$ respectively before co-precipitation with glutathione sepharose beads. A total of 3% of the input (left) and 30% of the eluates (right) were analyzed on 15% SDS-PAGE gels and visualized by staining with coomassie brilliant blue. (d) ITC experiment of MTR4ΔN with $NVL^{167-216}$. The filled squares show reference corrected titration of $NVL^{167-216}$ into the MTR4ΔN containing cell. The number of calculated binding sites (N), and dissociation constants (K_d) are shown in the inset.

Supplementary Figure 2: NMR analysis of the MTR4 KOW domain



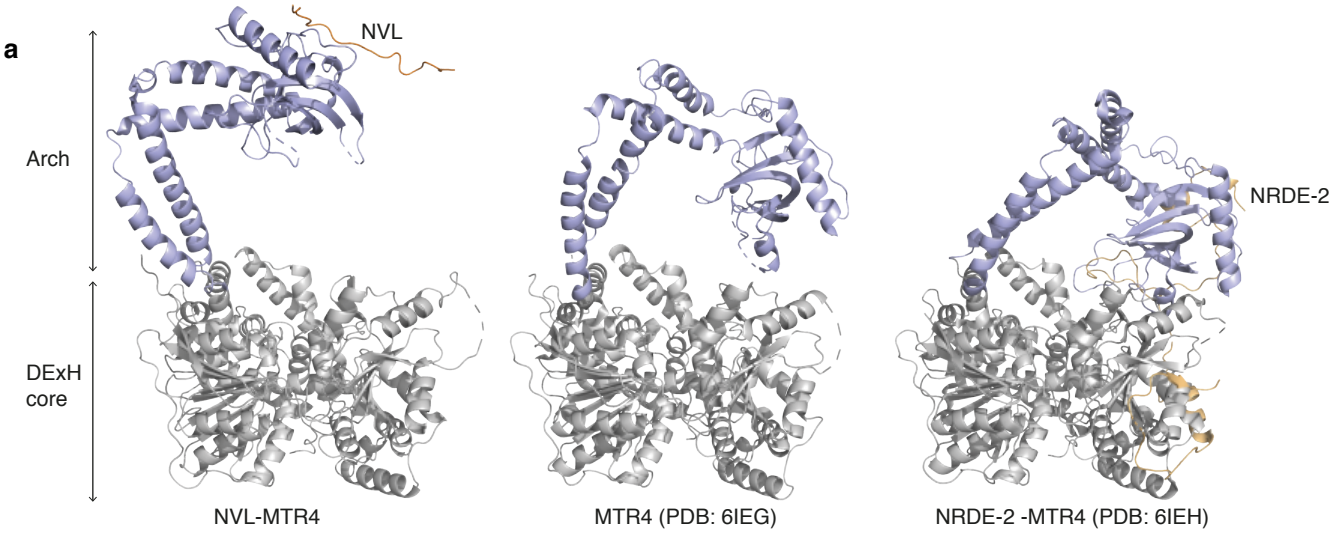
Supplementary Figure 2 (a) Upper bar chart shows secondary carbon chemical shifts of MTR4 KOW plotted against its primary sequence. The clusters of positive bars >2 represent α-helices and the clusters of negative bars >-2 represent β-strands. The scheme above shows a summary of secondary structure elements as derived from the analysis, and elements are labeled for comparison with panel b. The bottom chart shows a heteronuclear NOE plot of the MTR4 KOW demonstrating the residue-resolved rigidity along the primary sequence. Note that the two prominent dips are located within loops 1 and 2. (b) Secondary structure features of the KOW domain obtained from NMR analysis mapped on to the KOW domain from the crystal structure of MTR4 (PDB 6IEH)[3].

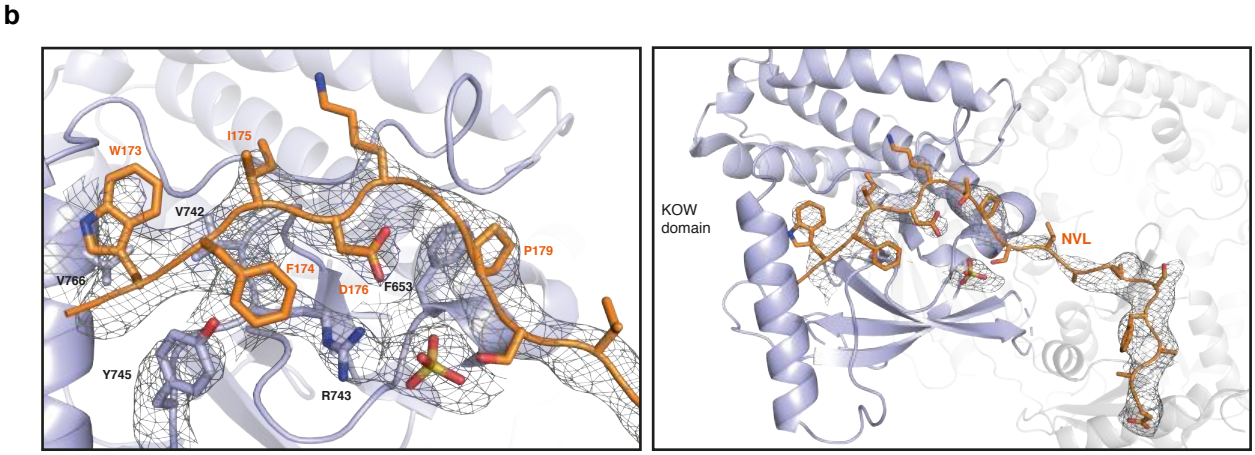
Supplementary Figure 3: NVL and Nop53 interact with MTR4 KOW in a similar manner

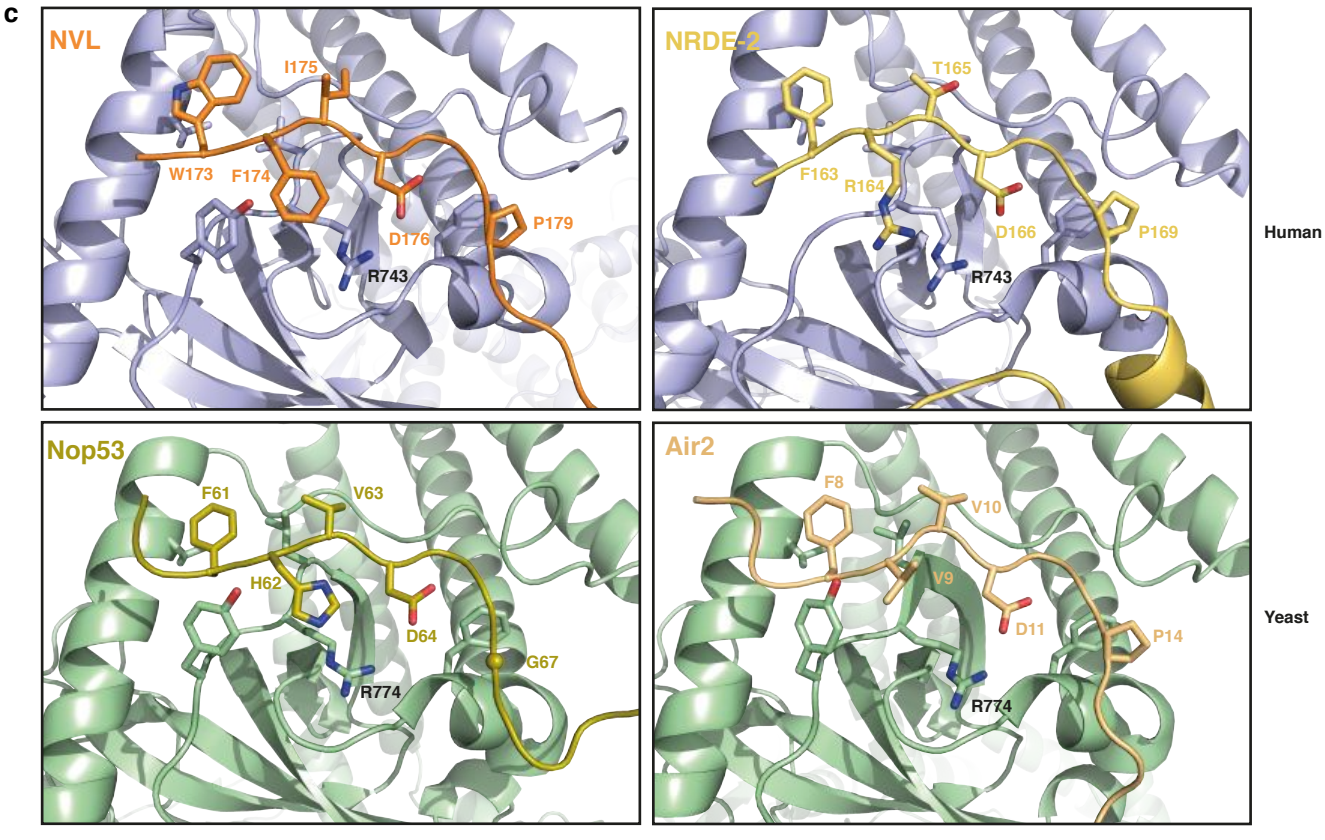


Supplementary Figure 3 (a) Structural model of MTR4 KOW (PDB 6IEH)[3] showing regions of significant backbone chemical shift perturbation (CSPs) upon NVL titration highlighted as red spheres. (b) yMtr4 KOW (PDB 5OOQ) showing significant CSPs (orange spheres) upon yNop53 titration as reported by Falk et al[4]. (c) Protein co-precipitations by pull down assays. GST tagged MTR4ΔN or the corresponding MTR4ΔN mutants were incubated with the AIM containing region of human Nop53 (Trx-hNop53 $^{84-123}$) before co-precipitation with glutathione sepharose beads. A total of 1% of the input (left) and 30% of the eluates (right) were analyzed on 15% SDS-PAGE gels and visualized by staining with coomassie brilliant blue. (d) ITC experiment of MTR4 KOW with hNop53 $^{84-123}$. The filled squares show reference corrected titration of hNop53 $^{84-123}$ into the MTR4 KOW containing cell. The number of calculated binding sites (N), and dissociation constants (K_d) are shown in the inset.

Supplementary Figure 4: Features of the NVL-MTR4 crystal structure

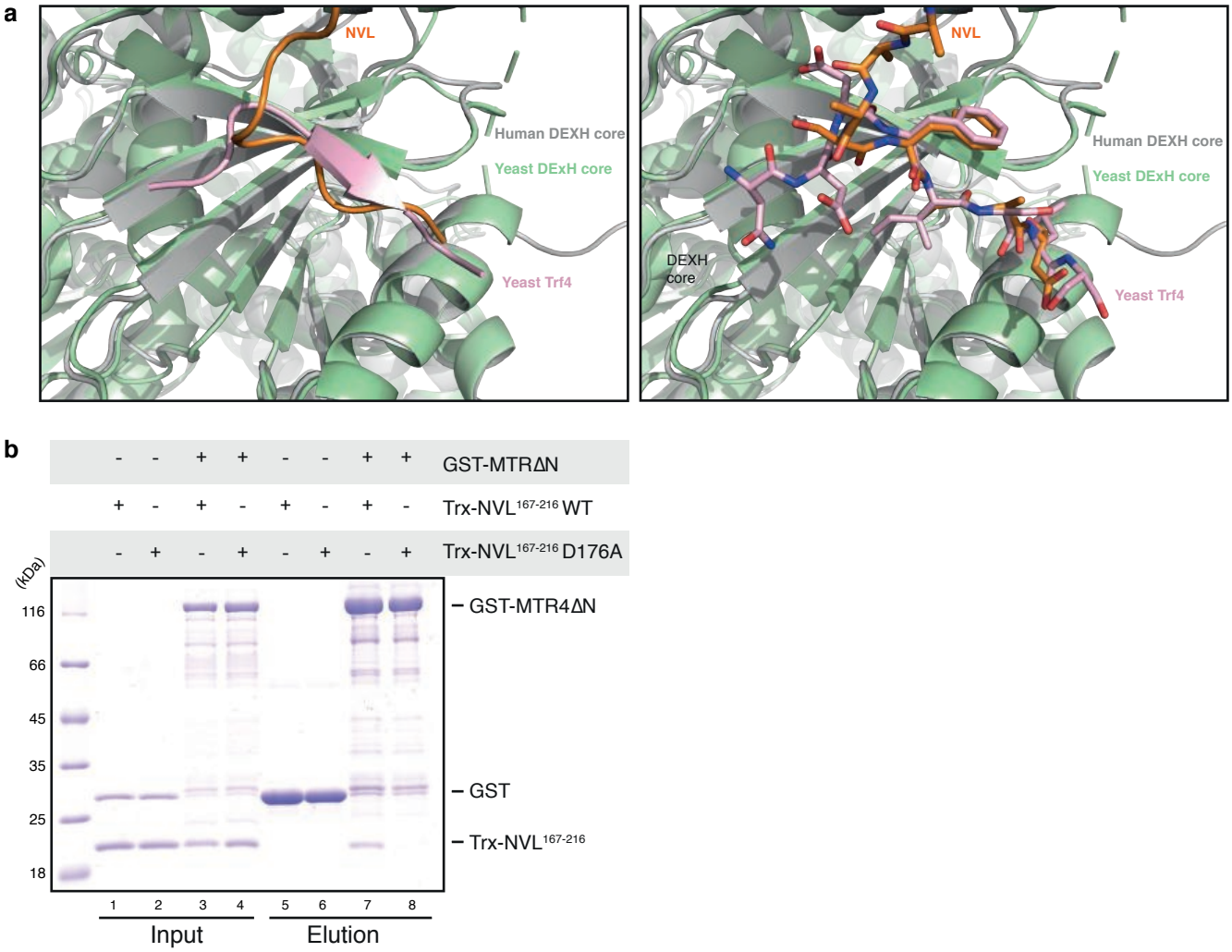






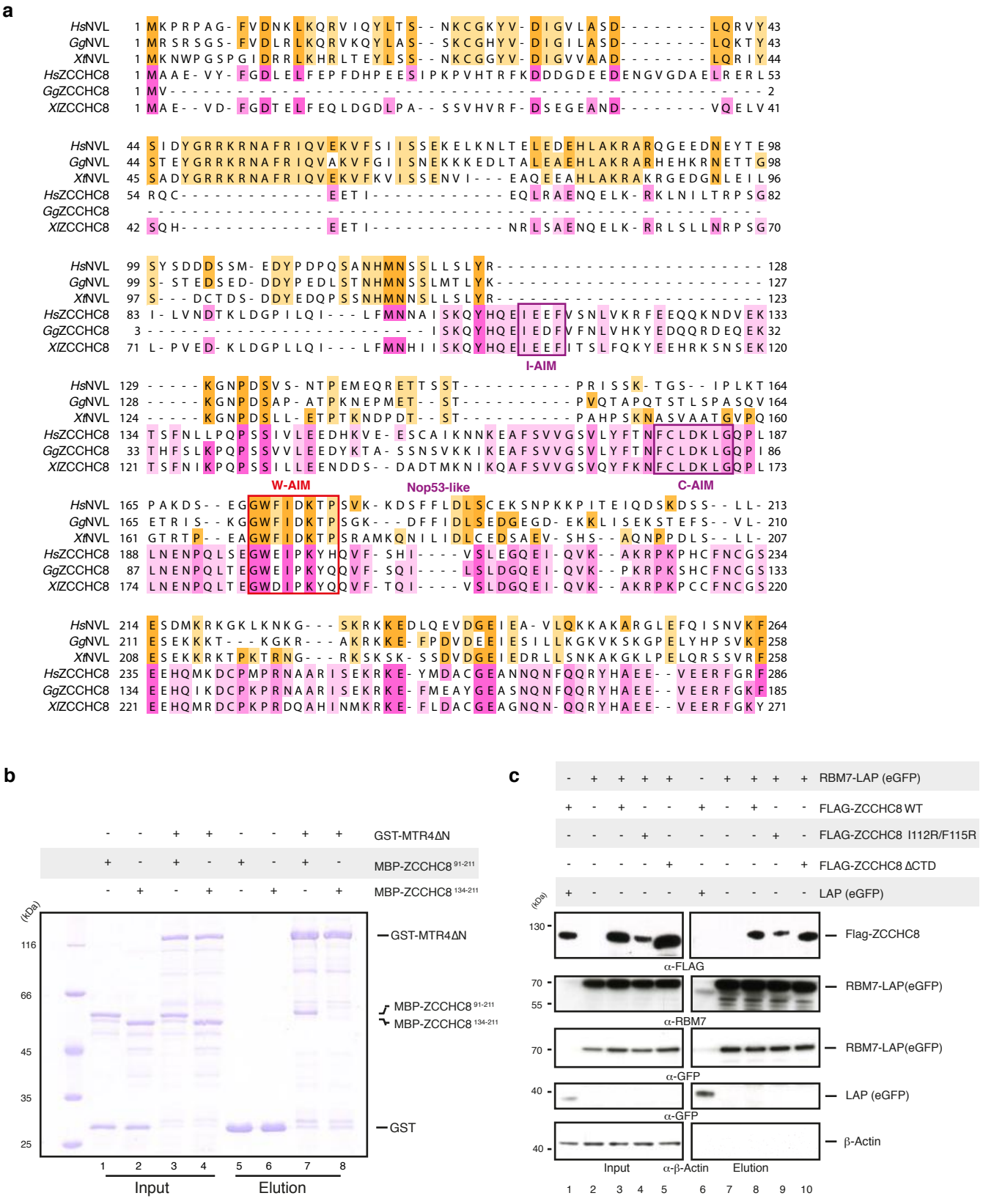
Supplementary Figure 4 (a) Comparison of the crystal structure of the NVL-MTR4 complex with crystal structure of MTR4 alone and the NRDE-2-MTR4 complex (PDB 6IEG and 6IEH)[3]. The structures were aligned based on the DExH core region (colored in grey) to represent the difference in orientation of the arch region (colored in light blue) with respect to the DExH core. (b) Zoom-in view of the interactions between MTR4 KOW domain (light blue) and NVL (orange) as displayed in Fig. 3d. The model is overlaid with the refined 2mFo-DFc map (grey mesh) showing the density for NVL and the interacting residues of MTR4. The residues of interest are labeled and the map is contoured at 1.0σ. (c) Zoom in view of MTR4-NVL crystal structure showing the ordered region of NVL (orange). The model is overlaid with the 2mFo-DFc omit map calculated in PHENIX omitting the NVL residues. The map is contoured at 1.0σ. (d) Zoom-in view of the KOW-AIM interfaces in NVL-MTR4, NRDE-2-MTR4 (PDB 6IEH[3]), Nop53-Mtr4 (PDB 5OOQ[4]) and Trf4-Air2-Mtr4 (PDB 4U4C[5]) crystal structures. KOW domains of the human MTR4 and yeast Mtr4 are colored in light blue and green respectively. NVL, NRDE-2, Nop53 and Air2 are colored in orange, yellow orange, olive and pale orange respectively and the residues of interest are labeled.

Supplementary Figure 5: Features of the NVL-MTR4 crystal structure and structure based mutagenesis



Supplementary Figure 5 (a) Superposition of DExH cores NVL-MTR4 (grey) and Trf4-Air2-Mtr4 (light green; PDB 4U4C[5]) showing that the NVL fragment (orange) from a symmetry related molecule aligns with the Trf4 fragment in Trf4-Air2-Mtr4 structure. NVL and Trf4 are shown in cartoon representation (left panel) and in stick representation (right panel) for clarity. (b) Protein co-precipitations by pull down assays. GST tagged MTR4ΔN was incubated with Trx-NVL$^{167-216}$ WT or Trx-NVL$^{167-216}$ D176A before co-precipitation with glutathione sepharose beads. A total of 1% of the input (left) and 30% of the eluates (right) were analyzed on 12% SDS-PAGE gels and visualized by staining with coomassie brilliant blue.

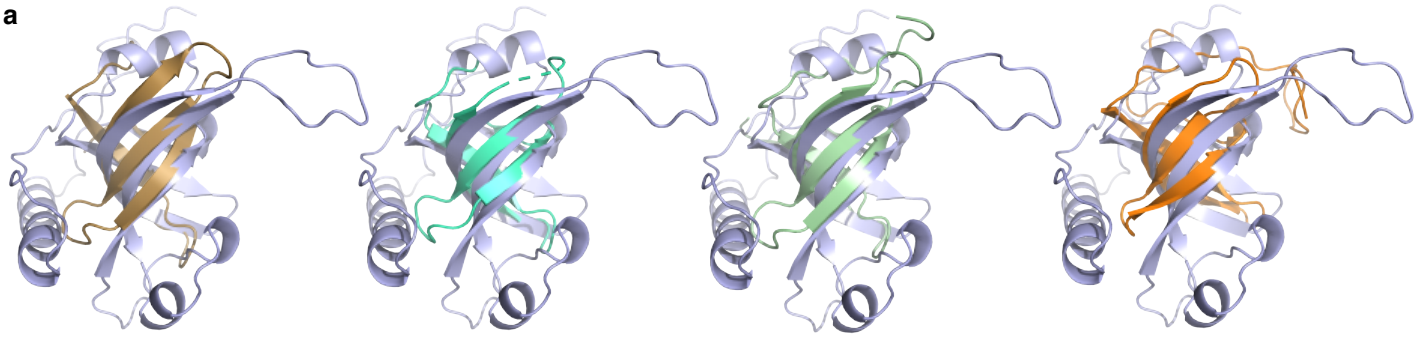
Supplementary Figure 6: ZCCHC8 harbors both canonical and non-canonical AIMs



Supplementary Figure 6 (a) Sequence alignment of the N-termini of NVL (orange) and ZCCHC8 (pink), highlighting W-AIM (red box), the respective Nop53-like C-AIM (LFXϕD) and ZCCHC8 specific I-AIM (purple boxes). The sequences were obtained from the UniProt database and aligned using the T-coffee server[1]. *Hs* stands for *Homo sapiens*, Gg for *Gallus gallus*, *Xt* for *Xenopus tropicalis and, Xl* for *Xenopus laevis*. (b) Protein co-precipitations by pull down assays. GST tagged MTR4ΔN was incubated with MBP-ZCCHC8$^{91-211}$ or MBP-ZCCHC8$^{134-211}$ before co-precipitation with glutathione sepharose beads. A total of 1% of the input (left) and 30% of the eluates (right) were analyzed on 10% SDS-PAGE gels and visualized by staining with coomassie brilliant blue. (c) Cellular co-IP assay. FLAG-tagged ZCCHC8 constructs (WT, IF-mutant, CTD-deletion) were transiently expressed in cells stably expressing RBM7-LAP. After precipitation of RBM7 taking advantage of the LAP tag, a total of 0.5% of the input (left) and 8.0% of the eluates (right) were analyzed on 4-12% SDS-PAGE gel followed by western blotting. The primary antibody used is indicated below the panel.

Supplementary Figure 7: Putative ligand binding phenylalanine in tudor domains is conserved in MTR4 KOW

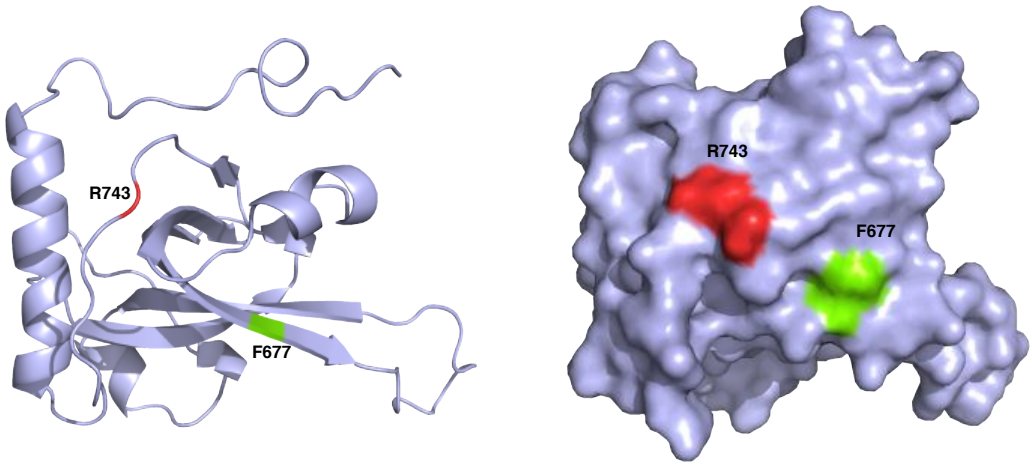
a



b

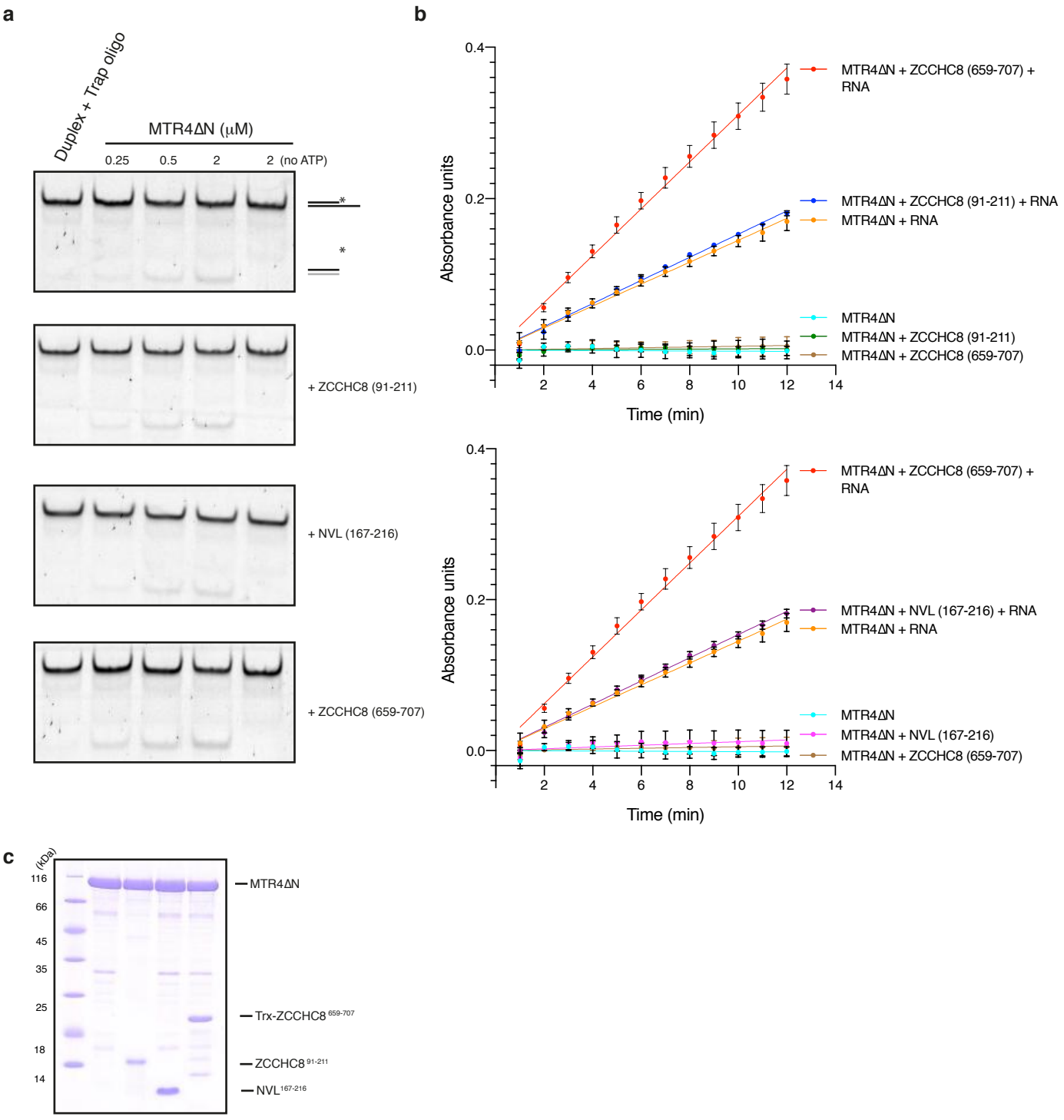
```
HsMTR4 KOW     645 - - - - - - - HKPKYC - - LPFL - - - - - - - - - - QPGRLVKVKNEG - DDFGWGVVVNFSKKSNVKP 685
CeMTR4 KOW     632 - - - - - - - REPKYI - - VGFL - - - - - - - - - - HAGRLFKVKSGD - RDFKWGILNQFKKEQNPD - 671
DmMTR4 KOW     661 - - - - - - - TKPQYL - - LPFL - - - - - - - - - - QPGRLVKVAAGS - QEYDWGIVLNFKKQDQSRK 701
ScMtr4 KOW     665 - - - - - - - THPANA - - LSFL - - - - - - - - - - QPGRLVEISVNGKDNYGWGAVVDFAKRINKRN 706
HsSMN Tudor     82 K - - - - - - - KNTAASLQQWKVGDKCSAIWSEDGCIYPATIASIDFKRETCVVVYTG - - - - - - 129
HsTDRD3 Tudor  555 M - - - - - - - - - - - - - - - - - WKPGDECFALYWEDNKFYRAEVEALHSSGMTAVVKFID - - - - - - 593
HsSPF30 Tudor   65 A - - - - - - - STQP - - THSWKVGDKCMAVWSEDGQCYEAEIEEIDEENGTAAITFAG - - - - - - 110
DmPcl Tudor    339 GAMAPPVAAPSP - - AVTYALQEDV - FIKCNDGRFYLGTI - - IDQTSDQYLIRFDD - - - - - - 388

HsMTR4 KOW     686 NSGE - L - DPLYVVEVLLRCS - - - - - - - KES - LKNSATEAAKPAKPDEKGEMQVVPVLVHLL 736
CeMTR4 KOW     672 - - DR - N - DQIYLCDMMIAIN - - - - - - - TEGRFDPTNPATLVPGFDLPKRRWIRVPMTIDRI 721
DmMTR4 KOW     702 NPLK - A - EPSVTIDVLLHVS - - - - - - - EAA - AKTG - - - DTEPCKPNERGCMEVVPVAHTLI 749
ScMtr4 KOW     707 PSAVYTDHESYIVNVVVNTMYIDSPVNLLKPFNPTLPEGIRPAEEGEKSICAVIPITLDSI 767
HsSMN Tudor    130 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - YGNREEQNLS - DLL 142
HsTDRD3 Tudor  594 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - YGNYEEVLLS - N - I 605
HsSPF30 Tudor  111 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - YGNAEVTPLL - N - L 122
DmPcl Tudor    389 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - QSEQWCEP - D - K - L 399

HsMTR4 KOW     737 SAISSVRLYIPKDLRPVDNRQSVLKSIQEVQKRFPDGIPLLDPIDDMGIQD - 787
CeMTR4 KOW     722 TAISAVRLKVPADIDKPDGQMRLDGMMAAATKRFGNQIPLLDPIQDMEIKTV 773
DmMTR4 KOW     750 TQISSIRVYFPNDLRSADNRRAVLKTIQEAKKRFPLGPPVLNPIDDMNIKD - 800
ScMtr4 KOW     768 KSIGNLRLYMPKDIRASGQKETVGKSLREVNRRFPDGIPVLDPVKNMKIED - 818
HsSMN Tudor    143 SPICEVANNIEQNAQENENESQ - - - - - - - - - - - - - - - - - - - - - - VSTD - E 169
HsTDRD3 Tudor  606 KPI - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 608
HsSPF30 Tudor  123 KPVEEG - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 128
DmPcl Tudor    400 RKLGGG - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 40
```

c



Supplementary Figure 7 (a) Structural superposition of MTR4 KOW model (light blue) (PDB 6IEH[3]) with structures of SMN tudor domain (gold) (PDB 1G5V)[6], TDRD3 tudor domain (turquoise) (PDB 3S6W)[7], SPF30 tudor domain (smudge green) (PDB 4A4F)[8] and Pcl tudor domain (orange) (PDB 2XK0)[9]. The superpositions were performed in PyMOL graphics system, version 2.2 (Schrödinger LLC). (b) Sequence alignment of representative MTR4 KOW (blue) and tudor (purple) domains, highlighting hydrophobic core residues (brown box) and putative substrate binding residues (red box) of tudor domains. The hydrophobic core resides conserved between KOW and tudor sequences are marked with brown rectangles (▪). The residues of interest in MTR4 KOW, Phe677 and Arg743, are marked by green and red asterisks respectively. The sequences were obtained from UniProt database and aligned using the T-coffee server[1]. *Hs* stands for *Homo sapiens*, *Ce* for *Caenorhabditis elegans*, *Dm* for *Drosphila melanogaster*, and *Sc* for *Saccharomyces cerevisiae*. (c) Cartoon and surface representation of MTR4 KOW (light blue) (PDB 6IEH)[3] highlighting the AIM interacting arginine (red) and putative ligand binding phenylalanine (green).

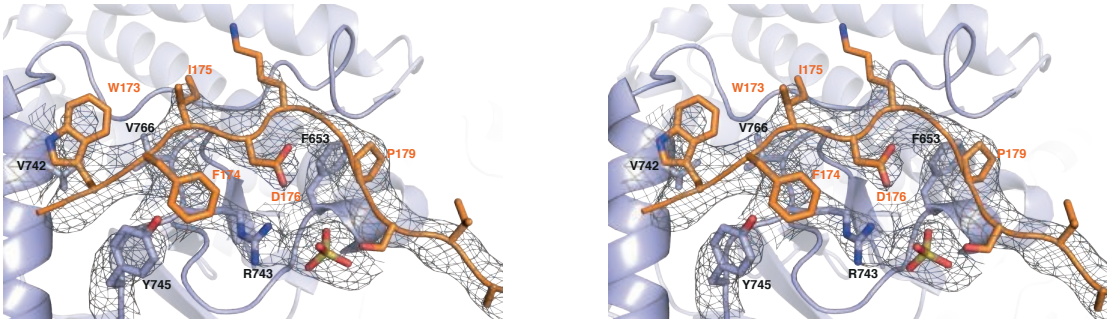
Supplementary Figure 8 : Arch interacting regions of NVL and ZCCHC8 do not influence MTR4 activity.



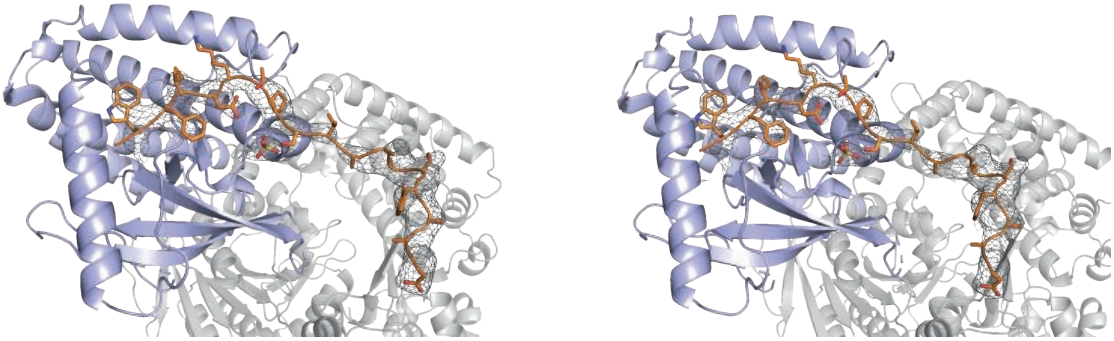
Supplementary Figure 8 (a) End point helicase activity assay of MTR4ΔN, MTR4ΔN with KOW-binding regions of NVL and ZCCHC8, and MTR4 with C-terminal domain of ZCCHC8 on RNA duplex substrates. (b) Time course of ATP hydrolysis by MTR4ΔN alone or in presence of RNA and KOW-binding regions of ZCCHC8 (upper panel) and NVL (lower panel). The data show mean (n=3) with standard deviation plotted as error bars. ZCCHC8 C-terminal domain, which is known to stimulate ATPase activity of MTR4[10], is used as positive control (c) Coomassie stained SDS-PAGE gel showing the proteins used in the helicase and ATPase assays.

Supplementary Figure 9 : Stereo view of electron density at NVL-MTR4 interface

a



b



Supplementary Figure 9 (a) Stereo zoom-in view of the interactions between MTR4 KOW domain (light blue) and NVL (orange) as displayed in Fig. 3d. The model is overlaid with the refined 2mFo-DFc map (grey mesh) showing the density for NVL and the interacting residues of MTR4. The residues of interest are labeled and the map is contoured at 1.0σ. (b) Stereo zoom in view of MTR4-NVL crystal structure showing the ordered region of NVL (orange). The model is overlaid with the 2mFo-DFc omit map calculated in PHENIX omitting the NVL residues. The map is contoured at 1.0σ.

Supplementary Figure 10: Uncropped gels & blots

Supplementary Figure 10 (a) Fig.1b (b) Fig.2d (c) Fig.3b (d) Fig.3e (e) Fig.4b (f) Fig.4d (g) Fig.5b (h) Fig.5c (i) Fig.5d (j) Supplementary Fig.1c (k) Supplementary Fig.3c (l) Supplementary Fig.5b (m) Supplementary Fig.6b (n) Supplementary Fig.6c (o) Supplementary Fig.8a (p) Supplementary Fig.8c.

Supplementary Table 1: List of primers for generating constructs used in this study.

Primer	Sequence (5'-3')
MTRΔN-Fwd	ccaggggcccgactcgatgatttttggaaagaagcccaggatagaagagtc
MTRΔN-Rev	cagaccgccaccgactgcttacaagtagaggctggcagcaaacacaatatctctcttg
MTRΔNΔarch-Fwd	gctttttcagttccagaatgttattagctctggctcgggacaggccgttattcagctggatgacc
MTRΔNΔarch-Rev	ggtcatccagctgaataacggcctgtcccgagccagagctaataacattctggaactgaaaaaagc
MTR4 KOW-Fwd	ccaggggcccgactcgatgcacaaaccaaaatactgcttacctttctac
MTR4 KOW-Rev	cagaccgccaccgactgcttaatcttgaatgcccatatcatcaataggg
NVL(1-266)-Fwd	ccaggggcccgactcgatgaagcccagacctgcagggttcg
NVL(1-266)-Rev	cagaccgccaccgactgcttaatcttcaaacttcacgttggagatctgg
NVL(167-216)-Fwd	ccaggggcccgactcgatgaaagattctgaaggaggatggtttattgac
NVL(167-216)-Rev	cagaccgccaccgactgcttaatcactctccaaaagagaagaatcttttgaatcc
NVL(167-216)-eYFP-cterm fusion-Fwd	tcaaaagattcttctcttttggagagtgatggcagcatggtgagcaagggcgag
NVL(167-216) eYFP cterm fusion-Rev	ctcgcccttgctcaccatgctgccatcactctccaaaagagaagaatcttttga
MTRΔN F677E-Fwd	tgactttggctggggagtagtggtgaatgagtcaaaaaagtcaaatgttaag
MTRΔN F677E-Rev	cttaacatttgacttttttgactcattcaccactactccccagccaaagtca
MTRΔN R743E-Fwd	gtctttaggaatgtaaagctcaacactgctgatagcagacaggag
MTRΔN R743E-Rev	ctcctgtctgctatcagcagtgttgagctttacattcctaaagac
MTRΔN for crystallization-Fwd	ccaggggcccgactcggaaaacctgtatttccagggaacagatgaacccatttttggaaagaagc
NVL(167-216) F186A/D189R-Fwd	gtgtaaagaaagacagtgctttcttgcgcctgtcatgtgagaaaagtaatcc
NVL(167-216) F186A/D189R-Rev	ggattacttttctcacatgacaggcgcaagaaagcactgtctttctttacac
NVL(167-216) W173A-Fwd	ccaggggcccgactcgaaagattctgaaggaggagcgtttattgac
NVL(167-216) I175E-Fwd	ccaggggcccgactcgaaagattctgaaggaggatggtttgaagac
NVL FL-Fwd	taagcagatatcatgaagcccagacctgcag
NVL FL-Rev	tgcttagcggccgcccggctgagggactcct
NVL FL W173A/I175E-Fwd	ctgccaaagattctgaaggaggagcgtttgcggacaaaaccccaagtgtaaag
NVL FL W173A/I175E-Rev	ctttacacttggggttttgtccgcaaacgctcctccttcagaatctttggcag
NVL FL ΔW-AIM-Fwd	ccattcccttgaagacccctgccaaagattctgaaggaagtgtaaagaaagacagttttttcttggacctg
NVL FL ΔW-AIM-Rev	caggtccaagaaaaaactgtctttctttacacttccttcagaatctttggcaggggtcttcaagggaatgg
ZCCHC8(91-211)-Fwd	aagttctgttccaggggcccatggatggacctatattacagattctattcatgaacaatg
ZCCHC8(91-211)-Rev	ccccagaacatcaggttaatggcgttaaacaatgtggctgaagacttgatggtacttg
ZCCHC8(91-211)-eYFP cterm fusion-Fwd	catcaagtcttcagccacattgttggcagcatggtgagcaagggcgag
ZCCHC8(91-211)-eYFP cterm fusion-Rev	ctcgcccttgctcaccatgctgccaacaatgtggctgaagacttgatg
ZCCHC8(91-211) F178A/D181R-Fwd	gaagcggttgccccaatttacgaaggcaagcattagtaaaatacaggacacttcctacaaca
ZCCHC8(91-211) F178A/D181R-Rev	tgttgtaggaagtgtcctgtattttactaatgcttgccttcgtaaattggggcaaccgcttc
ZCCHC8(91-211) W198A/K202E-Rev	gcaaagcaccgtcgttaaacaatgtggctgaagacttgatggtactcgggtatttcagctccttcggaaagctg
ZCCHC8(91-211)/FL I112R/F115R-Fwd	ttcatgaacaatgctatttcaaagcaatatcatcaagaaagagaggaacgtgtatcaaatttagtaaaaagatttgag
ZCCHC8(91-211)/FL I112R/F115R-Rev	ctcaaatctttttactaaatttgatacacgttcctctctttcttgatgatattgctttgaaatagcattgttcatgaa
ZCCHC8 FL-Fwd	taagcagatatcatggccgcagaggtgtattt
ZCCHC8 FL-Rev	tgcttagcggccgcttcagaggccttttttgtttttctg

yMtr4-Fwd	ccaggggcccgactcgatggattctactgatctgttcgatgttttcgagg
yMtr4-Rev	cagaccgccaccgactgcttataaatacaaagaaccagcagatacgatatctctatg
Rix7(1-206)-Fwd	ccagggagcagcctcgatggttaaagtaaagtcgaaaaagaactcatt
Rix7(1-206)-Rev	gcaaagcaccggcctcgttaggatttcagagacgaattaggtggagatc
hNop53(84-123)-Fwd	ccaggggcccgactcgatggaaaaactcttcttcgtggacactg
hNop53(84-123)-Rev	cagaccgccaccgactgcttattctcgaggatgaggtcaacc
NVL D176A-Fwd	ccaggggcccgactcgatgaaagattctgaaggaggatggtttattgccaaaac
ZCCHC8(134-211)-Fwd	aagttctgttccaggggcccatgacttcctttaatcttttgccccagc
ZCCHC8 ΔCTD-Rev	tgcttagcggccgcgctatgaattttagtggccgttg
ZCCHC8 (659-707)-Fwd	ccaggggcccgactcgatgcctatacctgacatgagcaaatttgcaac
ZCCHC8 (659-707)-Rev	cagaccgccaccgactgcttattattcagaggccttttgtttttctgc

*All primers contain overhangs to facilitate ligation independent cloning

Supplementary References

1. Di Tommaso, P. et al. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res* **39**, W13-7 (2011).
2. Oates, M.E. et al. D(2)P(2): database of disordered protein predictions. *Nucleic Acids Res* **41**, D508-16 (2013).
3. Wang, J. et al. NRDE2 negatively regulates exosome functions by inhibiting MTR4 recruitment and exosome interaction. *Genes Dev* (2019).
4. Falk, S. et al. Structural insights into the interaction of the nuclear exosome helicase Mtr4 with the preribosomal protein Nop53. *RNA* **23**, 1780-1787 (2017).
5. Falk, S. et al. The molecular architecture of the TRAMP complex reveals the organization and interplay of its two catalytic activities. *Mol Cell* **55**, 856-867 (2014).
6. Selenko, P. et al. SMN tudor domain structure and its interaction with the Sm proteins. *Nat Struct Biol* **8**, 27-31 (2001).
7. Liu, K. et al. Crystal structure of TDRD3 and methyl-arginine binding characterization of TDRD3, SMN and SPF30. *PLoS One* **7**, e30375 (2012).
8. Tripsianes, K. et al. Structural basis for dimethylarginine recognition by the Tudor domains of human SMN and SPF30 proteins. *Nat Struct Mol Biol* **18**, 1414-20 (2011).
9. Friberg, A., Oddone, A., Klymenko, T., Muller, J. & Sattler, M. Structure of an atypical Tudor domain in the Drosophila Polycomblike protein. *Protein Sci* **19**, 1906-16 (2010).
10. Puno, M.R. & Lima, C.D. Structural basis for MTR4-ZCCHC8 interactions that stimulate the MTR4 helicase in the nuclear exosome-targeting complex. *Proc Natl Acad Sci U S A* **115**, E5506-E5515 (2018).

3.2 CHARACTERIZATION OF THE SMG8-SMG9 INTERACTION

Liang.L., Lingaraju.M., Basquin.C., Basquin.J., Conti.E. (2017). Structure of a SMG8-SMG9 complex identifies a G-domain heterodimer in the NMD effector proteins. RNA 23, 1028-1034.

The study presents the first crystal structures of the *C. elegans* SMG8-SMG9 complex. The findings demonstrate that the complex resembles a G-domain heterodimer and that the nucleotide binding state of the proteins might play a role in regulating the activity of the SMG1 kinase.

The work was performed under the supervision of Prof. Conti E. Lingaraju M. performed the co-immunoprecipitation experiments on human SMG8 and SMG9 validating the dimer interface. Detailed author contributions are included in the article attached.

Structure of a SMG8–SMG9 complex identifies a G-domain heterodimer in the NMD effector proteins

LIANG LI, MAHESH LINGARAJU, CLAIRE BASQUIN, JÉROME BASQUIN, and ELENA CONTI
Department of Structural Cell Biology, Max-Planck-Institute of Biochemistry, D-82152 Martinsried, Germany

ABSTRACT

Nonsense-mediated mRNA decay (NMD) is a eukaryotic mRNA degradation pathway involved in surveillance and post-transcriptional regulation, and executed by the concerted action of several *trans*-acting factors. The SMG1 kinase is an essential NMD factor in metazoans and is associated with two recently identified and yet poorly characterized proteins, SMG8 and SMG9. We determined the 2.5 Å resolution crystal structure of a SMG8–SMG9 core complex from *C. elegans*. We found that SMG8–SMG9 is a G-domain heterodimer with architectural similarities to the dynamin-like family of GTPases such as Atlastin and GBP1. The SMG8–SMG9 heterodimer forms in the absence of nucleotides, with interactions conserved from worms to humans. Nucleotide binding occurs at the G domain of SMG9 but not of SMG8. Fitting the GDP-bound SMG8–SMG9 structure in EM densities of the human SMG1–SMG8–SMG9 complex raises the possibility that the nucleotide site of SMG9 faces SMG1 and could impact the kinase conformation and/or regulation.

Keywords: NMD; post-transcriptional regulation; *C. elegans*; G domain

INTRODUCTION

Nonsense-mediated mRNA decay (NMD) is a eukaryotic surveillance mechanism that degrades aberrant mRNAs containing premature translation termination codons (PTCs) (Popp and Maquat 2013; Lykke-Andersen and Bennett 2014; Karousis et al. 2016). In addition, NMD is a post-transcriptional regulatory mechanism that modulates the expression of physiological mRNAs, affecting the stability of ~10% of the transcriptome (Lykke-Andersen and Jensen 2015). A universal requirement for NMD is a 5′–3′ RNA unwinding activity that is exerted by the helicase UPF1 and regulated by two associated factors, UPF2 and UPF3. In metazoans, UPF1 is additionally regulated by phosphorylation at the N- and C-terminal regions, a decisive event that creates the binding platform for recruiting SMG6 and SMG5–SMG7, which then target the transcript for degradation (Popp and Maquat 2013; Karousis et al. 2016).

UPF1 phosphorylation is catalyzed by the SMG1 kinase (Yamashita et al. 2001). In human cells, SMG1 copurifies in a complex with SMG8 and SMG9 (Yamashita et al. 2009). Human and nematode SMG8 and SMG9 proteins affect the stability of PTC-containing mRNAs in NMD reporter assays (Yamashita et al. 2009). Consistently, inhibition of human SMG-8 has been shown to ameliorate NMD-exacerbated mutant phenotypes (Usuki et al. 2013). However, general impairment of NMD on natural PTC-containing targets was not detected in *smg-8* mutants in *C. elegans* (Rosains and Mango 2012) and in human subjects carrying homozygous loss-of-function *SMG9* mutations (Shaheen et al. 2016). Human patients with *SMG9* deficiency display widespread transcriptional dysregulation, suggesting a predominant role of SMG9 in post-transcriptional regulation rather than in surveillance (Shaheen et al. 2016).

SMG8 and SMG9 interact with each other and inhibit the kinase activity of SMG1 in vitro (Yamashita et al. 2009; Fernández et al. 2010). Electron microscopy studies have revealed the overall architecture of the SMG1–SMG8–SMG9 complex and the central position of SMG8–SMG9 in this trimeric assembly (Arias-Palomo et al. 2011; Melero et al. 2014; Deniaud et al. 2015). However, the limited resolution of the EM maps and the absence of atomic models have so far hampered a molecular understanding of the mechanisms. In this work, we set out to obtain an atomic model of SMG8–SMG9.

RESULTS AND DISCUSSION

Using bioinformatics analyses and proteolysis experiments, we identified regions *C. elegans* (*C.e.*) full-length SMG8 (873 residues) and SMG9 (385 residues) as sufficient to form a stable heterodimeric core complex (SMG8c, residues 1–423 and SMG9c, residues 59–375, Fig. 1A) and to yield

Corresponding author: conti@biochem.mpg.de
Article is online at http://www.rnajournal.org/cgi/doi/10.1261/rna.061200.117. Freely available online through the *RNA* Open Access option.

diffracting crystals. After overcoming crystal lattice defects (detailed in Materials and Methods), we solved the structure and refined it at 2.5 Å resolution with R_{free} of 26.0% (Table 1).

SMG8c and SMG9c contain a similar globular fold with characteristic architecture of G domains along with additional secondary structure elements (Fig. 1B,C). G domains are centered at a mixed β-sheet surrounded by α-helices on the concave and convex surfaces (α1, α5 and α2, α3, α4, respectively) (Wittinghofer and Vetter 2011). The major structural difference between SMG8c and SMG9c is the presence in the former of a helical bundle of three C-terminal helices (α7–α9) that forms a stalk-like protrusion reminiscent of the stalk domain found in GTPases of the dynamin family, such as Atlastin and GBP1 (Fig. 1B; Supplemental Fig. S1; Daumke and Praefcke 2016).

The G domains of SMG8c and SMG9c face each other and interact with part of their convex surfaces (Fig. 1B). In particular, SMG8c helix α2A interacts with SMG9c helices α4 and α3 (patch 1) (in particular Val83^{SMG8}, Ile86^{SMG8} with Leu258^{SMG9}, Leu261^{SMG9}) (Fig. 1D). In addition, the stalk domain of SMG8c folds back on the convex surface of SMG9c (patch 2). Here, SMG8c stalk helices α2B and α7 interact with SMG9c helices α7 and α3 (e.g., Ile335^{SMG8} and Phe338^{SMG8} with Val212^{SMG9} and Tyr358^{SMG9}) (Fig. 1E). Many of the hydrophobic interface residues observed in the *C. elegans* SMG8c–SMG9c structure are conserved in the human orthologs (Supplemental Figs. 2, 3), suggesting a similar overall structure. To test this prediction, we engineered mutations in human full-length SMG9 (hSMG9) by substituting Met390 (corresponding to *C. elegans* Leu258^{SMG9}) and Tyr515 (corresponding to *C. elegans* Tyr358^{SMG9}). We transiently coexpressed full-length HA-tagged hSMG8 and Flag-HA-tagged hSMG9 (wild-type, M390R and M390R, Y515R mutants) in HEK293T cells and carried out coimmunoprecipitation assays with Anti-Flag affinity beads, probing with an anti-HA antibody. We found that the interaction of hSMG8 and hSMG9 observed with the wild-type proteins was indeed strongly impaired by the hSMG9 M390R mutant and almost abolished with the hSMG9 M390R, Y515R double mutant (Fig. 1F).

The relative position of the G-like domains in the SMG8c–SMG9c heterodimer is remarkably similar to that observed in active dimeric GTPases of the dynamin family (Supplemental Fig. S1; Daumke and Praefcke 2016), with the two G domains converging at the loops that are known to harbor the nucleotide-binding motifs (G motifs) in canonical GTPases. However, SMG8 lacks the characteristic residues of G motifs. Another difference is that the single-stalk domain in SMG8c–SMG9c has a different position as compared to the conformations observed in dynamin-like proteins (Supplemental Fig. S1; Byrnes et al. 2013). Finally, the SMG8c–SMG9c heterodimer is formed irrespective of nucleotides, while proteins such as Atlastin or GBP1 dimerize in the presence of GTP analogs (Ghosh et al. 2006; Bian et al. 2011; Byrnes and Sondermann 2011).

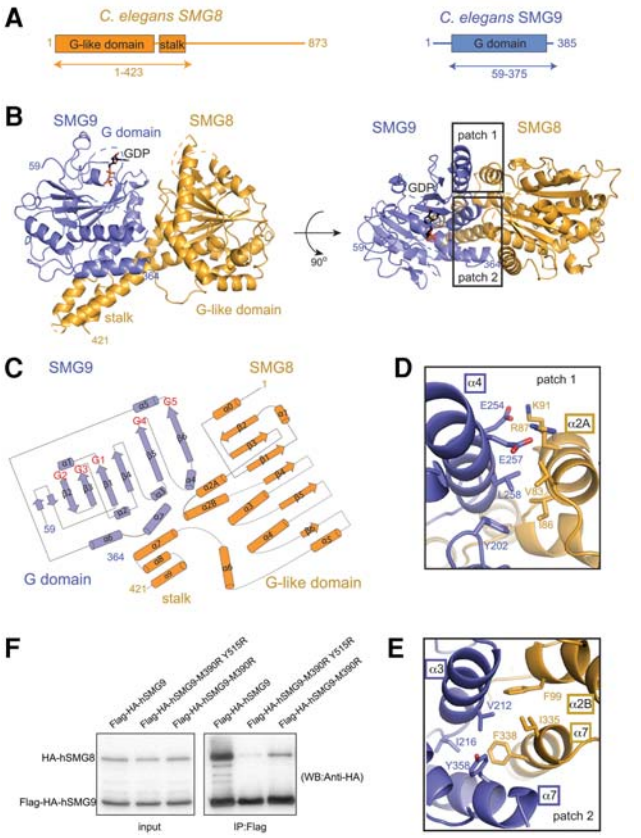


FIGURE 1. Structure of the conserved core of *C. elegans* SMG8–SMG9. (*A*) Schematic representation of the domain organization of *C. elegans* SMG8 (in orange) and SMG9 (in blue). Domains with a structured fold are shown as rectangles and labeled. Predicted low-complexity regions are shown as lines. The arrows *below* the diagram highlight the parts of the proteins that were crystallized. (*B*) Two views of the crystal structure of the *C. elegans* SMG8c–SMG9c core complex, with the molecules shown in orange and blue, respectively. The two views are related by a 90° clockwise rotation around a *horizontal* axis. The G-like domains and the stalk domain are indicated, as well as the N- and C-terminal residues with ordered electron density. The GDP moiety bound to the SMG9 G domain is shown in stick representation. Disordered loops are highlighted with dotted lines. On the *right*, the two rectangles highlight the two main interaction interfaces (patches 1 and 2) that are shown in more detail *below* in panels *D* and *E*. (*C*) Topological diagram of SMG8c and SMG9c (β-strands shown as arrows and α-helices as cylinders). Loops between secondary SMG9 feature similarities in the so-called G motifs as compared to other G domains. The positions of the G motifs in the loops between secondary structure elements are indicated in red. Note that SMG8c and SMG9c feature additional elements as compared to canonical G domains (α2A, α6 and α6, α7, respectively). (*D*) Zoomed-in view of the interacting residues at patch 1. The molecule is shown after ~180° rotation around a *horizontal* axis with respect to the view in panel *A*. SMG8c helix α2A and SMG9c helices α3 and α4 are labeled. (*E*) Zoomed-in view of the interacting residues at patch 2. The molecule is shown in a similar orientation as in panel *C*. SMG8c stalk helices α2B and α7 and SMG9c helices α3 and α7 are indicated. (*F*) Coimmunoprecipitation assays of human full-length HA-tagged hSMG8 and Flag-HA-tagged hSMG9 (wild-type or mutants) in transiently transfected HEK293T cells. Cell lysates (input) were immunoprecipitated with Flag binder and detected with an HA-antibody (precipitate) (12% SDS-PAGE gel). The mutated residues in human SMG9 (M390 and Y515) correspond to *C. elegans* SMG9 Leu258 (patch 1, panel *D*) and Tyr358 (patch 2, panel *E*). HA-SMG8 is 111.7 kDa and Flag-HA-SMG9 is 63.6 kDa.

TABLE 1. Data collection and refinement statistics

	CeSMG8-9-Apo	CeSMG8-9-GDP
Data set		
Wavelength (Å)	0.979	0.9785
Resolution range (Å)[a]	52.36–2.493 (2.583–2.493)	47.48–2.65 (2.734–2.65)
Space group	P 32 2 1	P 32 2 1
a, b, c (Å)	111.085, 111.085, 374.474	110.605, 110.605, 360.266
α, β, γ (°)	90 90 120	90 90 120
Total reflections[a]	3,768,766 (360,260)	764,628 (71,339)
Unique reflections[a]	94,467 (9210)	76,172 (7433)
Multiplicity[a]	39.9 (39.1)	10.0 (9.6)
Completeness (%)[a]	100 (98)	100 (99)
Mean *I*/sigma(*I*)[a]	27.25 (1.61)	12.39 (1.85)
R-merge[a]	0.1374 (2.587)	0.1198 (1.09)
CC1/2[a]	1 (0.783)	0.998 (0.793)
Refinement		
R-work	0.2247	0.2318
R-free	0.2602	0.2740
Average B-factor	89.31	81.5
Ligands		GDP, Mg, EDO
Stereochemistry		
RMS (bonds)	0.003	0.003
RMS (angles)	0.57	0.72
Ramachandran favored (%)	96	96
Ramachandran allowed (%)	3.6	4.1
Ramachandran outliers (%)	0	0

[a]Values in parentheses correspond to the highest-resolution shell.

We tested whether SMG8c–SMG9c can bind guanosine nucleotides. In fluorescence binding assays with mant-nucleotide derivatives, mant-GDP bound SMG8c–SMG9c and SMG9c with a dissociation constant (K_d) of 10 μM and 15 μM, respectively (Fig. 2A). Mant-GTPγS bound SMG8c–SMG9c with a K_d of 6.5 μM, suggesting a slightly tighter binding in the presence of the nucleotide γ-phosphate (Fig. 2A). In general, the low-micromolar binding affinities we measured for SMG8c–SMG9c are similar to those reported for GBP1 (Praefcke et al. 1999). We proceeded to obtain the structure of a nucleotide-bound SMG8c–SMG9c complex. Although the SMG8c–SMG9c crystals cracked when soaking GTP, GDP soaking experiments were successful. Diffraction data to 2.65 Å resolution (Table 1) showed the presence of well-defined electron density for a GDP moiety in SMG9c but not in SMG8c (Supplemental Fig. S4).

GDP binds SMG9c at a similar position as in Atlastin and GBP1, in particular with similarities at the phosphate-binding loops, e.g., at the motifs G1 (P loop), G2 (switch 1), and G3 (switch 2) (Fig. 2B,C). In SMG9c, the P loop residues Lys99^{SMG9} and Ser100^{SMG9} coordinate the phosphates of GDP. Although parts of the switch regions are disordered in our GDP-bound structure, the switch 2 residue Asp150^{SMG9} is at the position expected for coordinating the magnesium ion, while the switch 1 residue Thr135^{SMG9} is 10 Å away from the position expected upon γ-phosphate binding. There are two notable differences in the G1–G3 motifs of SMG9c as compared to the dynamin-like family. First, there is a conserved proline residue (Pro153^{SMG9}, disordered in the present structure) at the position of switch 2 typically occupied by a glycine (Fig. 2B,C). Second, there is a conserved glycine residue (Gly96^{SMG9}) in the P loop at the equivalent position of the so-called arginine "finger" (Arg77^{Atlastin}) (Fig. 2B,C). Consistent with the absence of such arginine (which stimulates the GTPase activity of dynamin-like proteins in *cis*), we did not detect convincing GTPase hydrolysis in vitro (data not shown). Another significant difference is at the G4 and G5 loops that bind the base of the nucleotide in dynamin-like proteins. The characteristic guanosine specificity determinant of Atlastin and GBP1 in the G4 motif is not present in SMG9 (Fig. 2B,C). At the corresponding position of Asp218^{Atlastin}, SMG9 features a conserved lysine residue (Lys241^{SMG9}) that stacks with its aliphatic portion on top of the guanine base. With the caveat that motif G5 is largely disordered, none of the interactions in the current structure engage guanine-specific moieties.

We used our coordinates to progress in the interpretation of cryo-EM structures of human SMG1–SMG8–SMG9 that have been recently resolved at ~20 Å resolution (Fig. 3; Arias-Palomo et al. 2011; Melero et al. 2014; Deniaud et al. 2015). We fitted a homology model of SMG1 with the kinase domain in the "head" region of the density and the N-terminal HEAT-repeat domain in the "arm" region, as in Deniaud et al. (2015). We positioned the *C. elegans* SMG8c–SMG9c structure in the remaining unoccupied density that is connected to the "arm," in a density previously shown to correspond to human SMG8–SMG9 (Arias-Palomo et al. 2011). Although the interpretation of low-resolution maps needs to be judged with caution, placing the atomic coordinates appeared to result in a remarkably good fit, whereby the G domain of SMG9 is at the center of the density, with the G-motif loops pointing toward the HEAT repeat region of SMG1 (Fig. 3). In this pseudo-atomic model, SMG8 has a more peripheral position, with the G-like domain approaching the N-terminal end of the "arm" while the stalk is exposed to solvent.

This pseudo-atomic model is generally in agreement with previous biochemical data (Yamashita et al. 2009; Deniaud et al. 2015). The start of the G domain of *C. elegans* SMG9 (residue 59) is near the density of the SMG1 HEAT repeat "arm." Consistently, the low-complexity N-terminal region

of human SMG9 has been shown to interact with the SMG1 HEAT-repeat domain in co-IP assays (Yamashita et al. 2009) and in crosslinking-mass spectrometry experiments (Deniaud et al. 2015). The end of the folded domain of *C. elegans* SMG8 (residue 421) points toward the SMG1 C-terminal "head." Consistently, the low-complexity C-terminal region of human SMG8 has been shown to contact an insertion domain present in the C-terminal domain of human SMG1 (Deniaud et al. 2015). Finally, the β5–α4 loop of SMG8 faces the density of the SMG1 N-terminal arch. Consistently, the corresponding loop of human SMG8 (residues 290–293) has been shown to contact the SMG1 N terminus in crosslinking-mass spectrometry experiments (Deniaud et al. 2015). Although parts of the SMG9 G motifs as well as the low-complexity regions described above are not present in the current SMG8c–SMG9c crystal structure, the fitting suggests that they might become ordered upon SMG1 binding. In summary, the pseudo-atomic model not only rationalizes how SMG9 recruits the more peripheral SMG8 to the SMG1 complex (Deniaud et al. 2015), but also has predictive value because it raises the hypothesis that the nucleotide-binding state of SMG9 might impact on the entire complex.

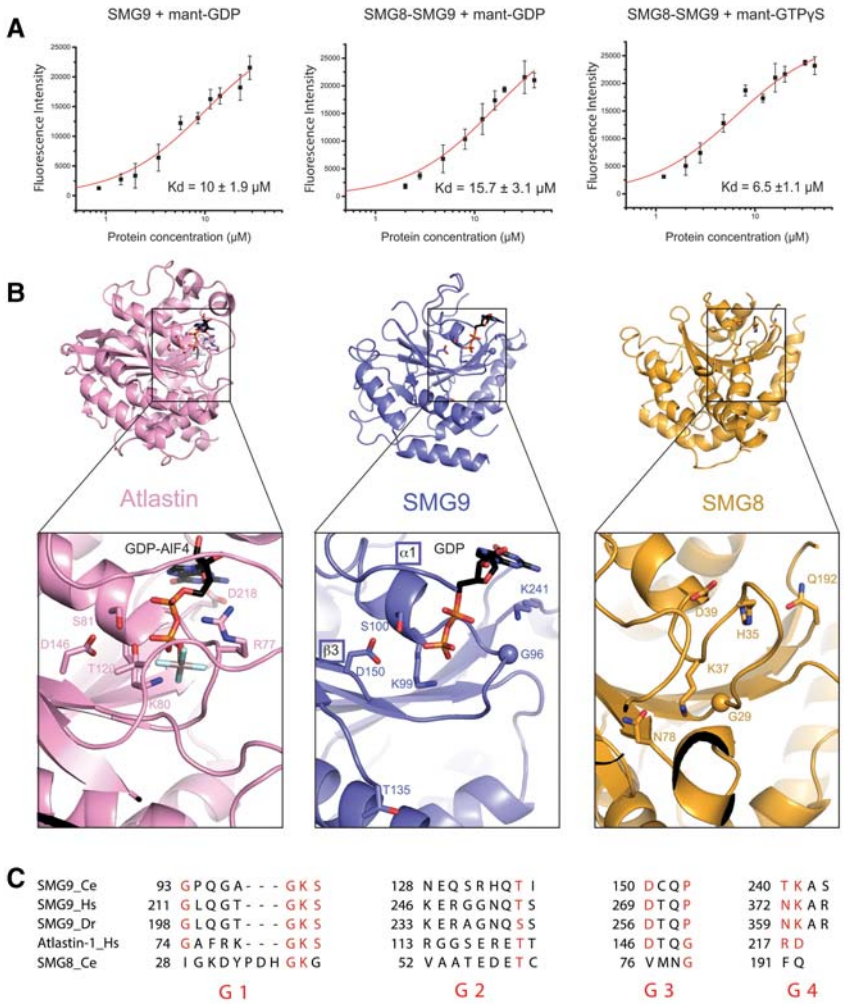


FIGURE 2. The nucleotide-binding site of SMG9. (*A*) Fluorescence measurements of binding affinities of guanosine-nucleotides to SMG8c–SMG9c and SMG9c using mant-labeled GDP and GTP. The data were fitted to a binding equation describing a single-site binding model to obtain the dissociation constants (K_d). The best fit was plotted as a solid line. The K_d values and their corresponding errors are the mean and standard deviation of a minimum of three independent experiments. (*B*) Zoomed-in view at the nucleotide-binding site from the structure of SMG8c–SMG9c bound to GDP. The G domain of SMG9 is shown in the same orientation as in Figure 1B, left panel. The G domain of SMG8 and, as comparison, the G domain of Atlastin (bound to the GDP–AlF4 transition-state analog, ref) are shown in a similar orientation after optimal superposition. The nucleotides and important residues at the nucleotide-binding pockets of SMG9 and Atlastin are shown in ball-and-stick representation. Note that Thr135 in GDP-bound SMG9 (*center* panel) corresponds to Thr120 in GDP–AlF4-bound Atlastin (*left* panel). In SMG8, the equivalent site is incompatible with nucleotide binding: His35 and Gln192 would sterically clash with the ribose and base moieties, respectively, and Asp39 would lead to electrostatic repulsion with the phosphates. (*C*) Alignment of the G1–G4 motif sequences of SMG9 from *C. elegans* (Ce), *H. sapiens* (Hs), and *D. rerio* (Dr), and comparison with human Hs Atlastin and Ce SMG8. The position of the G motifs is schematized in Figure 1B: G1 (or P loop) in the β1–α1 loop, G2 (or switch 1) in α1–β2, G3 (or switch 2) in β3–α2, G4 in β5–α4, and G5 in β6–α5. The G5 motif is disordered in the present structure and divergent in sequence and therefore cannot be compared at present.

MATERIALS AND METHODS

Protein expression and purification

We analyzed the amino acid sequence of SGM8 and SMG9 proteins from different species in an effort to identify orthologs that would be best suited for crystallization. We selected the *C. elegans* (*C.e.*) proteins since they are 10%–25% smaller and therefore likely more compact than their human counterparts. *C.e.* SMG8 (873 residues) and SMG9 (385 residues) were subcloned from a *C.e.* cDNA library with standard PCR protocols in a single MultiBac expression vector (pFL) (Fitzgerald et al. 2006). SMG8 was cloned into the multiple cloning site 1 (MCS1) of the pFL vector using Xma1 and Nhe1, while SMG9 was cloned into the multiple cloning site 2 (MCS2) using BamHI and SalI. Coexpression was crucial to obtain the heterodimer: Although SMG9 could be expressed and purified in a soluble form, SMG8 was insoluble when in isolation (data not shown). Rounds of limited proteolysis and optimization of the expression constructs narrowed down the SMG8c–SMG9c core complex (*C.e.* SMG8 1–423 and SMG9 59–375). SMG8c–SMG9c were coexpressed in baculovirus-infected Hi-Five insect cells (Invitrogen) at 26°C for 70 h. Cells were lysed in 25 mM Tris pH 8.0 with 300 mM NaCl,

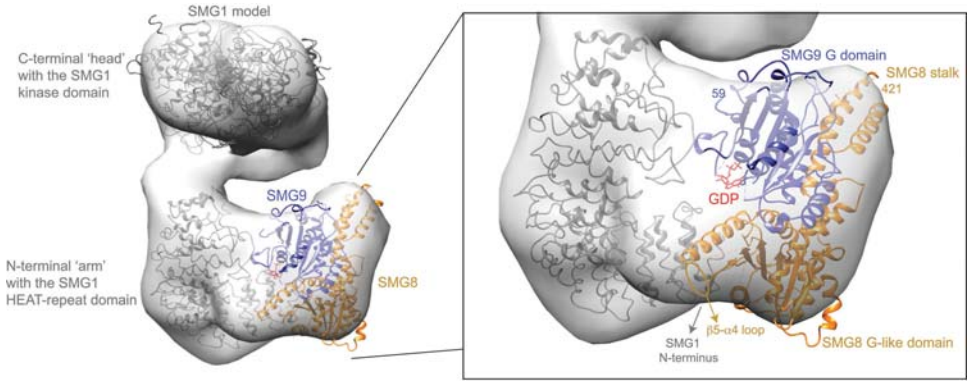


FIGURE 3. Pseudo-atomic model of a SMG1–SMG8–SMG9 complex. In gray is the EM density of a human SMG1–SMG8–SMG9 complex fitted with a model of human SMG1, as in Deniaud et al. (2015). The remaining density was fitted with the crystal structure of the *C. elegans* SMG8–SMG9 core complex (which lacks the low-complexity SMG9 N-terminal and SMG8 C-terminal regions). In red is the GDP molecule bound to SMG9. The fitting was done in Chimera (Pettersen et al. 2004).

and 20 mM imidazole was supplemented to the supernatant before loading onto the nickel column. The complex was purified by nickel-based affinity chromatography via a C-terminal hexahistidine tag on *C.e.* SMG8, and subsequent ion exchange (Heparin HiTrap) and gel-filtration chromatography (Superdex200, equilibrated with 25 mM Tris, 300 mM NaCl, pH 8.0). SelenoMethionine (SeMet) substituted proteins were expressed in insect cells with similar protocols that we reported previously (Halbach et al. 2013). The purification procedure of the SeMet-substituted complex was the same as for the native protein, except that all buffers were degassed and 4 mM β-mercaptoethanol and 2 mM DTT were added before and after elution from the Ni^{2+}–NTA resin, respectively. Mass spectrometry analysis showed the presence of ~60% SeMet incorporation in the purified complex.

Crystallization and structure determination

C.e. SMG8c–CeSMG9c crystallized by vapor diffusion in several PEG conditions at pH 8.0 and 10°C. These initial crystals diffracted to ~3.0 Å resolution and could be processed in a hexagonal space-group, but analysis of the cumulative intensity distribution showed the presence of merohedral twinning with a twin fraction close to 0.5. Additive screening allowed us to identify yttrium chloride as an effective chemical compound to overcome the twinning problem. The best un-twinned crystals were grown by hanging-drop vapor diffusion in drops formed by equal volumes (1.5 µL) of protein (6.8 mg/mL in gel filtration buffer supplemented with 0.11 mM YCl_3) and crystallization buffer (10% PEG3350, 0.1M Tris pH 8.5). SeMet crystals were obtained using the same conditions, but adding tris(2-carboxyethyl)phosphine (TCEP, to limit SeMet oxidation) and covering the reservoir buffer with paraffin oil (to slow drop evaporation and increase crystal size). All crystals were cryoprotected with the crystallization buffer supplemented with 25% ethylene glycol prior to cryo-cooling and data collection.

Diffraction data were collected at 100K at the Swiss Light Source (SLS) beamline PXII. Diffraction data were collected at the selenium K-edge peak wavelength and were processed with XDS (Kabsch 2010). The crystals belong to a trigonal P3221 space group with three copies of the complex in the asymmetric unit related by non-crystallographic symmetry. We used SHELX for phasing (Sheldrick 2010) and phenix.autobuild for initial model building (Adams et al. 2010). We completed the model with iterative rounds of manual building in Coot and refinement with phenix.refine. The three independent copies of the complex in the asymmetric unit are very similar and contain most of the polypeptide chains, except disordered loop regions. The copy of SMG8c–SMG9c described in the text contains SMG8 residues 1–416 (with the exception of disordered loops between residues 193–211, 256–288, and 356–386) and SMG9 residues 59–363 (with the exception of disordered loops between residues 124–134, 152–172, and 284–311) (Table 1).

Native crystals were soaked with 10 mM GDP for 5 min prior to freezing. The structure of *C.e.* SMG8c–SMG9c–GDP was determined by molecular replacement with Phaser using the SeMet-derivatized CeSMG8–9 structure as a search model. The model was completed with Coot (Emsley et al. 2010) and refined with phenix.refine (Adams et al. 2010).

Nucleotide-binding experiments

The affinities for GDP were determined by fluorescence measurements on an Infinite M1000 Pro (Tecan). Experiments were carried out at 21°C in a buffer containing 25 mM Tris pH 8.5, 150 mM NaCl, and 5 mM $MgCl_2$. Increasing protein concentrations were incubated with 1.67 µM of methylanthraniloyl (mant) labeled GDP for 30 min at room temperature. The experiments were carried out with the fragments crystallized, since the full-length proteins were prone to degradation of the low complexity sequences. Fluorescence of mant-GDP was excited at 355 nm and emission spectra were then monitored from 400 to 500 nm, with emission maxima detected at 448 nm. The intrinsic protein fluorescence as well as the mant-nucleotide background was subtracted from the curves. Curve fittings were done with Origin with a one-to-one binding model and are consistent with the presence of one molecule of nucleotide per heterodimer. Curves were done in triplicate. Similar approaches were used to determine the binding affinities for GTPγS.

Coimmunoprecipitation assays

Both the SMG8 and SMG9 were cloned in a vector containing the EF-1 α promoter and with an N-terminal Flag tag and N-terminal

HA tag using EcoRI and NotI restriction sites. HEK293T cells were cultured in Dulbecco's modified Eagle medium containing 10% fetal bovine serum (Gibco), 100 U/mL penicillin, and 0.1 mg/mL streptomycin (Gibco) at 32°C/5% CO_2. Plasmids were transfected with polyethyleneimine (Polysciences Inc., 1 mg/mL) for protein interaction studies. HEK293T cells were collected from confluent six-well plates after 72 h of transient transfection. Cells were lysed in 0.5 mL of lysis buffer containing 50 mM Tris, pH 7.4, 150 mM NaCl, supplemented with protease inhibitor cocktail (Roche) and DNase I. The lysate was centrifuged at 16,000*g* for 30 min at 4°C. Twelve microliters of Anti-Flag M2 sepharose beads (Sigma) were added to supernatant for 1 h at 4°C. Beads were washed four times with 1 mL of buffer containing 50 mM Tris, pH 7.4, 300 mM NaCl, and proteins were eluted with 25 µL of lysis buffer supplemented with 100 µg/mL flag peptide (Sigma-Aldrich, F3290). Eluted proteins were run on 12% polyacrylamide gels and transferred onto polyvinylidene difluoride membrane (0.45 µm pore size) (Millipore Immobilon-P) for Western blotting. Anti-HA (Covance, MMS-101 R) antibody and horseradish peroxidase–coupled goat anti-mouse (Millipore, AQ502A) secondary antibody were used in combination with ECL prime Western blotting detection reagent (GE healthcare) for detection of Flag-HA and HA-tagged proteins via Western blotting.

DATA DEPOSITION

The coordinates have been deposited in the Protein Data Bank with accession codes 5NKM (SMG8-SMG9) and 5NKK (SMG8-SMG9-GDP).

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We thank the staff of the MPIB Crystallization Facility for robotic crystal screening and optimization setups and of the Core Facility of MPIB for mass spectrometry analysis. We thank Christiane Schaffitzel for the coordinates of the human SMG1 model (Deniaud et al. 2015). We also thank the staff of the PX beamlines of the SLS synchrotron for assistance in data collection. This study was supported by the Max Planck Gesellschaft, the European Commission (ERC Advanced Investigator Grant 294371 and Marie Curie ITN RNPnet), and the Deutsche Forschungsgemeinschaft (DFG SFB646, SFB1035, GRK1721, FOR1680, and CIPSM) to E.C.

Author contributions: L.L. purified, crystallized, and solved the structure together with J.B.; M.L. carried out the co-IP assays; C.B. carried out biophysical and biochemical assays; E.C. and J.B. supervised the project; E.C. and L.L. wrote the manuscript.

Received February 22, 2017; accepted April 4, 2017.

REFERENCES

Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, et al. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66:** 213–221.

Arias-Palomo E, Yamashita A, Fernández IS, Núñez-Ramírez R, Bamba Y, Izumi N, Ohno S, Llorca O. 2011. The nonsense-mediated mRNA decay SMG-1 kinase is regulated by large-scale conformational changes controlled by SMG-8. *Genes Dev* **25:** 153–164.

Bian X, Klemm RW, Liu TY, Zhang M, Sun S, Sui X, Liu X, Rapoport TA, Hu J. 2011. Structures of the atlastin GTPase provide insight into homotypic fusion of endoplasmic reticulum membranes. *Proc Natl Acad Sci* **108:** 3976–3981.

Byrnes LJ, Sondermann H. 2011. Structural basis for the nucleotide-dependent dimerization of the large G protein atlastin-1/SPG3A. *Proc Natl Acad Sci* **108:** 2216–2221.

Byrnes LJ, Singh A, Szeto K, Benvin NM, O'Donnell JP, Zipfel WR, Sondermann H. 2013. Structural basis for conformational switching and GTP loading of the large G protein atlastin. *EMBO J* **32:** 369–384.

Daumke O, Praefcke GJK. 2016. Invited review: mechanisms of GTP hydrolysis and conformational transitions in the dynamin superfamily. *Biopolymers* **105:** 580–593.

Deniaud A, Karuppasamy M, Bock T, Masiulis S, Huard K, Garzoni F, Kerschgens K, Hentze MW, Kulozik AE, Beck M, et al. 2015. A network of SMG-8, SMG-9 and SMG-1 C-terminal insertion domain regulates UPF1 substrate recruitment and phosphorylation. *Nucleic Acids Res* **43:** 7600–7611.

Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66:** 486–501.

Fernández IS, Yamashita A, Arias-Palomo E, Bamba Y, Bartolomé RA, Canales MA, Teixidó J, Ohno S, Llorca O. 2010. Characterization of SMG-9, an essential component of the nonsense-mediated mRNA decay SMG1C complex. *Nucleic Acids Res* **39:** 347–358.

Fitzgerald DJ, Berger P, Schaffitzel C, Yamada K, Richmond TJ, Berger I. 2006. Protein complex expression by using multigene baculoviral vectors. *Nat Methods* **3:** 1021–1032.

Ghosh A, Praefcke GJK, Renault L, Wittinghofer A, Herrmann C. 2006. How guanylate-binding proteins achieve assembly-stimulated processive cleavage of GTP to GMP. *Nature* **440:** 101–104.

Halbach F, Reichelt P, Rode M, Conti E. 2013. The yeast ski complex: crystal structure and RNA channeling to the exosome complex. *Cell* **154:** 814–826.

Kabsch W. 2010. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr* **66:** 133–144.

Karousis ED, Nasif S, Mühlemann O. 2016. Nonsense-mediated mRNA decay: novel mechanistic insights and biological impact. *WIREs RNA* **7:** 661–682.

Lykke-Andersen J, Bennett EJ. 2014. Protecting the proteome: eukaryotic cotranslational quality control pathways. *J Cell Biol* **204:** 467–476.

Lykke-Andersen S, Jensen TH. 2015. Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nat Rev Mol Cell Biol* **16:** 665–677.

Melero R, Uchiyama A, Castaño R, Kataoka N, Kurosawa H, Ohno S, Yamashita A, Llorca O. 2014. Structures of SMG1-UPFs complexes: SMG1 contributes to regulate UPF2-dependent activation of UPF1 in NMD. *Structure* **22:** 1105–1119.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* **25:** 1605–1612.

Popp MW-L, Maquat LE. 2013. Organizing principles of mammalian nonsense-mediated mRNA decay. *Annu Rev Genet* **47:** 139–165.

Praefcke GJ, Geyer M, Schwemmle M, Robert Kalbitzer H, Herrmann C. 1999. Nucleotide-binding characteristics of human guanylate-binding protein 1 (hGBP1) and identification of the third GTP-binding motif. *J Mol Biol* **292:** 321–332.

Rosains J, Mango SE. 2012. Genetic characterization of smg-8 mutants reveals no role in *C. elegans* nonsense mediated decay. *PLoS ONE* **7:** e49490.

Shaheen R, Anazi S, Ben-Omran T, Seidahmed MZ, Caddle LB, Palmer K, Ali R, Alshidi T, Hagos S, Goodwin L, et al. 2016. Mutations in SMG9, encoding an essential component of nonsense-mediated decay machinery, cause a multiple congenital anomaly syndrome in humans and mice. *Am J Hum Genet* **98:** 643–652.

Sheldrick GM. 2010. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* **66:** 479–485.

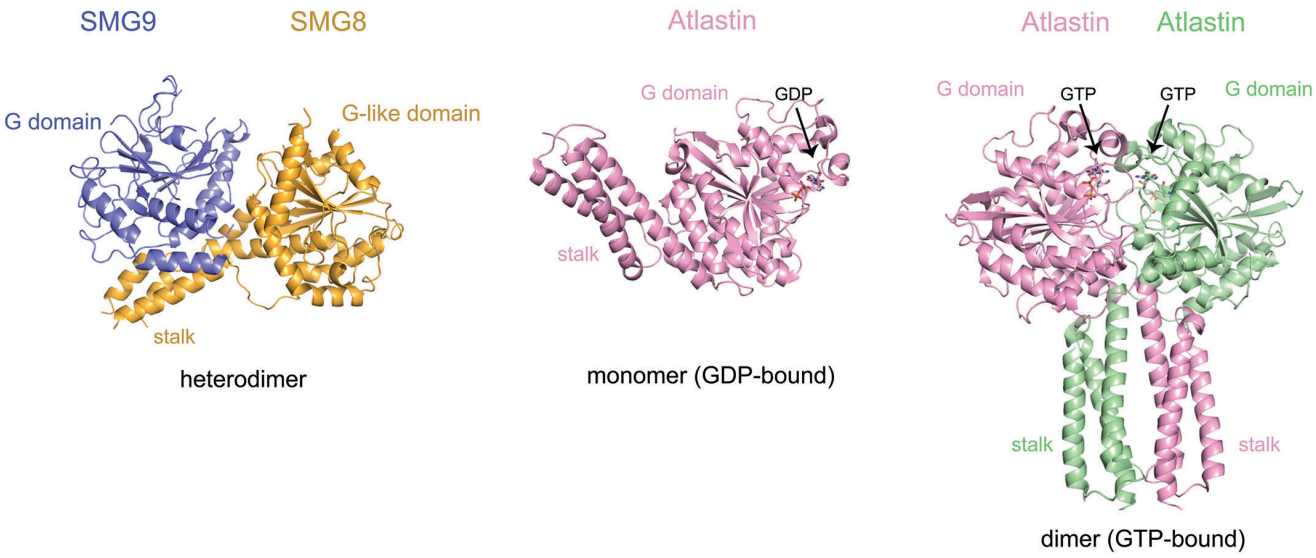
Usuki F, Yamashita A, Shiraishi T, Shiga A, Onodera O, Higuchi I, Ohno S. 2013. Inhibition of SMG-8, a subunit of SMG-1 kinase, ameliorates nonsense-mediated mRNA decay-exacerbated mutant phenotypes without cytotoxicity. *Proc Natl Acad Sci* **110:** 15037–15042.

Wittinghofer A, Vetter IR. 2011. Structure-function relationships of the G domain, a canonical switch motif. *Biochemistry* **80:** 943–971.

Yamashita A, Ohnishi T, Kashima I, Taya Y, Ohno S. 2001. Human SMG-1, a novel phosphatidylinositol 3-kinase-related protein kinase, associates with components of the mRNA surveillance complex and is involved in the regulation of nonsense-mediated mRNA decay. *Genes Dev* **15:** 2215–2228.

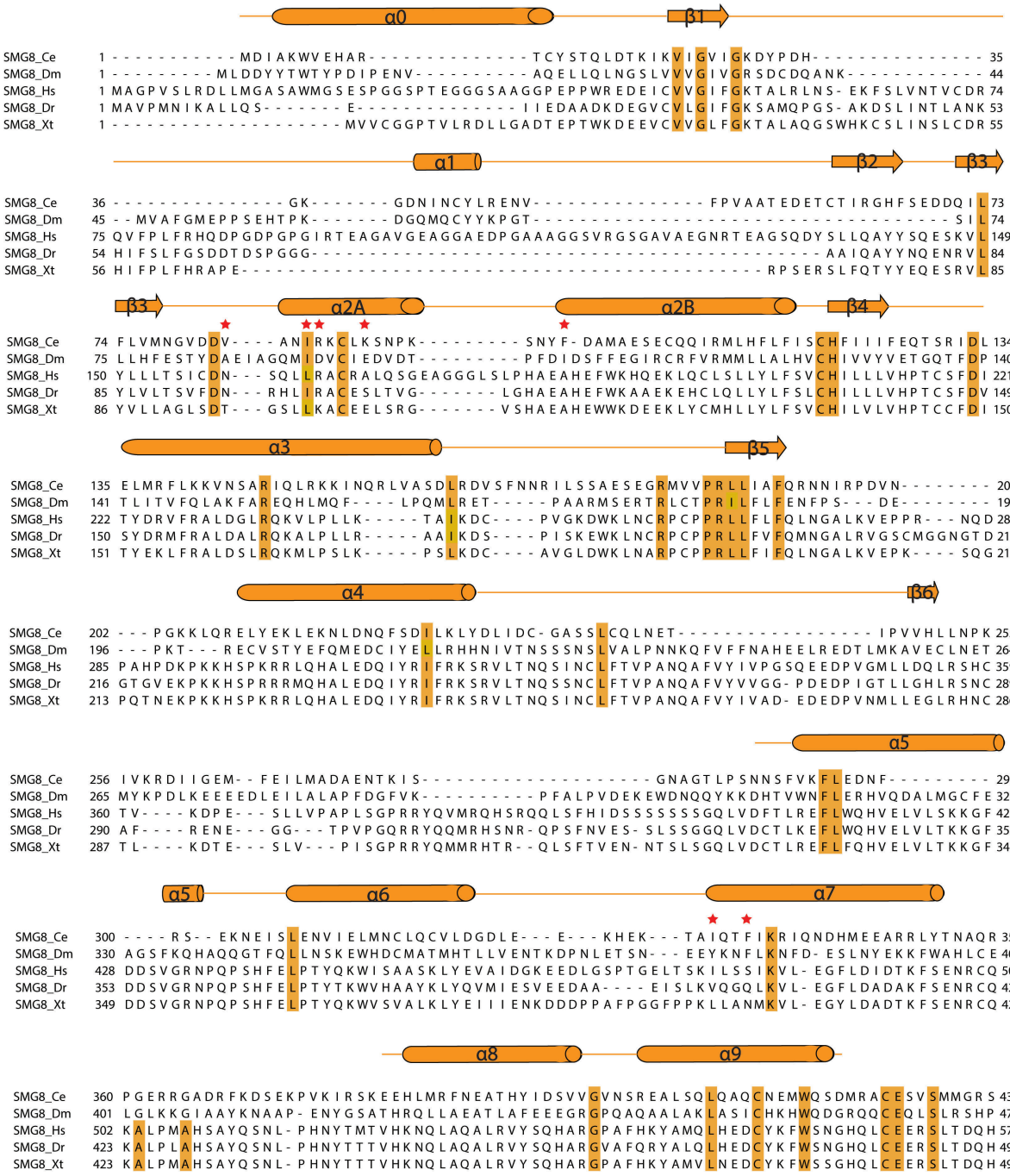
Yamashita A, Izumi N, Kashima I, Ohnishi T, Saari B, Katsuhata Y, Muramatsu R, Morita T, Iwamatsu A, Hachiya T, et al. 2009. SMG-8 and SMG-9, two novel subunits of the SMG-1 complex, regulate remodeling of the mRNA surveillance complex during nonsense-mediated mRNA decay. *Genes Dev* **23:** 1091–1105.

Supplementary figure 1: Comparison of the structures of SMG8-SMG9 and the dynamin-like GTPase Atlastin

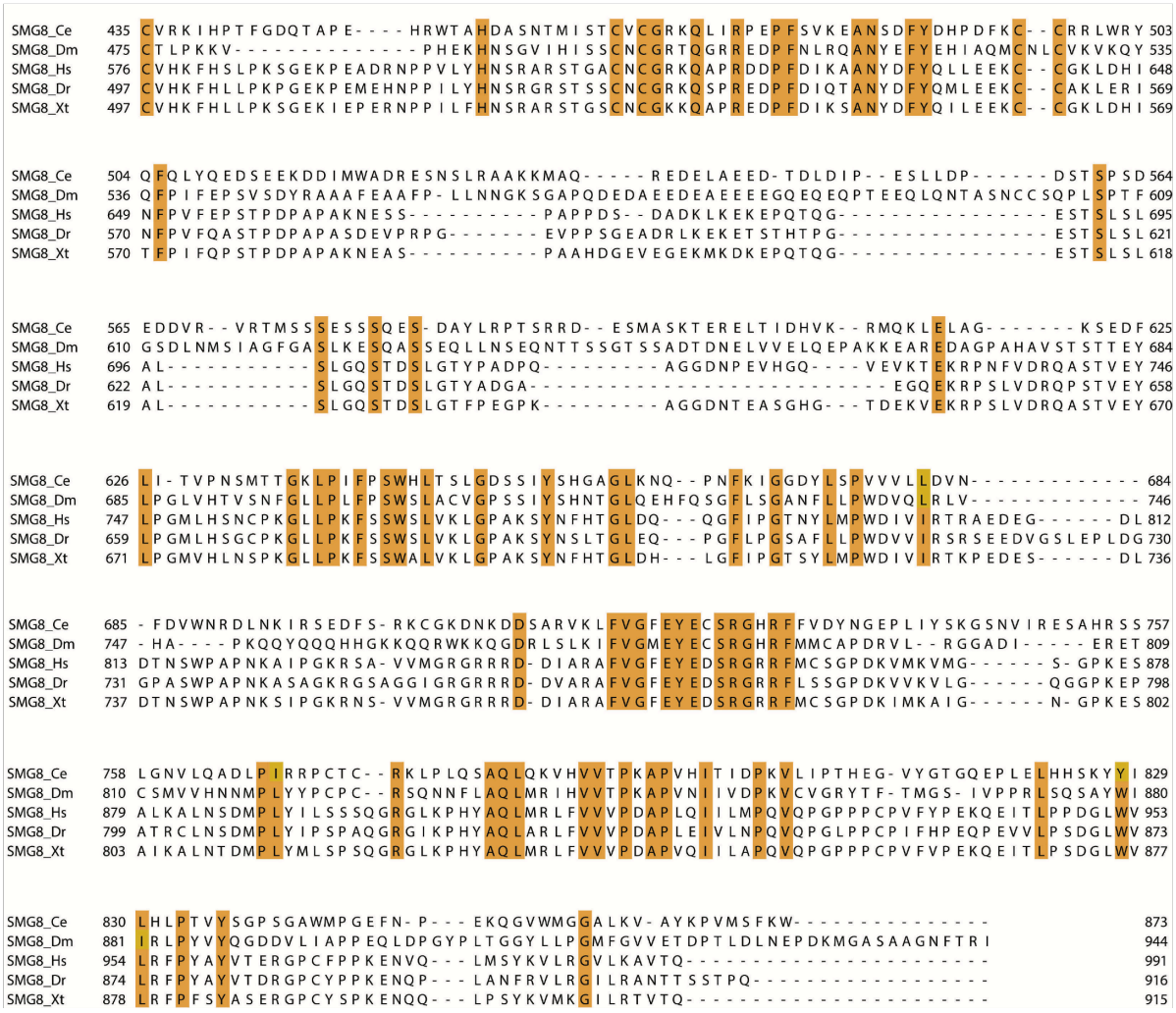


The C.e. SMG8-SMG9 heterodimer is shown on the left, as in Figure 1A. The monomeric form of Atlastin (GDP-bound) is shown in the central panel, with the G domain in the same orientation as the G domain of SMG9 as shown in the left panel. The homo-dimeric form of Atlastin (GTP bound) is shown on the right panel, with protomer in pink shown in the same orientation as in the central panel, after optimal superposition of their G domains. The nucleotides are shown in ball-and-stick representation. Note the different position of the stalk domains.

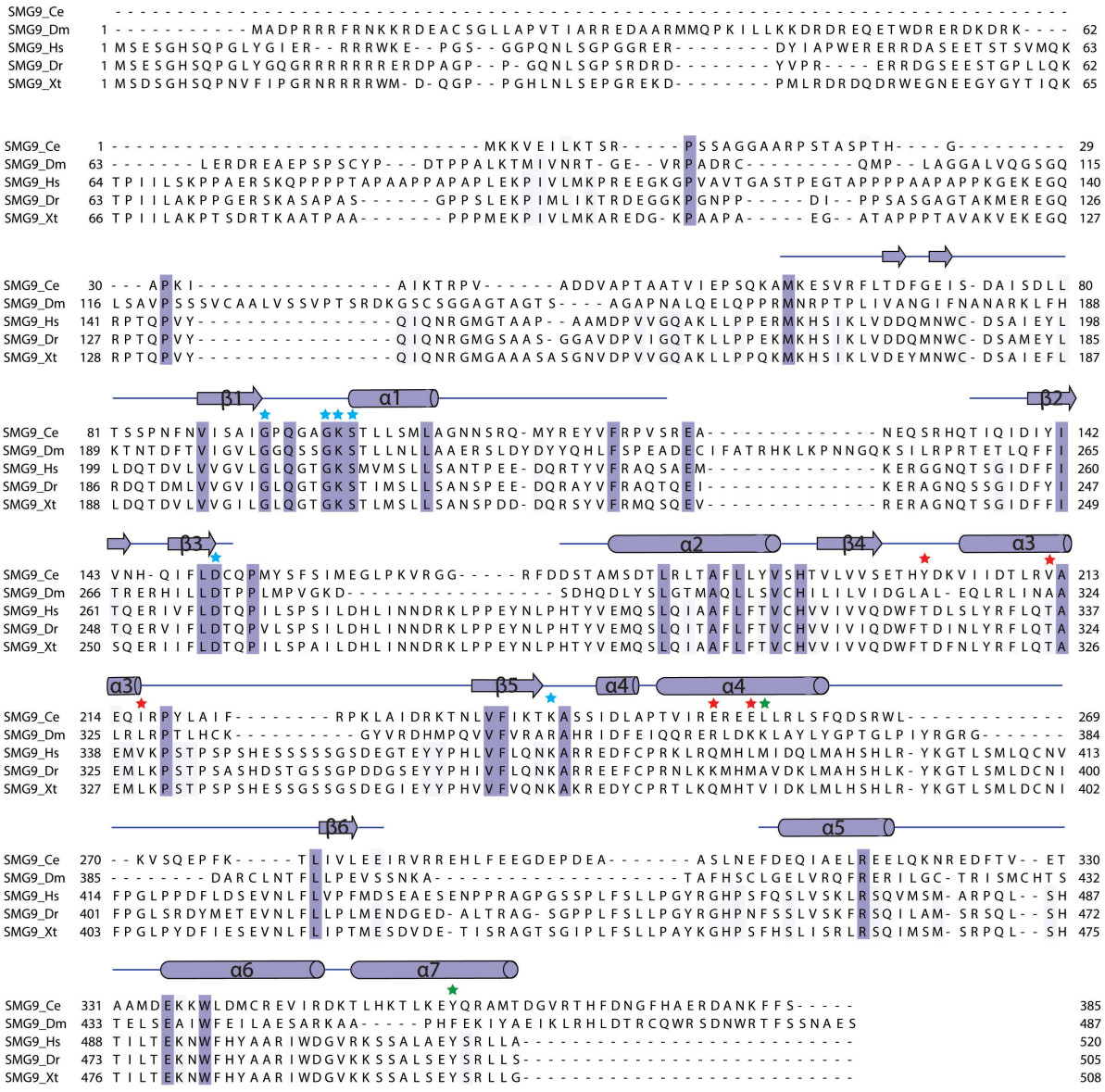
Supplementary figure 2: Evolutionary conservation of metazoan SMG8



```
SMG8_Ce    1 -----------MDIAKWVEHAR--------------TCYSTQLDTKIKVIGVIGKDYPDH--------------- 35
SMG8_Dm    1 ---------MLDDYYTWTYPDIPENV----------AQELLQLNGSLVVVGIVGRSDCDQANK------------ 44
SMG8_Hs    1 MAGPVSLRDLLMGASAWMGSESPGGSPTEGGGSAAGGPEPPWREDEICVVGIFGKTALRLNS-EKFSLVNTVCDR 74
SMG8_Dr    1 MAVPMNIKALLQS-------E--------------IIEDAADKDEGVCVLGIFGKSAMQPGS-AKDSLINTLANK 53
SMG8_Xt    1 --------------------MVVCGGPTVLRDLLGADTEPTWKDEEVCVVGLFGKTALAQGSWHKCSLINSLCDR 55

SMG8_Ce   36 ---------------GK-------GDNINCYLRENV---------------FPVAATEDETCTIRGHFSEDDQIL 73
SMG8_Dm   45 --MVAFGMEPPSEHTPK-------DGQMQCYYKPGT------------------------------------SIL 74
SMG8_Hs   75 QVFPLFRHQDPGDPGPGIRTEAGAVGEAGGAEDPGAAAGGSVRGSGAVAEGNRTEAGSQDYSLLQAYYSQESKVL 149
SMG8_Dr   54 HIFSLFGSDDTDSPGGG--------------------------------------------AAIQAYYNQENRVL 84
SMG8_Xt   56 HIFPLFHRAPE---------------------------------------------RPSERSLFQTYYEQESRVL 85

SMG8_Ce   74 FLVMNGVDDV----ANIRKCLKSNPK---------SNYF-DAMAESECQQIRMLHFLFISCHFIIIFEQTSRIDL 134
SMG8_Dm   75 LLHFESTYDAEIAGQMIDVCIEDVDT---------PFDIDSFFEGIRCRFVRMMLLALHVCHIVVYVETGQTFDP 140
SMG8_Hs  150 YLLLTSICDN---SQLLRACRALQSGEAGGGLSLPHAEAHEFWKHQEKLQCLSLLYLFSVCHILLLVHPTCSFDI 221
SMG8_Dr   85 YLVLTSVFDN---RHLIRACESLTVG-------LGHAEAHEFWKAAEKEHCLQLLYLFSLCHILLLVHPTCSFDV 149
SMG8_Xt   86 YVLLAGLSDT---GSLLKACEELSRG-------VSHAEAHEWWKDEEKLYCMHLLYLFSVCHILVLVHPTCCFDI 150

SMG8_Ce  135 ELMRFLKKVNSARIQLRKKINQRLVASDLRDVSFNNRILSSAESEGRMVVPRLLIAFQRNNIRPDVN-------- 201
SMG8_Dm  141 TLITVFQLAKFAREQHLMQF----LPQMLRET-----PAARMSERTRLCTPRILFLFENFPS---DE-------- 195
SMG8_Hs  222 TYDRVFRALDGLRQKVLPLLK-----TAIKDC----PVGKDWKLNCRPCPPRLLFLFQLNGALKVEPPR---NQD 284
SMG8_Dr  150 SYDRMFRALDALRQKALPLLR-----AAIKDS----PISKEWKLNCRPCPPRLLFVFQMNGALRVGSCMGGNGTD 215
SMG8_Xt  151 TYEKLFRALDSLRQKMLPSLK-----PSLKDC----AVGLDWKLNARPCPPRLLFIFQLNGALKVEPK----SQG 212

SMG8_Ce  202 ---PGKKLQRELYEKLEKNLDNQFSDILKLYDLIDC-GASSLCQLNET-----------------IPVVHLLNPK 255
SMG8_Dm  196 ---PKT---RECVSTYEFQMEDCIYELLRHHNIVTNSSSNSLVALPNNKQFVFFNAHEELREDTLMKAVECLNET 264
SMG8_Hs  285 PAHPDKPKKHSPKRRLQHALEDQIYRIFRKSRVLTNQSINCLFTVPANQAFVYIVPGSQEEDPVGMLLDQLRSHC 359
SMG8_Dr  216 GTGVEKPKKHSPRRRMQHALEDQIYRIFRKSRVLTNQSSNCLFTVPANQAFVYVVGG-PDEDPIGTLLGHLRSNC 289
SMG8_Xt  213 PQTNEKPKKHSPKRRLQHALEDQIYRIFRKSRVLTNQSINCLFTVPANQAFVYIVAD-EDEDPVNMLLEGLRHNC 286

SMG8_Ce  256 IVKRDIIGEM--FEILMADAENTKIS--------------------GNAGTLPSNNSFVKFLEDNF--------- 299
SMG8_Dm  265 MYKPDLKEEEEDLEILALAPFDGFVK----------PFALPVDEKEWDNQQYKKDHTVWNFLERHVQDALMGCFE 329
SMG8_Hs  360 TV----KDPE---SLLVPAPLSGPRRYQVMRQHSRQQLSFHIDSSSSSSSGQLVDFTLREFLWQHVELVLSKKGF 427
SMG8_Dr  290 AF----RENE---GG---TPVPGQRRYQQMRHSNR-QPSFNVES-SLSSGGQLVDCTLKEFLWQHVELVLTKKGF 352
SMG8_Xt  287 TL----KDTE---SLV---PISGPRRYQMMRHTR--QLSFTVEN-NTSLSGQLVDCTLREFLFQHVELVLTKKGF 348

SMG8_Ce  300 ----RS--EKNEISLENVIELMNCLQCVLDGDLE---E---KHEK---TAIQTFIKRIQNDHMEEARRLYTNAQR 359
SMG8_Dm  330 AGSFKQHAQQGTFQLLNSKEWHDCMATMHTLLVENTKDPNLETSN---EEYKNFLKNFD-ESLNYEKKFWAHLCE 400
SMG8_Hs  428 DDSVGRNPQPSHFELPTYQKWISAASKLYEVAIDGKEEDLGSPTGELTSKILSSIKVL-EGFLDIDTKFSENRCQ 501
SMG8_Dr  353 DDSVGRNPQPSHFELPTYTKWVHAAYKLYQVMIESVEEDAA----EISLKVQGQLKVL-EGFLDADAKFSENRCQ 422
SMG8_Xt  349 DDSVGRNPQPSHFELPTYQKWVSVALKLYEIIIENKDDDPPAFPGGFPPKLLANMKVL-EGYLDADTKFSENRCQ 422

SMG8_Ce  360 PGERRGADRFKDSEKPVKIRSKEEHLMRFNEATHYIDSVVGVNSREALSQLQAQCNEMWQSDMRACESVSMMGRS 434
SMG8_Dm  401 LGLKKGIAAYKNAAP-ENYGSATHRQLLAEATLAFEEEGRGPQAQAALAKLASICHKHWQDGRQQCEQLSLRSHP 474
SMG8_Hs  502 KALPMAHSAYQSNL-PHNYTMTVHKNQLAQALRVYSQHARGPAFHKYAMQLHEDCYKFWSNGHQLCEERSLTDQH 575
SMG8_Dr  423 KALPLAHSAYQSNL-PHNYTTTVHKNQLAQALRVYSQHARGVAFQRYALQLHEDCYKFWSNGHQLCEERSLTDQH 496
SMG8_Xt  423 KALPMAHSAYQSNL-PHNYTTTVHKNQLAQALRVYSQHARGPAFHKYAMVLNEDCYKFWSSGHQLCEERSLTDQH 496
```


SMG8_Ce 435 CVRKIHPTFGDQTAPE----HRWTAHDASNTMISTCVCGRKQLIRPEPFSVKEANSDFYDHPDFKC--CRRLWRY 503
SMG8_Dm 475 CTLPKKV---------------PHEKHNSGVIHISSCNCGRTQGRREDPFNLRQANYEFYEHIAQMCNLCVKVKQY 535
SMG8_Hs 576 CVHKFHSLPKSGEKPEADRNPPVLYHNSRARSTGACNCGRKQAPRDDPFDIKAANYDFYQLLEEKC--CGKLDHI 648
SMG8_Dr 497 CVHKFHLLPKPGEKPEMEHNPPILYHNSRGRSTSSCNCGRKQSPREDPFDIQTANYDFYQMLEEKC--CAKLERI 569
SMG8_Xt 497 CVHKFHLLPKSGEKIEPERNPPILFHNSRARSTGSCNCGKKQAPREDPFDIKSANYDFYQILEEKC--CGKLDHI 569
SMG8_Ce 504 QFQLYQEDSEEKDDIMWADRESNSLRAAKKMAQ-----REDELAEED-TDLDIP--ESLLDP------DSTSPSD 564
SMG8_Dm 536 QFPIFEPSVSDYRAAAFEAAFP-LLNNGKSGAPQDEDAEEDEAEEEEGQEQEQPTEEQLQNTASNCCSQPLSPTF 609
SMG8_Hs 649 NFPVFEPSTPDPAPAKNESS----------PAPPDS--DADKLKEKEPQTQG----------------ESTSLSL 695
SMG8_Dr 570 NFPVFQASTPDPAPASDEVPRPG-------EVPPSGEADRLKEKETSTHTPG----------------ESTSLSL 621
SMG8_Xt 570 TFPIFQPSTPDPAPAKNEAS----------PAAHDGEVEGEKMKDKEPQTQG----------------ESTSLSL 618
SMG8_Ce 565 EDDVR--VRTMSSSESSSQES-DAYLRPTSRRD--ESMASKTERELTIDHVK--RMQKLELAG-------KSEDF 625
SMG8_Dm 610 GSDLNMSIAGFGASLKESQASSEQLLNSEQNTTSSGTSSADTDNELVVELQEPAKKEAREDAGPAHAVSTSTTEY 684
SMG8_Hs 696 AL----------SLGQSTDSLGTYPADPQ---------AGGDNPEVHGQ----VEVKTEKRPNFVDRQASTVEY 746
SMG8_Dr 622 AL----------SLGQSTDSLGTYADGA----------EGQEKRPSLVDRQPSTVEY 658
SMG8_Xt 619 AL----------SLGQSTDSLGTFPEGPK---------AGGDNTEASGHG---TDEKVEKRPSLVDRQASTVEY 670
SMG8_Ce 626 LI-TVPNSMTTGKLPIFPSWHLTSLGDSSIYSHGAGLKNQ--PNFKIGGDYLSPVVVLLDVN----------684
SMG8_Dm 685 LPGLVHTVSNFGLLPLFPSWSLACVGPSSIYSHNTGLQEHFQSGFLSGANFLLPWDVQLRLV----------746
SMG8_Hs 747 LPGMLHSNCPKGLLPKFSSWSLVKLGPAKSYNFHTGLDQ---QGFIPGTNYLMPWDIVIRTRAEDEG------DL 812
SMG8_Dr 659 LPGMLHSGCPKGLLPKFSSWSLVKLGPAKSYNSLTGLEQ---PGFLPGSAFLLPWDVVIRSRSEEDVGSLEPLDG 730
SMG8_Xt 671 LPGMVHLNSPKGLLPKFSSWALVKLGPAKSYNFHTGLDH---LGFIPGTSYLMPWDIVIRTKPEDES------DL 736
SMG8_Ce 685 -FDVWNRDLNKIRSEDFS-RKCGKDNKDDSARVKLFVGFEYECSRGHRFFVDYNGEPLIYSKGSNVIRESAHRSS 757
SMG8_Dm 747 -HA----PKQQYQQQHHGKKQQRWKKQGDRLSLKIFVGMEYECSRGHRFMMCAPDRVL--RGGADI-----ERET 809
SMG8_Hs 813 DTNSWPAPNKAIPGKRSA-VVMGRGRRRD-DIARAFVGFEYEDSRGRRFMCSGPDKVMKVMG------S-GPKES 878
SMG8_Dr 731 GPASWPAPNKASAGKRGSAGGIGRGRRRD-DVARAFVGFEYEDSRGRRFLSSGPDKVVKVLG------QGGPKEP 798
SMG8_Xt 737 DTNSWPAPNKSIPGKRNS-VVMGRGRRRD-DIARAFVGFEYEDSRGRRFMCSGPDKIMKAIG------N-GPKES 802
SMG8_Ce 758 LGNVLQADLPIRRPCTC--RKLPLQSAQLQKVHVVTPKAPVHITIDPKVLIPTHEG-VYGTGQEPLELHHSKYYI 829
SMG8_Dm 810 CSMVVHNNMPLYYPCPC--RSQNNFLAQLMRIHVVTPKAPVNIIVDPKVCVGRYTF-TMGS-IVPPRLSQSAYWI 880
SMG8_Hs 879 ALKALNSDMPLYILSSSQGRGLKPHYAQLMRLFVVVPDAPLQIILMPQVQPGPPPCPVFYPEKQEITLPPDGLWV 953
SMG8_Dr 799 ATRCLNSDMPLYIPSPAQGRGIKPHYAQLARLFVVVPDAPLEIVLNPQVQPGLPPCPIFHPEQPEVVLPSDGLWV 873
SMG8_Xt 803 AIKALNTDMPLYMLSPSQGRGLKPHYAQLMRLFVVVPDAPVQIILAPQVQPGPPPCPVFVPEKQEITLPSDGLWV 877
SMG8_Ce 830 LHLPTVYSGPSGAWMPGEFN-P---EKQGVWMGGALKV-AYKPVMSFKW---------------873
SMG8_Dm 881 IRLPYVYQGDDVLIAPPEQLDPGYPLTGGYLLPGMFGVVETDPTLDLNEPDKMGASAAGNFTRI 944
SMG8_Hs 954 LRFPYAYVTERGPCFPPKENVQ---LMSYKVLRGVLKAVTQ----------------------991
SMG8_Dr 874 LRFPYAYVTDRGPCYPPKENQP---LANFRVLRGILRANTTSSTPQ----------------916
SMG8_Xt 878 LRFPFSYASERGPCYSPKENQQ---LPSYKVMKGILRTVTQ----------------------915

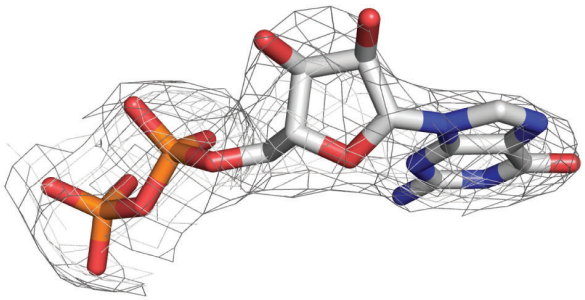
Supplementary figure 3: Evolutionary conservation of metazoan SMG9



Evolutionary conservation of metazoan SMG8 & SMG9 (Supplementary figures 2 & 3)

Sequence alignment of SMG8 and SMG9 from C. elegans (Ce), D. melanogaster (Dm), H. sapiens (Hs), D. rerio (Dr) and X. tropicalis (Xt). Conserved residues of SMG8 and SMG9 are highlighted in orange and blue, respectively. Above the sequences are the secondary structure elements, labeled as in the topology diagram in Figure 1B. The G motifs of SMG9 are indicated.

Supplementary figure 4: Electron density of the GDP molecule bound to SMG9



SMG9-bound GDP

Zoom-in view of the GDP molecule bound to SMG9.The model is overlaid with the refined 2mFo-DFc map (grey mesh) showing the density for GDP and the map is contoured at 1.0σ

3.3 STRUCTURE OF SMG1-SMG8-SMG9

Gat.Y., Schuller.J.M., Lingaraju.M., Weyher.E., Bonneau.F., Strauss.M., Murray.P.J., Conti.E. InsP6 binding to PIKKs revealed by the cryo-EM structure of a SMG1-SMG8-SMG9 complex. Nat. Struct. Mol. Biol 12. 1089-1093.

The study presents the first high-resolution structure of the human SMG1-SMG8-SMG9 complex. The study shows that IP_6 is a structural co-factor of SMG1 and possibly other PIKKs and is required for their activity.

The work was performed under the supervision of Prof. Conti E. Lingaraju M. along with Bonneau F. cloned the constructs, established the mammalian expression system and initial purification protocols for the complex. Detailed author contributions are included in the article attached.

InsP6 binding to PIKKs revealed by the cryo-EM structure of a SMG1-SMG8-SMG9 complex

Authors: Yair Gat[1]†, Jan Michael Schuller[1]†, Mahesh Lingaraju[1], Elisabeth Weyher[2], Fabien Bonneau[1], Mike Strauss[3] ‡, Peter J. Murray[4], Elena Conti[1*].

Affiliations:

[1]Department of Structural Cell Biology, MPI for Biochemistry, Munich, Germany.

[2]Biochemistry Core Facility, MPI for Biochemistry, Munich, Germany.

[3]Cryo-EM Facility, MPI for Biochemistry, Munich, Germany.

[4]Immunoregulation group, MPI for Biochemistry, Munich, Germany.

*Correspondence to: conti@biochem.mpg.de.

†These authors contributed equally to the work.

‡Present address: Department of Anatomy and Cell Biology, McGill University, Montreal, Canada.

Short Title: Cryo-EM structure of the SMG1 complex

Abstract

We report the 3.45-Å resolution cryo-EM structure of human SMG1-SMG8-SMG9, a phosphatidylinositol-3-kinase (PI3K)-related protein kinase (PIKK) complex central to mRNA surveillance. Structural and mass spectrometry analyses reveal the presence of inositol hexaphosphate (IP_6) in the SMG1 kinase. We show that the IP_6-binding site is conserved in mTOR and potentially other PIKK members, and that it is required for optimal in vitro phosphorylation of both SMG1 and mTOR substrates.

The SMG1 kinase performs a crucial phosphorylation event in the nonsense-mediated mRNA decay (NMD) pathway, a surveillance process that targets and degrades defective mRNAs containing premature translation termination codons (PTCs)[1,2]. Although the precise mechanisms of NMD are not fully understood, a general consensus posits that upon premature translation termination, this PIKK phosphorylates Ser/Thr motifs of the NMD factor UPF1[3-6], enabling the recruitment of downstream effectors that promote degradation of PTC-containing transcripts[1,2]. SMG1 constitutively binds to two cofactors, SMG8 and SMG9[7]. Unlike other PIKKs (the cytoplasmic mTOR and the nuclear proteins DNA-PK, ATM, ATR and TRRAP)[8,9], structural data for the SMG1-SMG8-SMG9 complex is currently limited by low resolution[10-12].

To gain mechanistic insights into this important NMD factor and relationships between PIKKs, we first generated stable mammalian cell lines expressing streptavidin-tagged versions of wild-type (WT) human SMG1 (for biochemical analyses) or an inactive Asp2335Ala mutant (for cryo-EM analyses). Both co-purified with endogenous SMG8 and SMG9, though higher yields were obtained when co-expressing all three proteins (Supplementary Fig. 1). The resulting 3.45 Å resolution structure (Supplementary Fig. 2 and 3) allowed building of a *de novo* atomic model of SMG1-SMG8-SMG9 core complex (Fig. 1, Supplementary Table 1, Supplementary Fig. 4, 5).

SMG1 consists of a globular 'head' and an extended 'arch' (Fig. 1). Like for other PIKKs, the head contains the catalytic module (also referred to as FATKIN[8,9]), which consists of the two lobes typical of Ser-Thr protein kinases packed against the so-called FAT and FATC domains[13]. Remarkably, the well-ordered FATC domain of SMG1 follows a

disordered insertion spanning more than 1100 residues (Fig. 1a). The FAT domain includes a curved 'spine' of 8 irregular tetratricopeptide repeats (TPR) and another helical repeat region with a 'ring' structure (Fig. 1). As reflected in the name of this protein kinase family, the spine of SMG1 and other PIKKs is similar (but more extended) to that found in PI3K, a lipid kinase that binds and phosphorylates the headgroup of phosphatidylinositols[8]. Finally, attached to the globular head is the arch structure, a curved solenoid of 14 N-terminal HEAT repeats. In contrast to other PIKKs, the N-terminal solenoid of SMG1 is a single structural unit and does not mediate homodimerization. Similarly to other members of this kinase family, however, the N-terminal solenoid forms the binding platform for recruiting regulatory proteins, SMG8 and SMG9 (Fig. 2a).

Human SMG8 and SMG9 contain a globular core with architectural similarities to the dynamin-like GTPase family, as previously observed in the crystal structure of the *C. elegans* orthologues[14]. Two loop regions in the SMG8 and SMG9 G-folds provide the main SMG1-binding determinants. First, a helix of SMG8 binds the SMG1 arch, engaging the surface groove between HEAT 8 and HEAT 10 with conserved hydrophobic interactions (Fig. 2b). Second, an extended segment of SMG9 (residues 435-464) interacts at the parallel surface groove in the SMG1 arch formed between HEAT 10 and 11 (Fig. 2c). Finally, a short stretch of residues extends from SMG9 and contacts the ring region of the SMG1 FAT domain (Fig. 2a). Overall, the SMG8-SMG9 heterodimer wedges between the 'head' and the 'arch' of SMG1, positioning the G-fold of SMG9 to face the kinase active site of SMG1. Although this is reminiscent of the other cytoplasmic PIKK, mTOR, which binds a small RAS-superfamily GTPase (Supplementary Fig. 6)[15], the dynamin-like SMG8-SMG9 heterodimer has unusual biochemical properties. Only the SMG9 G-domain contains a nucleotide-binding site though it lacks residues required for hydrolysis activity[14]. Consistently, only SMG9 is bound to a nucleotide in our cryo-EM reconstruction (Supplementary Fig. 7a). As the resolution was not sufficient to unambiguously identify the endogenous ligand that co-purifies together with human SMG9, we set out to biochemically identify it. We purified SMG8-SMG9 (Supplementary Fig. 7b) and subjected it to electrospray ionization-mass spectroscopy (ESI-MS), coupled to reversed-phase ion-pair high-performance liquid chromatography (RPIP-HPLC). Surprisingly, the mass spectrometry analysis revealed the presence of a low-molecular weight mass corresponding to ATP (506 Da) and no GDP nor GTP

(Supplementary Fig. 7c). The biochemical basis and relevance for this unusual nucleotide in a G-fold domain awaits further investigations.

During model building, we noted that the polypeptide chain of the SMG1 PI3K-like module forms a cavity lined by conserved positively charged residues, containing a strong spherical electron density (Fig. 3a). To identify small molecules non-covalently bound to the ~578-kDa SMG1-SMG8-SMG9 complex, we subjected the purified sample to the same RPIP-HPLC - ESI-MS analysis described above. In the MS analysis of the ternary complex, we detected not only the low-molecular weight mass corresponding to the SMG9-bound ATP (506 Da), but also a prominent 659-Da peak (Fig. 3b). Coupled to collision-induced dissociation MS (Supplementary Fig. 8), we could unambiguously assign the 659 Da peak to inositol hexaphosphate (IP_6), a small negatively-charged metabolite that is present in the cytoplasm at micromolar concentrations[16]. Thus, IP_6 constitutively bound *in vivo* to SMG1 and remained associated with the protein complex during purification.

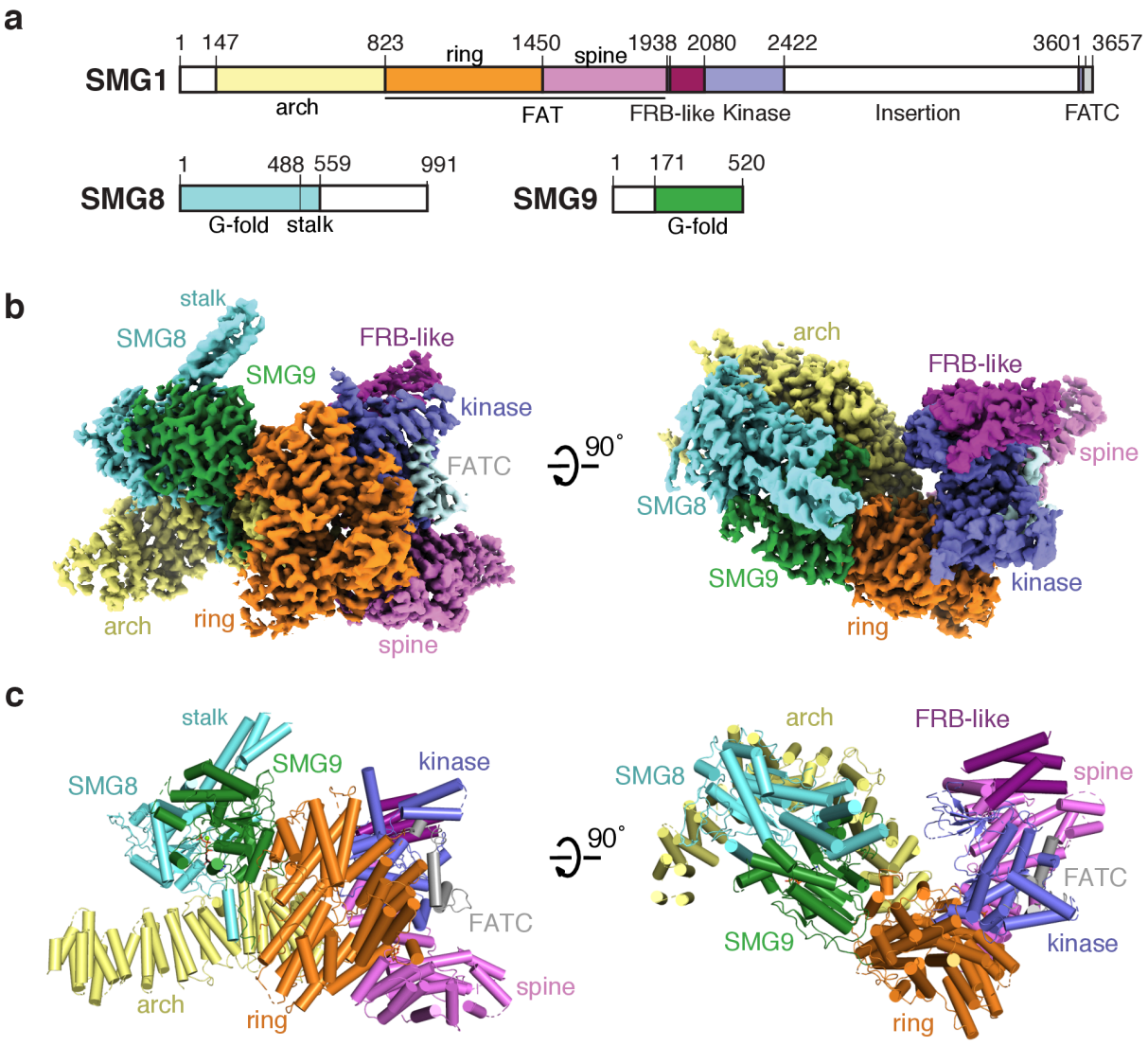
The IP_6 ligand fitted well in the distinctive properties of the electron density and chemical features of the protein environment (Fig. 3a). On one side, IP_6 phosphates interact with conserved positively-charged residues of the SMG1 FAT domain (Fig. 3a, Supplementary Fig. 5), notably starting from helical repeats of the spine that extend beyond those found in the PI3K lipid kinase. On the other side, IP_6 phosphates interact with the conserved hinge between the kinase lobes (Fig. 3a). Thus, IP_6 appears to play a structural role in locally stabilizing and/or configuring lobes of the protein kinase domains and may therefore be required for optimal catalytic activity. To test this, we expressed and purified a SMG1 mutant whereby we substituted a positively-charged IP_6-binding lysine residue of the helical spine (Lys1530, Fig. 3a) to a negatively charged glutamic acid (SMG1-K1530E) in order to disfavor IP_6 binding by electrostatic repulsion. We then assayed SMG1 kinase activity using its physiological substrate, UPF1. Unlike WT-SMG1, SMG1-K1530E only poorly phosphorylated UPF1, indicating that an intact IP_6-binding site is required for full SMG1 activity *in vitro* (Fig. 3c).

Superimposing the SMG1 structure with atomic models of other PIKK family members revealed a very similar positively-charged cavity at the corresponding position in

mTOR (Fig. 3d). Inspection of mTOR electron density maps[15,17,18] indeed showed un-modeled density that could be interpreted as a partially ordered IP_6 molecule (Fig. 3d). To identify the endogenous ligand bound to mTOR, we expressed and purified a complex of mTOR (residues 1376-2549) and its binding factor mLST8 (as described in [15], hereby defined $mTOR^{\Delta N}$) and subjected it to the RPIP-HPLC - ESI-MS analysis described above. In the MS data, we identified a 659-Da peak $mTOR^{\Delta N}$, corresponding to a constitutively bound IP_6 (Fig. 3e). To test the functional relevance of the bound IP_6 molecule, we engineered $mTOR^{\Delta N}$ mutants with a single K1753E substitution (corresponding to SMG1 K1530E) or with a double K1753E-K1788E substitution (referred to as KE and KKEE mutants), and assayed for kinase activity using an mTOR substrate, AKT1 (residues 450-480). In these assays, the phosphorylation activity of $mTOR^{\Delta N}$ WT was reduced in the $mTOR^{\Delta N}$ KE single mutant, and was undetectable in the $mTOR^{\Delta N}$ KKEE double mutant (Fig. 3F), suggesting that IP_6 is important for full $mTOR^{\Delta N}$ activity *in vitro*.

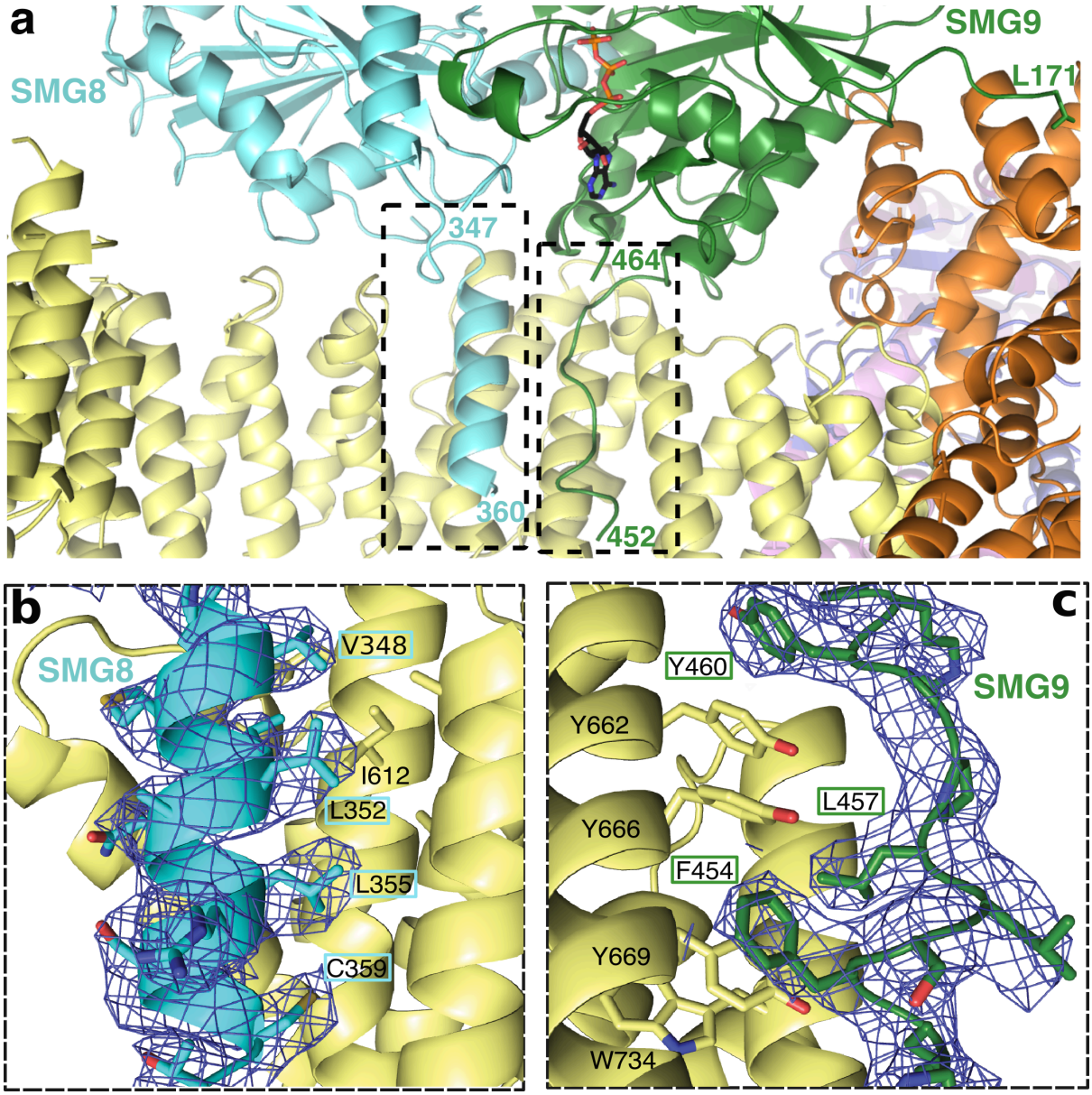
Although mTOR is the PIKK with the most extensive similarity to SMG1 at the IP_6-binding pocket, the available structural data suggest that at least some of the nuclear members of the PIKK family, ATM/ATR and TRRAP in particular, may also feature an IP_6 binding site (Supplementary Fig. 9). In contrast, DNA-PK appears to have evolved alternative structural elements that replace the architectural role of IP_6 (Supplementary Fig. 9, consistent with[19,20]). In summary, IP_6 is an integral structural element of the catalytic module of SMG1, mTOR and possibly other members of the phosphatidylinositol-3-kinase (PI3K)-related protein kinase family.

Figure 1



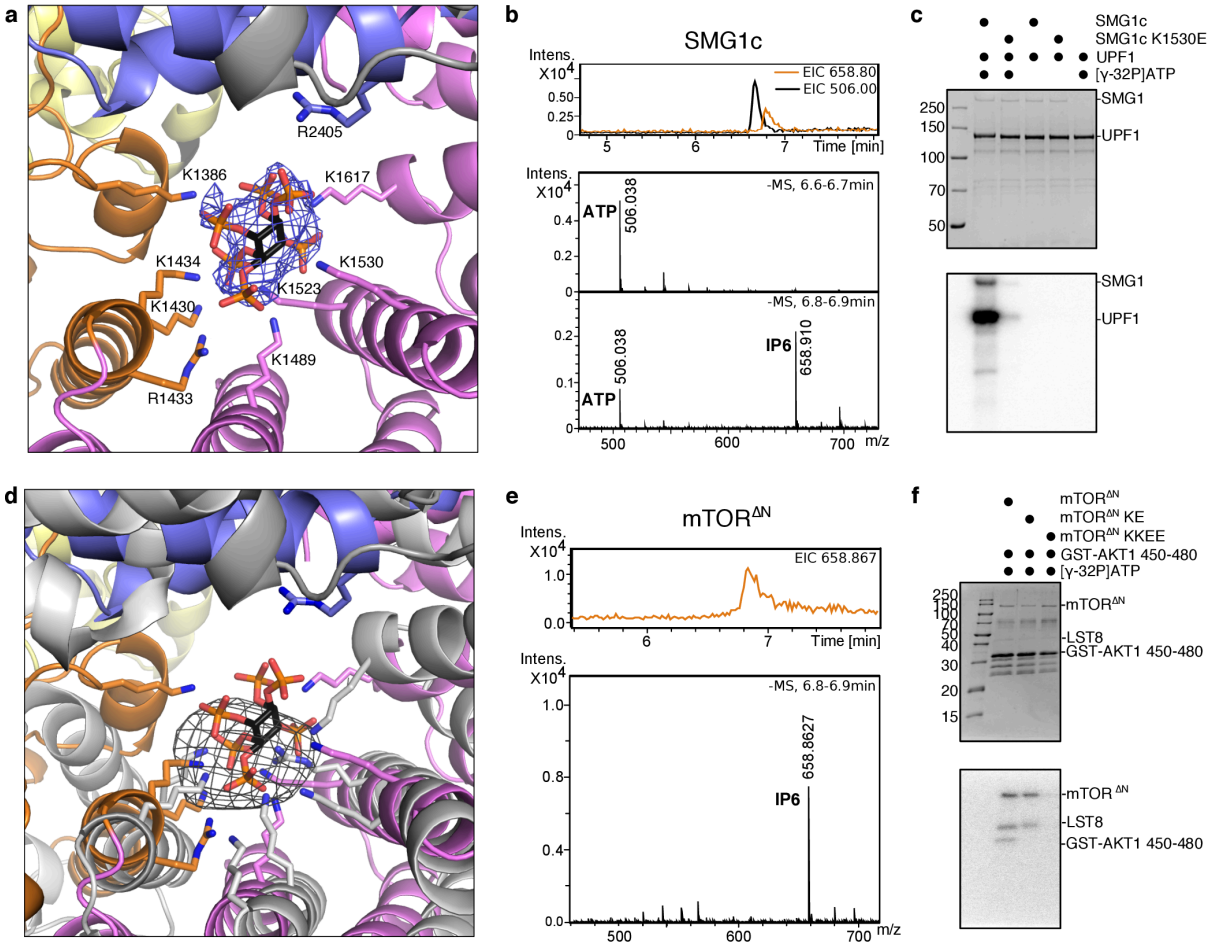
Cryo-EM structure of human SMG1-SMG8-SMG9. **(a)** Schematic representation of the domain arrangements. Regions lacking ordered density are in white. **(b)** Two orthogonal views of the cryo-EM density map, segmented and colored by domain, and **(c)** corresponding atomic model in cartoon representation.

Figure 2



Interactions between SMG1 and SMG8-SMG9 **(a)** Overall view, with protein domains colored as in Fig. 1. **(b, c)** Zoom-in views of interactions with cryo-EM densities superimposed.

Figure 3



IP_6-binding site of SMG1 and mTOR **(a)** Structure of SMG1 IP_6-binding site with density around IP_6. **(b)** Ion-pair HPLC-MS spectrum. HPLC extracted-ion chromatogram (top) and MS analysis of the peak fraction indicated (bottom). **(c)** *In vitro* kinase assay with purified SMG1-SMG8-SMG9 and UPF1 substrate. Coomassie-stained gel (top), phosphorylation assay with [γ^{32}P]-ATP (bottom). **(d)** Atomic model of the SMG1 IP_6-binding site superimposed with the corresponding region of mTOR[17] (in grey), with a previously unmodeled density feature (Fo-Fc, 4σ) ([18,15,17]). **(e)** Ion-pair HPLC-MS spectrum of $mTOR^{\Delta N}$ (as in panel b). **(f)** *In vitro* kinase assay with purified $mTOR^{\Delta N}$ complexes and GST-ATK1 450-480 substrate (as in panel **c**).

References and Notes:

1. Kurosaki, T. & Maquat, L. E. *J Cell Sci* **129,** 461–467 (2016).
2. Karousis, E. D. & Mühlemann, O. *Cold Spring Harb Perspect Biol* a032862 (2018). doi:10.1101/cshperspect.a032862
3. Yamashita, A., Ohnishi, T., Kashima, I., Taya, Y. & Ohno, S. *Genes Dev* **15,** 2215–2228 (2001).
4. Denning, G., Jamieson L., Maquat L.E., Thompson E.A., & Fields A.P. *J. Biol. Chem*. **276,** 22709-22714 (2001).
5. Ohnishi, T. *et al. Mol Cell* **12,** 1187–1200 (2003).
6. Yamashita, A. *Genes Cells* **18,** 161–175 (2013).
7. Yamashita, A. *et al. Genes Dev* **23,** 1091–1105 (2009).
8. Baretić, D. & Williams, R. L. *Curr Opin Struct Biol* **29,** 134–142 (2014).
9. Imseng, S., Aylett, C. H. & Maier, T. *Curr Opin Struct Biol* **49,** 177–189 (2018).
10. Arias-Palomo, E. *et al. Genes Dev* **25,** 153–164 (2011).
11. Deniaud, A. *et al. Nucleic Acids Res* **43,** 7600–7611 (2015).
12. Melero, R. *et al. Structure* **22,** 1105–1119 (2014).
13. Bosotti, R., Isacchi, A. & Sonnhammer, E. L. *Trends Biochem Sci* **25,** 225–227 (2000).
14. Li, L., Lingaraju, M., Basquin, C., Basquin & Conti, E. *RNA* **23,** 1028-1034 (2017).
15. Yang, H. *et al. Nature* **552,** 368–373 (2017).
16. Letcher, A. J., Schell, M. J. & Irvine, R. F. *Biochem J* **416,** 263–270 (2008).
17. Yang, H. *et al. Nature* **497,** 217–223 (2013).
18. Aylett, C. H. S. *et al. Science* **351,** 48–52 (2016).
19. Hanakahi, L. A. & West, S. C. (2002). *EMBO J.* **21,** 2038-2044.
20. Ma, Y. & Lieber, M. R. (2002). *J. Biol. Chem.* **277,** 10756-10759.

Acknowledgements: We thank Sabine Suppmann and Nagarjuna Nagaraj from the Max Planck Institute of Biochemistry (MPIB) Core Facilities for advise on protein expression and for data in Supplementary Fig. 1, Daniela Wartini for excellent support with mammalian cell culture, and the MPIB cryo-EM facility for microscope access. We also thank members of our group for comments and discussions, in particular Lukas Langer.

Funding: This study was supported by grants from the Max Planck Gesellschaft and the German Research Foundation (DFG SFB1035, GRK1721, SFB/TRR 237) to E.C. and a Boehringer Ingelheim Fonds fellowship to M.L.

Author contributions: J.M.S. and Y.G. expressed, purified, and prepared protein complex from mammalian cell lines established by M.L. and F.B; cryo-EM data were collected by M.S. and processed by J.M.S.; Y.G. built the atomic model and performed biochemical assays; E.W. conducted the MS experiments; P.J.M. contributed to plan mTOR experiments; Y.G., J.M.S., and E.C. analyzed the structure and wrote the manuscript.

Competing Interests: Authors declare no competing interests.

Data and Materials Availability: Cryo-EM density maps are deposited in the Electron Microscopy Data Bank, under accession numbers EMD-10347 and EMD-10348. The atomic model is deposited in the PDB, under accession number 6SYT.

Supplementary Data

InsP6 binding to PIKK kinases revealed by the cryo-EM structure of a SMG1-SMG8-SMG9 complex

Authors: Yair Gat[1]†, Jan Michael Schuller[1]†, Mahesh Lingaraju[1], Elisabeth Weyher[2], Fabien Bonneau[1], Mike Strauss[3] ‡, Peter J. Murray[4], Elena Conti[1*].

Affiliations:

[1]Department of Structural Cell Biology, MPI for Biochemistry, Munich, Germany.

[2]Biochemistry Core Facility, MPI for Biochemistry, Munich, Germany.

[3]Cryo-EM Facility, MPI for Biochemistry, Munich, Germany.

[4]Immunoregulation group, MPI for Biochemistry, Munich, Germany.

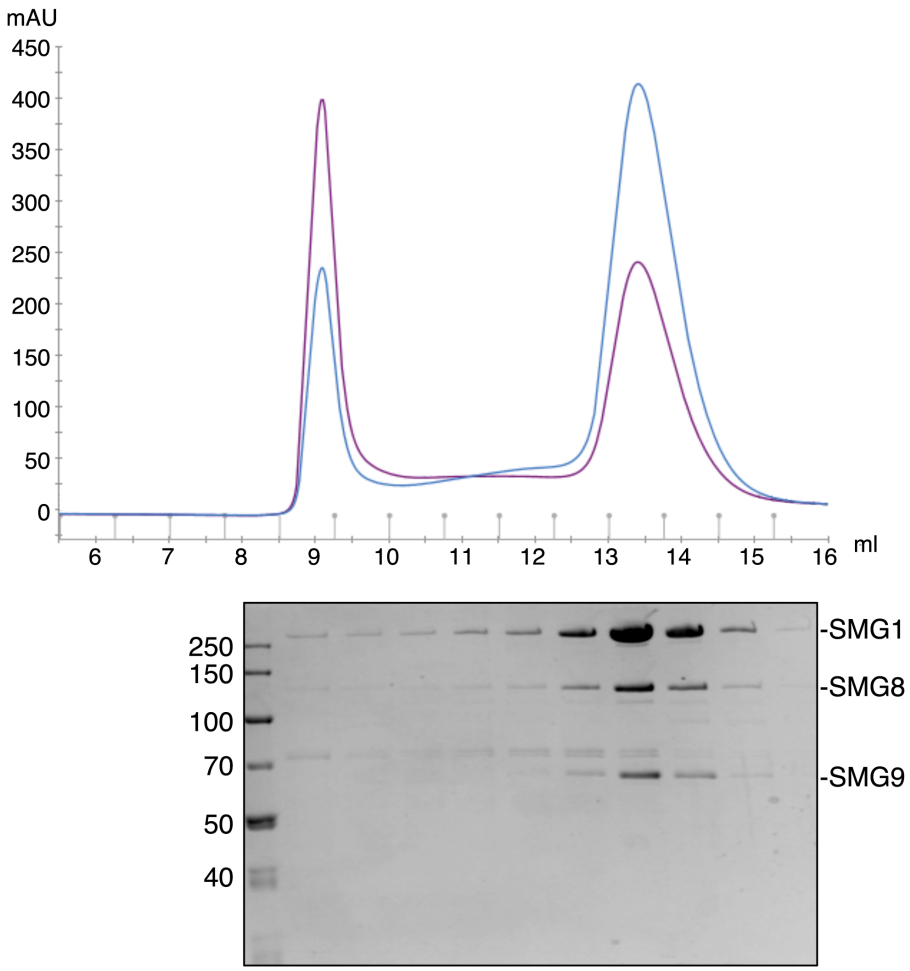
*Correspondence to: conti@biochem.mpg.de.

†These authors contributed equally to the work.

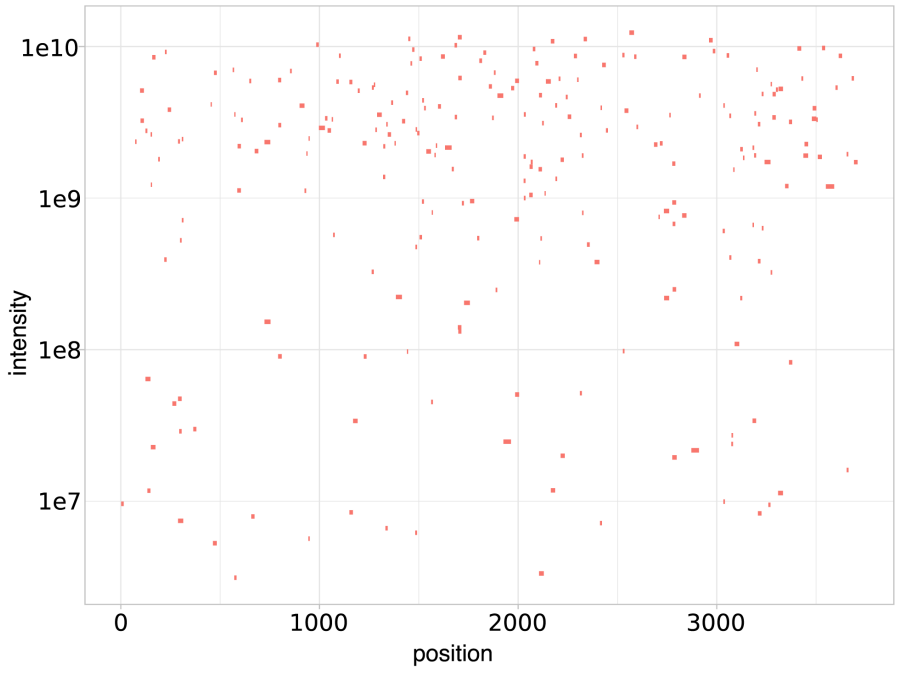
‡Present address: Department of Anatomy and Cell Biology, McGill University, Montreal, Canada.

Supplementary Figure 1

a

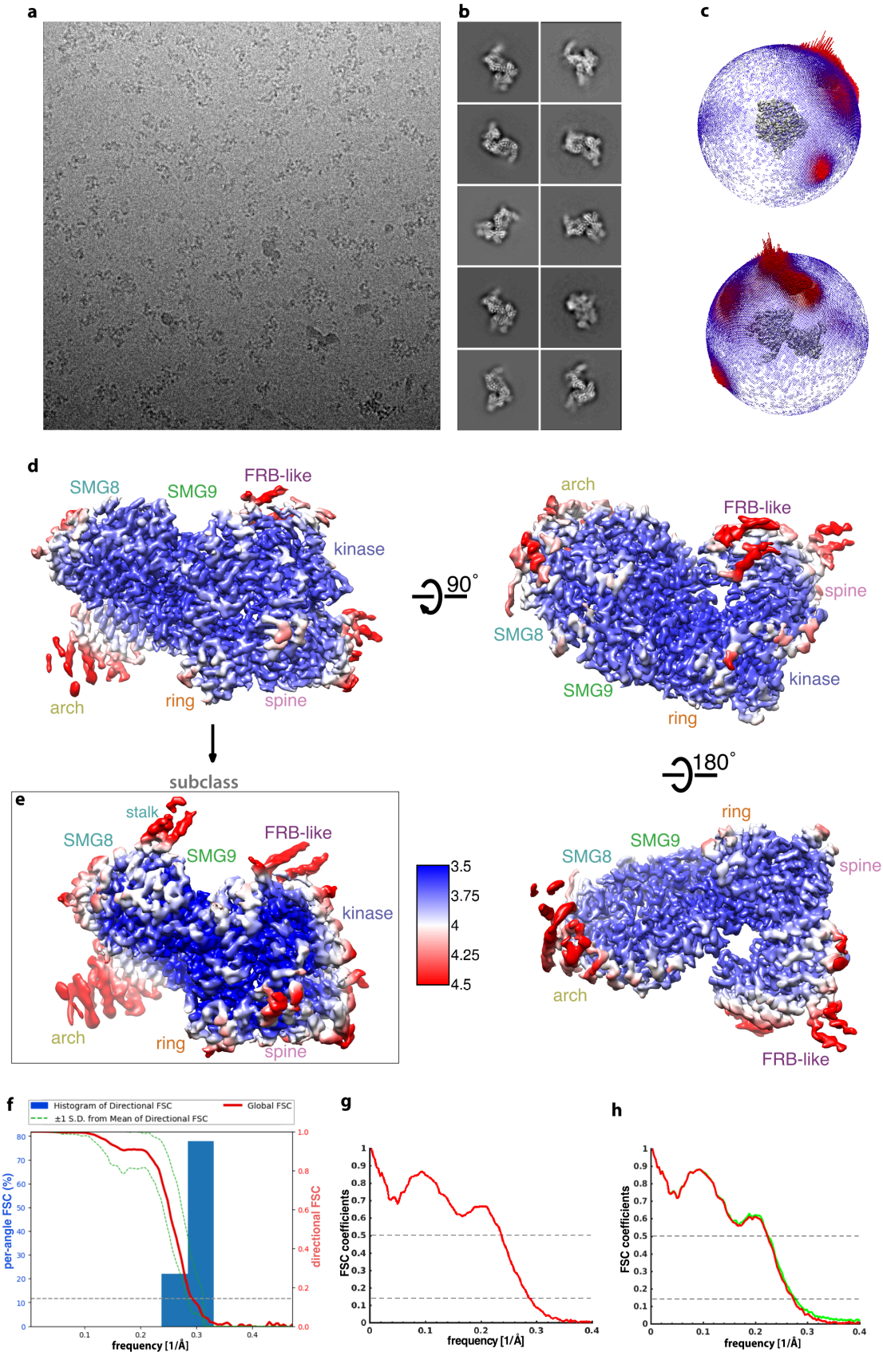


b



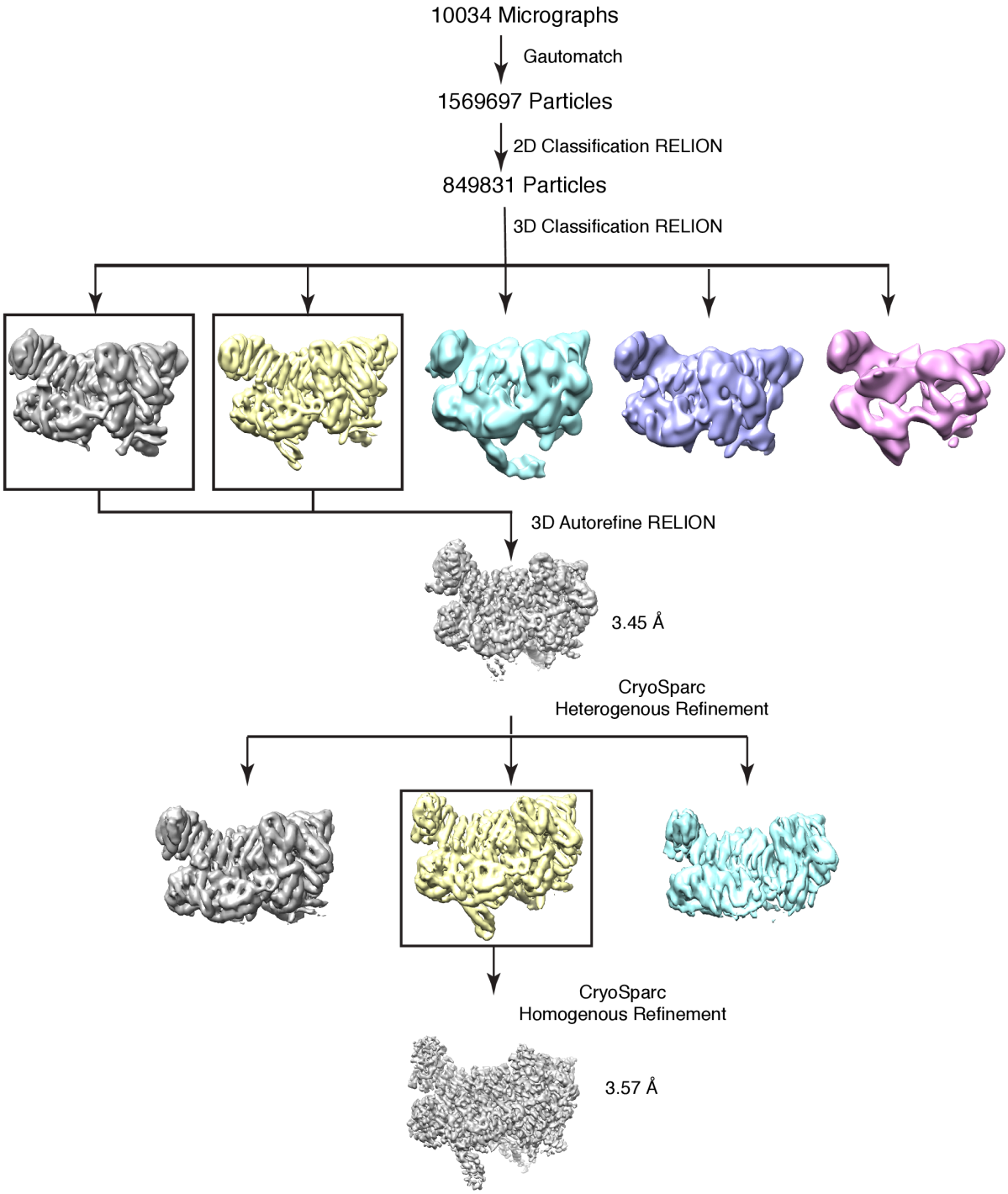
Biochemical characterization of recombinant SMG1-SMG8-SMG9 complex (SMG1c). **(a)** Size-exclusion chromatography assay showing the formation of homogeneous SMG1-SMG8-SMG9 ternary complex. After expression in mammalian cells, the complex was purified by size-exclusion column (Superose 6 Increase 10/300 GL, void volume of 8.0 ml). Top panel: chromatography profile, with absorbance at 280 nm and 260 nm shown as blue and purple traces, respectively. Bottom panel: Coomassie-stained SDS-PAGE gel, with samples from each fraction that eluted between 8.5–16 ml; the SMG1c eluted between 13–15 ml. **(b)** Mass spectrometry (MS) analysis of purified SMG1. Graphical representation shows the result of an in-gel peptide mass fingerprinting experiment. Identified peptides are indicated as red dashes at the corresponding position in the SMG1 sequence (horizontal axis) and are plotted against their intensity, as detected by MS (vertical axis). The uniform distribution of detected peptides over the entire amino acid sequence indicates that the entire SMG1 polypeptide is present in the purified sample.

Supplementary Figure 2



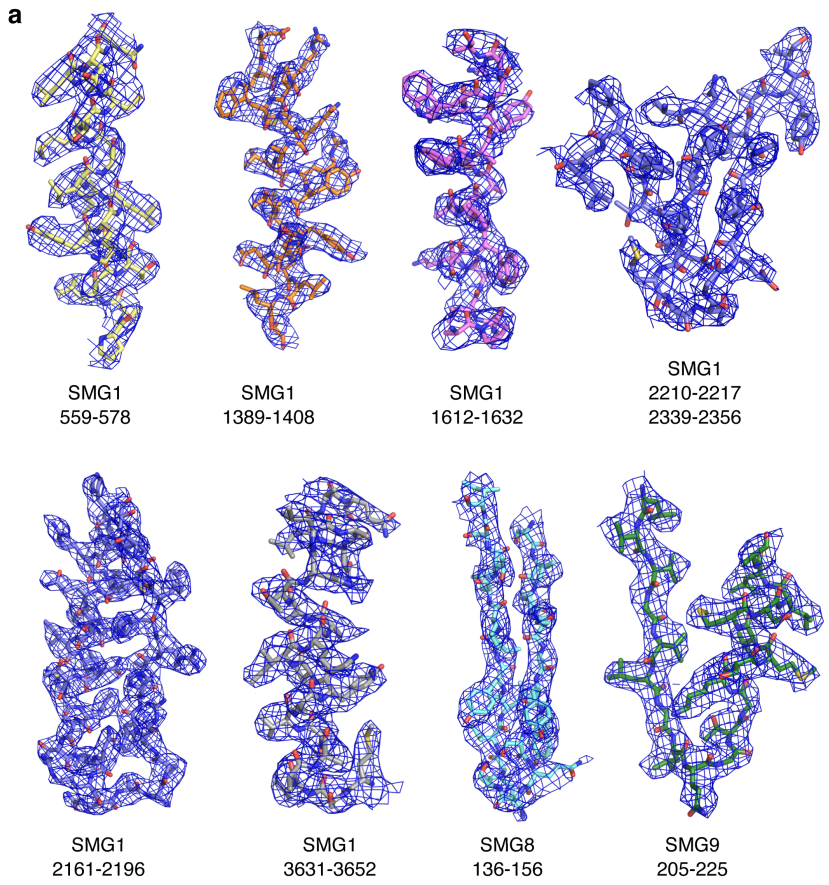
Cryogenic electron microscopy (Cryo-EM) data collection and analysis. **(a**) A representative cryo-EM micrograph collected on an FEI Titan Krios microscope, operated at 300 kV and equipped with a K2 Summit camera. (**b**) Representative reference-free 2D class averages are shown. (**c**) Angular distribution of the particles used for the final round of refinement. (**d**) Local-resolution analysis of the SMG1c. Map shows the variation in local resolution, as estimated by RELION. (**e**) Local-resolution analysis of the SMG1c (including the C-terminal region of SMG8). Map shows the variation in local resolution, as estimated by cryoSPARC. (**f**) 3D FSC and preferred orientation analysis of the dataset with the red line representing the estimated global FSC of 3.45 Å ± 1 SD (green dashed lines). A sphericity of 0.943 indicates a mostly isotropic map without preferred orientation bias. (**g**) Model vs. map FSC for the final PHENIX real-space refined model. (**h**) The model was probed for over-fitting by randomly perturbing the atoms by 0.5 Å and refining against the first of the two independent half-maps (work half-map, red). The resulting refined model was then used to calculate a model-map FSC against the second half-map (test half-map, green), which was not used for refinement. FSC-work and FSC-test curves show excellent agreement over the entire resolution range, validating the entire structure against over-fitting.

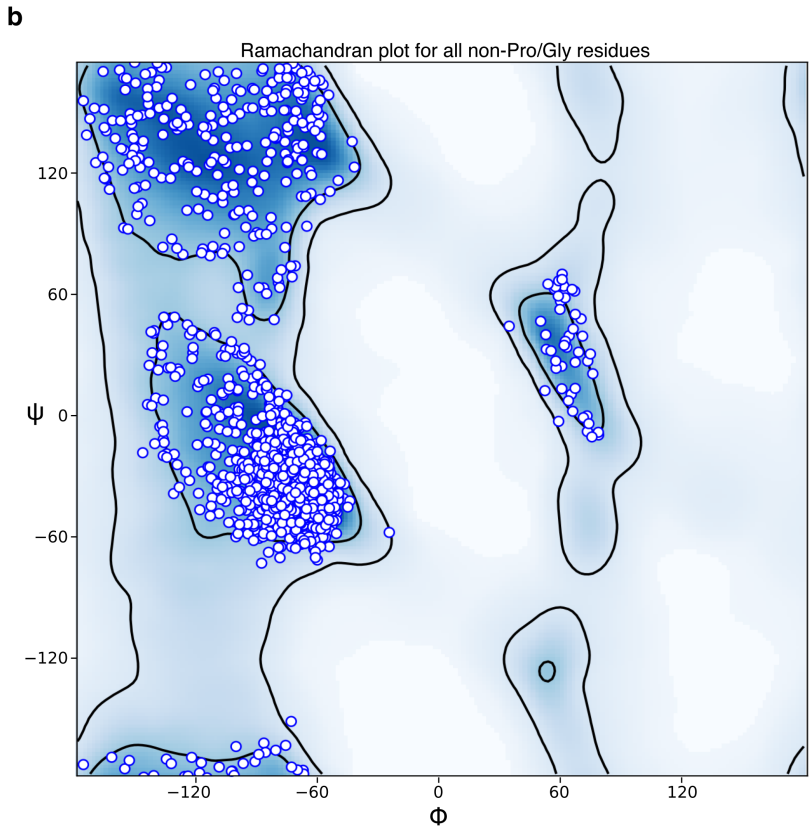
Supplementary Figure 3



Cryo-EM data processing scheme. The 849,831 particles from the 2D classification were initially 3D-classified into five classes. The two classes containing the best aligning particles were combined and auto-refined in RELION to a resolution of 3.45 Å. The dataset was further classified for the C-terminus of the SMG8 protein in cryoSPARC, using heterogeneous refinement. A class-showing density for the SMG8 C-terminal region was then selected and refined using homogenous refinement to a resolution of 3.57 Å.

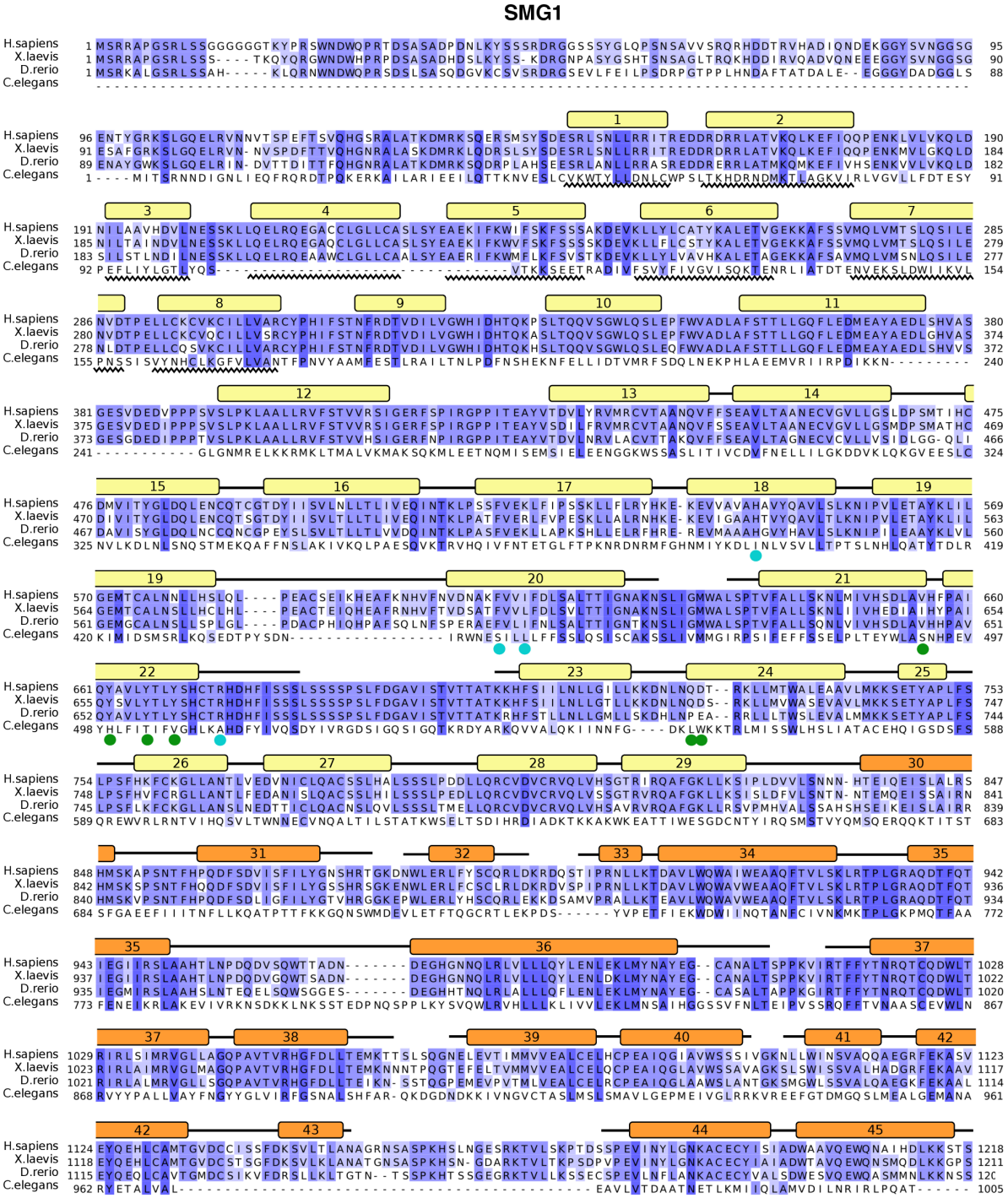
Supplementary Figure 4

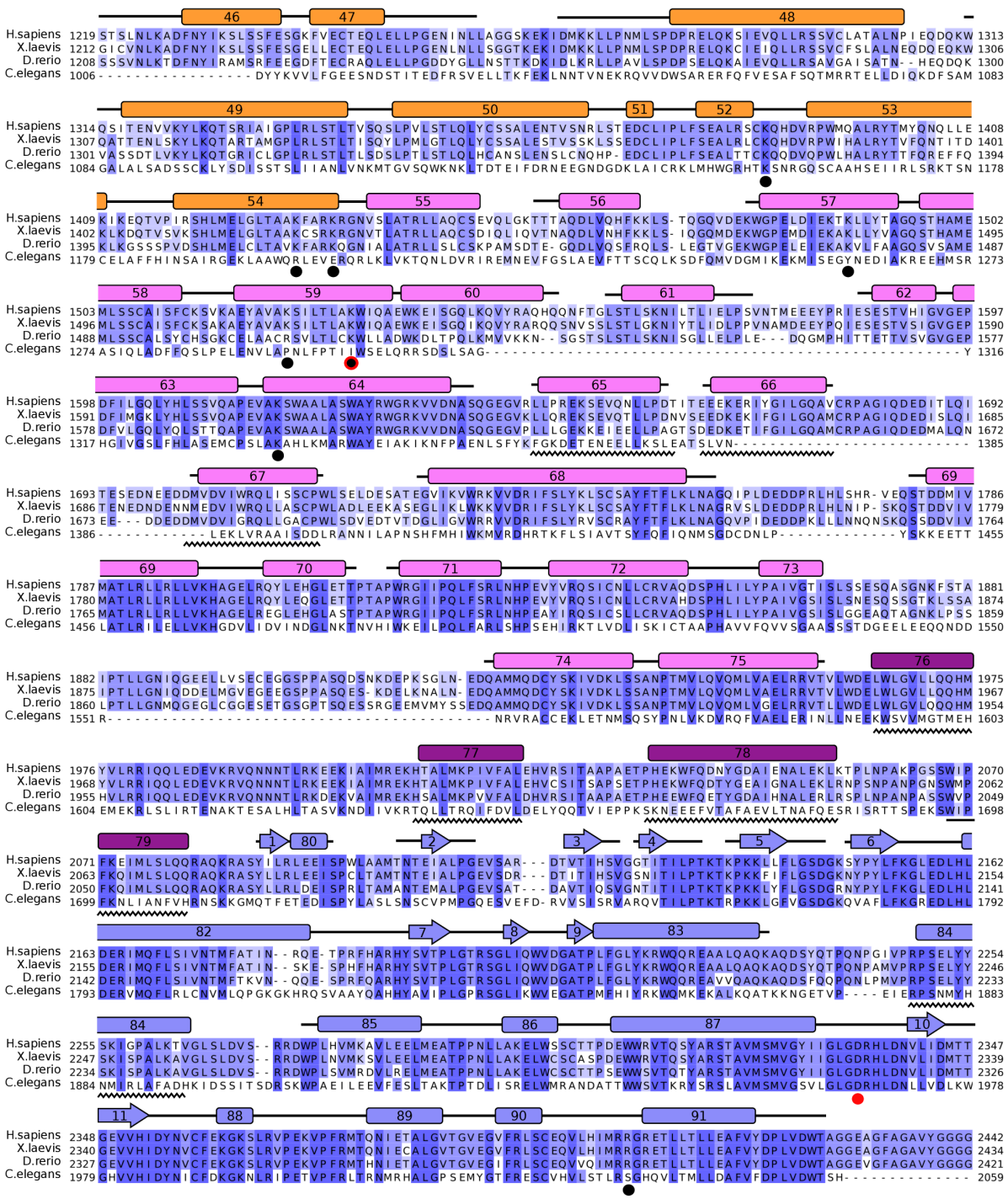


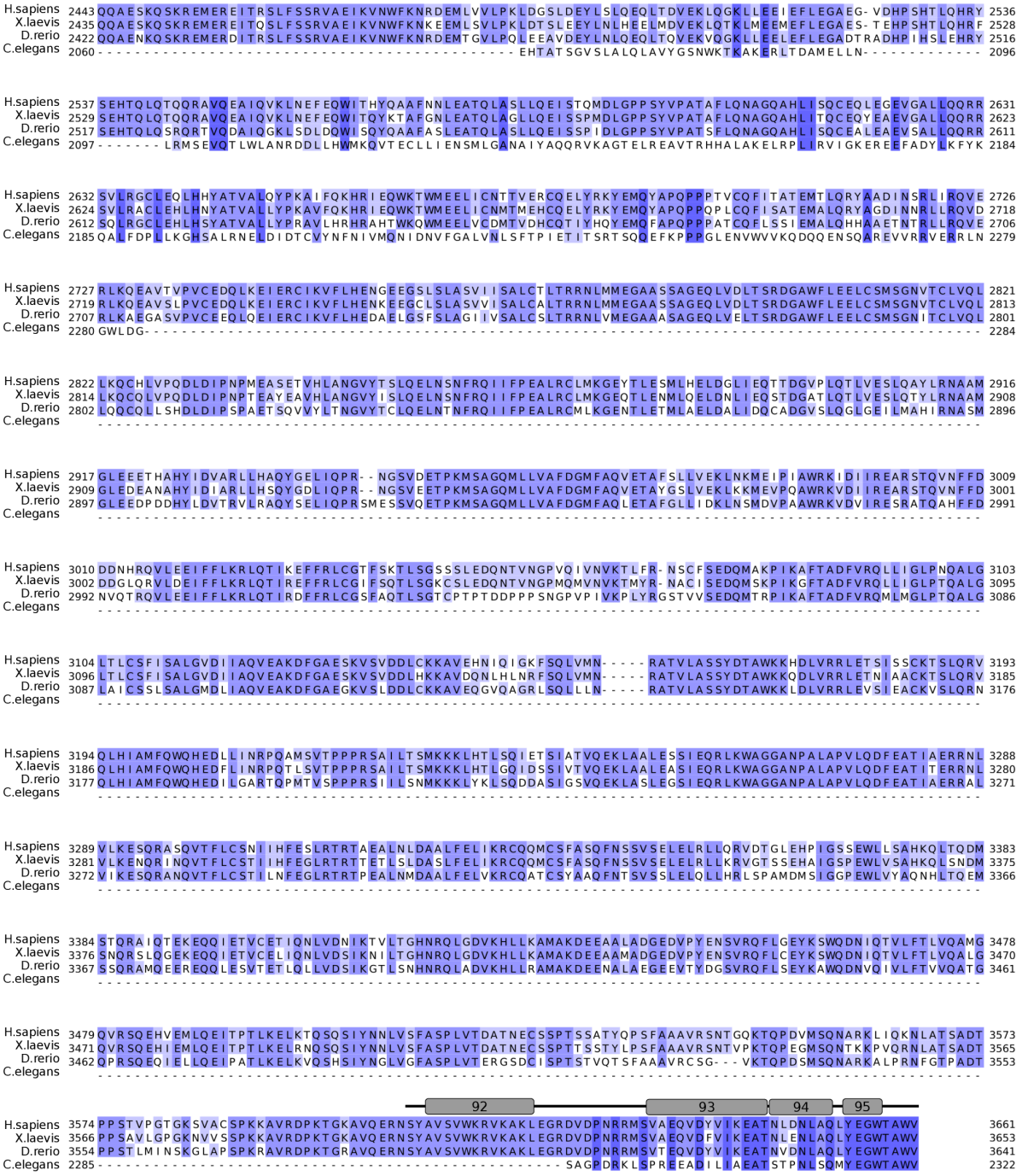


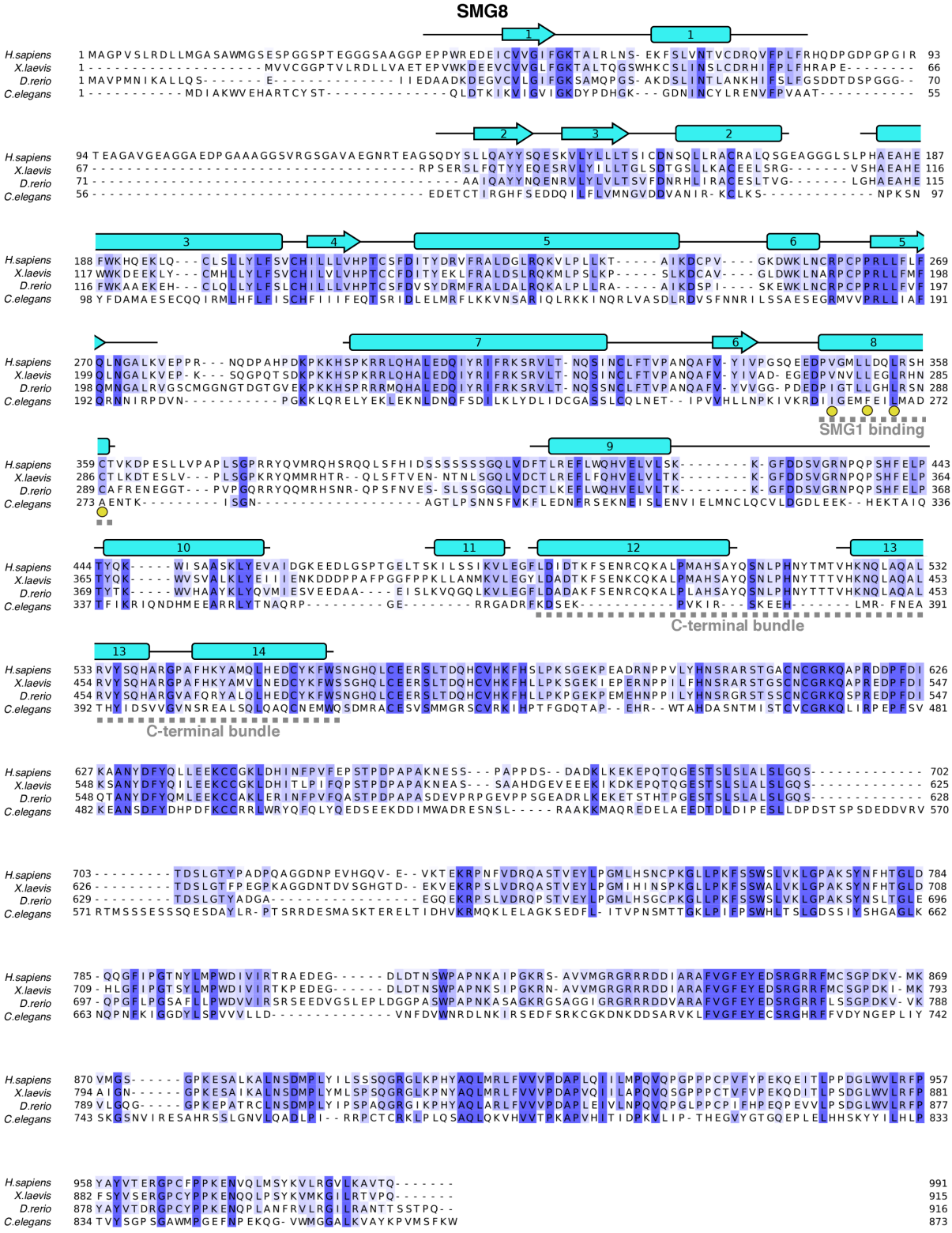
Quality of the structural model built *de novo* in the cryo-EM map. (**a**) Representative regions of the SMG1c and surrounding electron density are shown. (**b**) Ramachandran plot of the main chain ϕ/ψ-conformational angles of the SMG1-SMG8-SMG9 atomic model. Areas of favored ϕ/ψ-combinations are defined in dark blue (see also Supplementary Table 1).

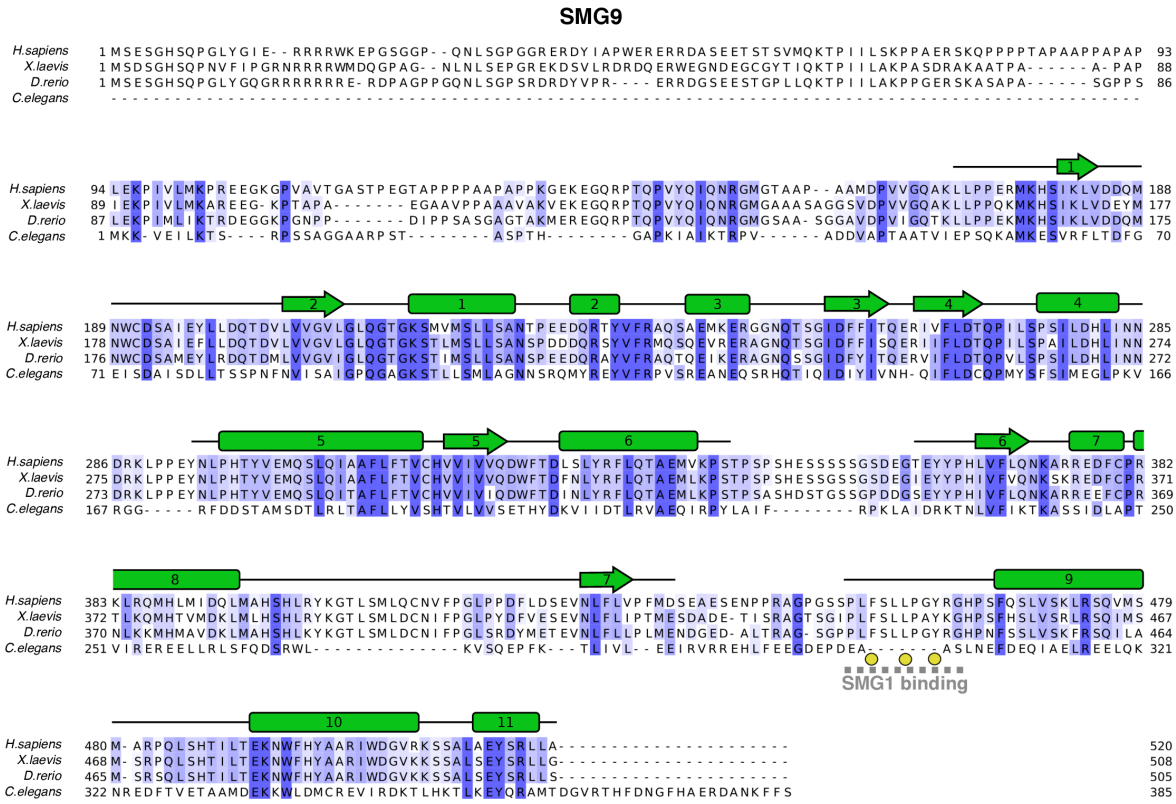
Supplementary Figure 5




H.sapiens
X.laevis
D.rerio
C.elegans

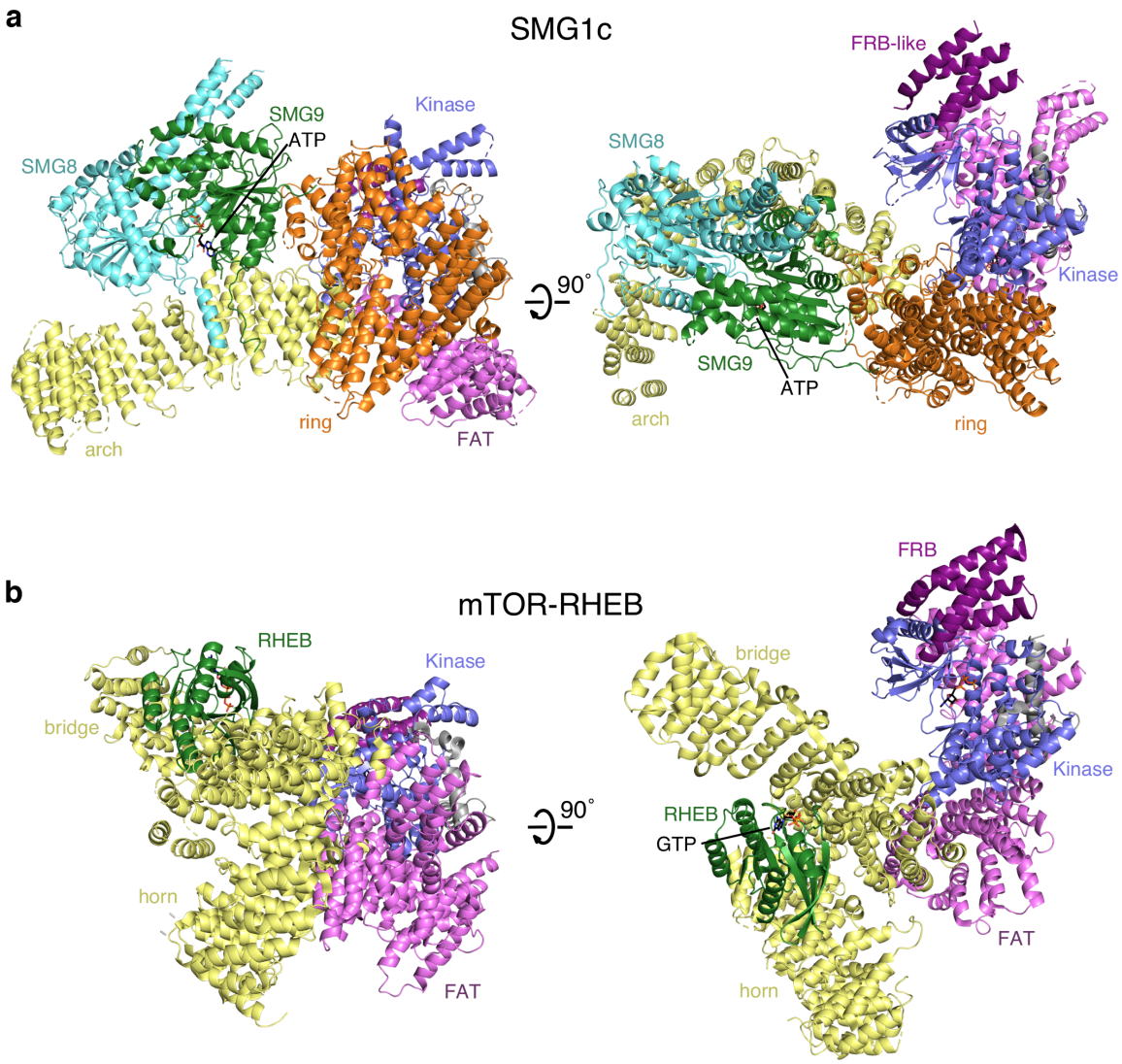

H.sapiens
X.laevis
D.rerio
C.elegans
92
93
94
95


SMG8
SMG1 binding
C-terminal bundle
C-terminal bundle



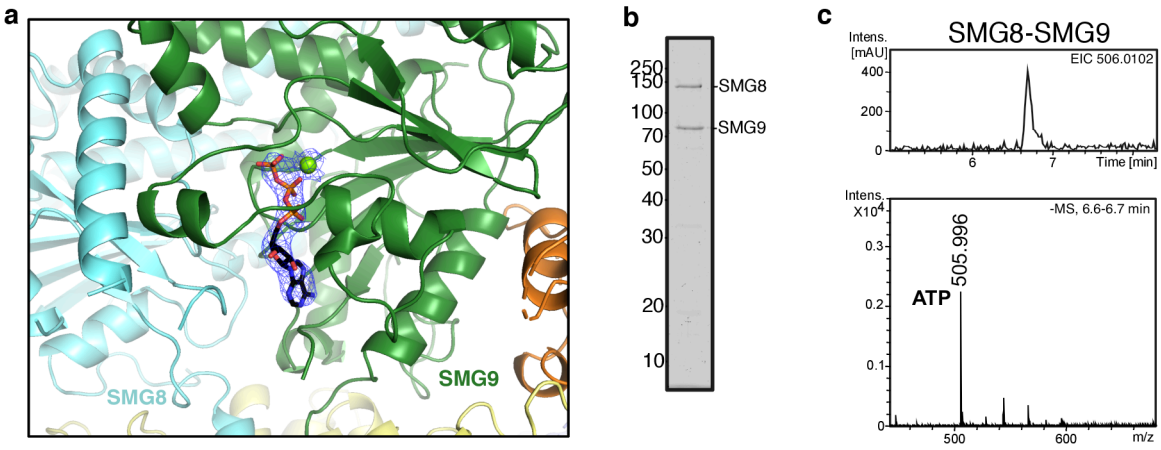
Structure-based sequence conservation. The structure-based sequence alignments of SMG1, SMG8 and SMG9 were performed using Clustal Omega and colored by conservation using Jalview. The SMG1 alignment includes orthologues from *Homo sapiens* (3661 amino acids), *X. laevis* (3653 aa), *Danio rerio* (3641 aa), and *Caenorhabditis elegans* (2322 aa). Secondary structure elements are shown above the sequences (rectangles for α-helices and arrows for β-strands), colored as in Fig. 1a. Helices and strands are numbered sequentially according to the primary modeled structure. The zigzag line below the sequences represent regions left as a poly-alanine-level model. Cyan and green circles indicate the residues interacting with SMG8 and SMG9, respectively. Red circles indicate the residue mutated at the kinase active site and the IP_6-binding site, and black circles indicate IP_6-binding residues. SMG8 and SMG9 are aligned, numbered and colored as SMG1. Yellow circles are shown below residues that interact with the SMG1 arch (as shown in Fig. 2b and 2c).

Supplementary Figure 6



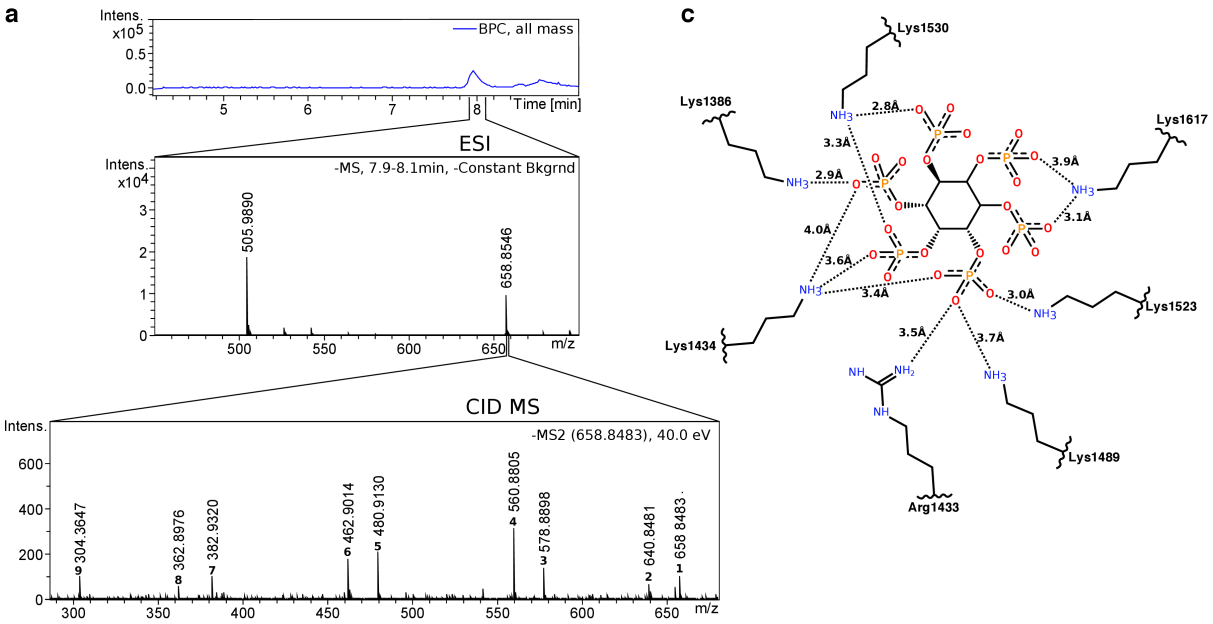
G-fold protein regulators of the cytoplasmic PIKK proteins, SMG1 and mTOR. Structures of SMG1 bound to the G-domain proteins, SMG8 and SMG9 **(a)**, and of mTOR bound to RHEB (PDB, code 6BCU) **(b)** are shown in a similar orientation after superimposition of the kinase domains of SMG1 and mTOR. Although the overall binding site is similar, there are major differences. First, RHEB is a bona fide GTPase, whereas the G-domain regulators of SMG9 binds ATP rather than GTP, and may well lack catalytic activity. Second, RHEB is positioned with its GTP domain roughly facing the catalytic cleft of mTOR, in an opposite orientation to that of SMG9. Unlike SMG9, RHEB binds to HEAT repeats proximal to the catalytic head ('bridge'), as well as the very N-terminal HEAT repeats of mTOR ('horn'). Significantly, the 'bridge' region of the mTOR HEAT repeats forms an extensive interface with the globular RHEB and largely replaces the equivalent position of SMG8 in the SMG1 structure.

Supplementary Fig. 7



SMG9 purifies bound to ATP. **(a)** Dimerization interface between the G-fold-like domains of SMG8 (cyan) and SMG9 (green), showing the electron density around the nucleotide and Mg ion bound in the SMG9 G-fold. **(b)** Coomassie-stained SDS gel of the human SMG8-SMG9 complex purified from HEK 293T cells. **(c)** Analysis of the purified SMG8-SMG9 dimer by ion-pair HPLC-mass spectrometry. Top: HPLC extracted-ion chromatogram (EIC) of SMG8-SMG9 dimer. Bottom: MS analysis of the peak fraction indicated above, with the mass corresponding to ATP labeled.

Supplementary Fig. 8

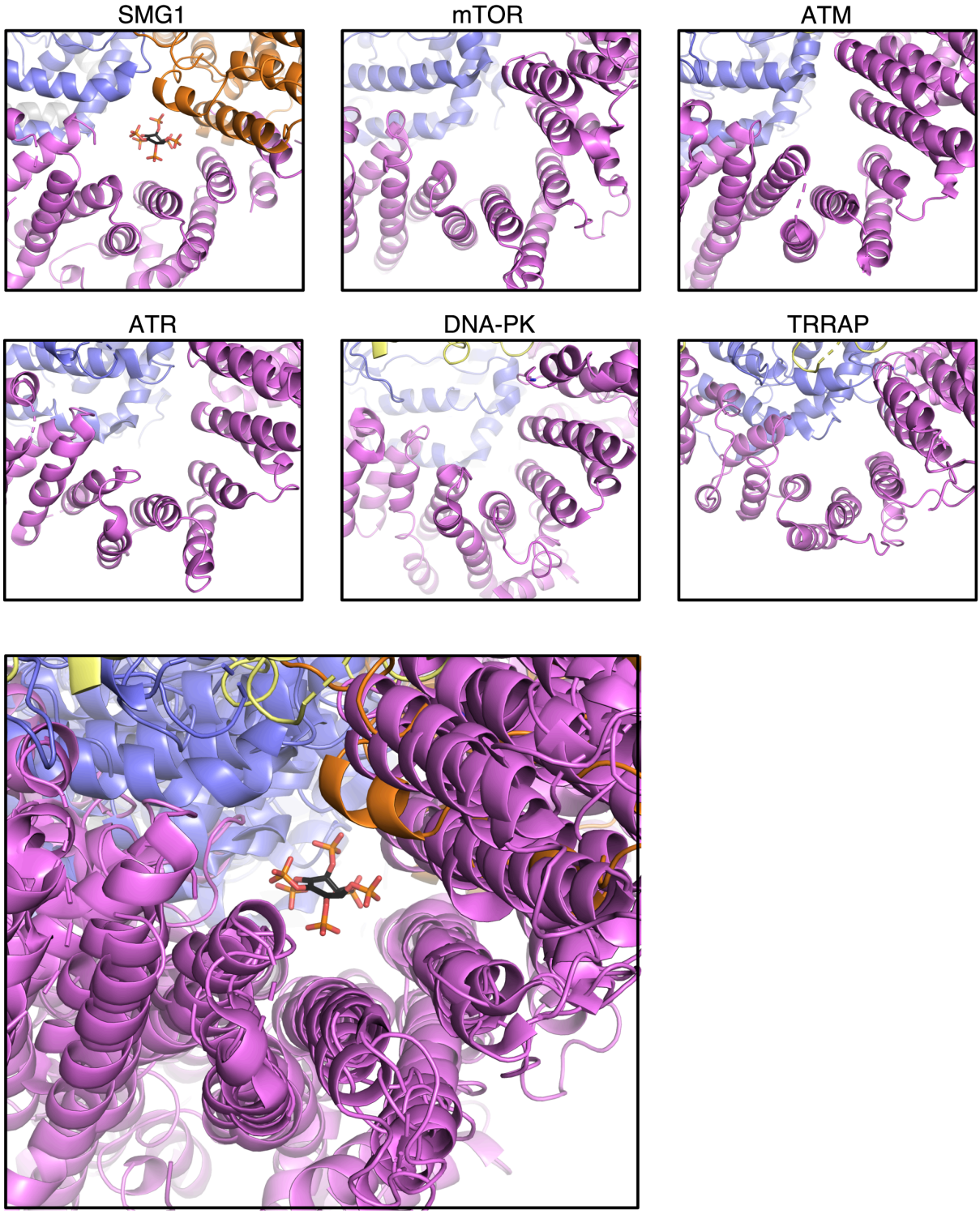


b

	Formula	composition	Expected mass	Observed mass
1.	$C_6H_6O_6(PO_3H_2)_6$, -H	$C_6H_{17}O_{24}P_6$	658.853	658.854
2.	$C_6H_6O_6(PO_3H_2)_6$, -H_2O, -H	$C_6H_{15}O_{23}P_6$	640.843	640.848
3.	$C_6H_7O_6(PO_3H_2)_5$, -H	$C_6H_{16}O_{21}P_5$	578.887	578.888
4.	$C_6H_7O_6(PO_3H_2)_5$, -H_2O, -H	$C_6H_{14}O_{20}P_5$	560.876	560.880
5.	$C_6H_8O_6(PO_3H_2)_4$, -H_2O, -H	$C_6H_{13}O_{17}P_4$	480.910	480.913
6.	$C_6H_8O_6(PO_3H_2)_4$, -2•(H_2O), -H	$C_6H_{11}O_{16}P_4$	462.899	462.901
7.	$C_6H_9O_6(PO_3H_2)_3$, -2•(H_2O), -H	$C_6H_{10}O_{13}P_3$	382.933	382.932
8.	$C_6H_9O_6(PO_3H_2)_3$, -3•(H_2O), -H	$C_6H_8O_{12}P_3$	364.922	362.897
9.	$C_6H_{10}O_6(PO_3H_2)_2$, -2•(H_2O), -H	$C_6H_9O_{10}P_2$	302.967	304.364

Validation of IP_6 identification and interactions. **(a)** The SMG1 complex was analyzed by ion pair HPLC-MS (top panel). To validate the identity of the 658.85 Da mass, this ligand was fragmented and the daughter ions analyzed by a second MS (bottom panel). As expected, the masses of the fragmented ions reflect a loss of individual groups of phosphate and water. **(b)** The table shows the chemical formula, the composition and the monoisotopic mass (both observed and expected) of each daughter ion. **(c)** Diagram of the chemical environment of IP_6 bound to SMG1, showing the distances between the IP6 phosphate oxygens and the amino groups of SMG1 lysine and arginine side chains measured from the atomic model. We note, however, that the precision of these measurements is limited by the resolution of the electron density map.

Supplementary Figure 9



Structural comparison of the FAT domains of SMG1 with that of the other PIKK family members. In this figure, all structures are shown in the same orientation after optimal manual superposition of their FAT domains (FAT in violet, catalytic domain in slate, N-terminal HEAT in yellow). mTOR (mammalian target of rapamycin) is from the PDB, code 4JSV, ATM (ataxia telangiectasia-mutated) is from the PDB, code 5MP0, ATR (ataxia- and Rad3-

related) is from the PDB, code 5X6O, DNA-PK (DNA protein kinase) is from the PDB, code 5LUQ, and the related protein TRRAP (transformation/transcription domain-associated protein) is from the PDB, code 5OJS. Despite poor structural conservation of the position and orientation of individual helices, the overall cleft-like feature that in SMG1 binds IP_6 appears to be conserved within the PIKK family. Comparing the contact potential of each protein (not shown) we observe that the surface of this cleft is distinctly positive in SMG1, mTOR, TRRAP and ATR. The contact potential of DNA-PK is not as positive, and ATM was not included in this comparison due to the low resolution of this region. Shown in the lower panels are zoom-in views at the IP_6-binding site for SMG1 and mTOR.

Supplementary Table 1.

Cryo-EM data collection, refinement, and validation statistics.

	SMG189 high res (EMDB-10347) (PDB 6SYT)	SMG189 (EMDB-10348)
Data collection and processing		
Magnification	135,000	135,000
Voltage (kV)	300	300
Electron exposure (e–/Å^2)	52.8	52.8
Defocus range (μm)	0.5-3.5	0.5-3.5
Pixel size (Å)	1.06	1.06
Symmetry imposed	C1	C1
Initial particle images (no.)	849 831	849 831
Final particle images (no.)	214 254	128 074
Map resolution (Å)	3.45	3.57
FSC threshold	0.143	0.143
Refinement		
Model resolution (Å)	3.78	3.92
FSC threshold	0.5	0.5
Map sharpening *B* factor (Å^2)	-144	-55
Model composition		
Non-hydrogen atoms	18370	
Protein residues	2582	
Ligands	IP_6 (1)	
	ATP (1)	
	Mg (1)	
B factors (Å^2)		
Protein	54.74	
Ligand	41.01	
RMS deviations		
Bond lengths (Å)	0.018	
Bond angles (°)	1.243	
Validation		
MolProbity score	1.54	
Clashscore	3	
Poor rotamers (%)	0.13	
Ramachandran plot		
Favored (%)	92.93	
Allowed (%)	6.99	
Disallowed (%)	0.08	

METHODS

Protein expression and purification

HEK 293T cell lines stably expressing Twin-Strep®-tagged SMG1, SMG8, and SMG9 were generated using the PiggyBac transposon system[21,22]. Cells were grown to a density of 1 × 10^6 cells/ml in FreeStyle 293 Expression Medium (Gibco, Thermo Fisher Scientific, Waltham, MA, USA) and induced using 1 µg/ml doxycycline for 72 h at 37˚C. Cells were resuspended in phosphate-buffered saline (PBS), containing 1 mM $MgCl_2$, 1 mM dithiothreitol (DTT), and a Complete Protease Inhibitor Cocktail Tablet (Roche, Basel, Switzerland). They were then lysed using a Dounce homogenizer and clarified by centrifugation. The supernatant was loaded onto a StrepTrap column (GE Healthcare, Chicago, IL, USA) and eluted with PBS and 5 mM desthiobiotin. The protein-containing fraction was further purified by size-exclusion chromatography using a Superose 6 Increase 10/300 GL column (GE Healthcare) in PBS, containing 1 mM $MgCl_2$ and 1 mM DTT. SMG1-SMG8-SMG9-containing fractions were pooled and concentrated using centrifugal concentrators (Sartorius, Göttingen, Germany). This protocol was used for wild-type (WT) SMG1, as well as for a catalytically inactive mutant (engineered for structural studies to prevent possible auto-phosphorylation and heterogeneity) and the IP_6-binding mutants.

The human SMG8-SMG9 dimer was purified using full length HA-tagged SMG8 and HA-flag tagged SMG9 expressed using the same system described above. HEK 293T cells were grown to 1 million cells/ml in 300 ml of FreeStyle 293 Expression Medium (Gibco, Thermo Fisher Scientific, Waltham, MA, USA). The cells were transfected using a 4:1 ratio of SMG8:SMG9 using PEI as the transfection reagent and grown for 72 h at 37˚C. Cells were washed with PBS and resuspended in PBS, 0.5% TRITON X-100 and Complete Protease Inhibitor Cocktail Tablet (Roche, Basel, Switzerland). Cells were lysed using a Dounce homogenizer and clarified by centrifugation. The supernatant was then incubated with ANTI-FLAG® M2 beads for 1 hour at 4˚C. The beads were washed with PBS 0.5% TRITON X-100 and finally eluted with the addition of 0.15 µg/µl 3X FLAG peptide incubated for 1 hour at 4˚C. The eluted protein was further purified by gel filtration or directly applied to a centrifugal concentrator to exchange the buffer to PBS and increase its concentration for MS analysis.

The SMG1 substrate, UPF1, was expressed in bacterial cells and purified using protocols similar to those previously reported[23]. Briefly, a plasmid expressing TEV-cleavable 6X-Histidine (His)-tagged UPF1 was transformed into competent BL21(DE3) *Escherichia coli* cells. Bacterial cultures were grown in Terrific Broth (TB) medium and induced with 0.5 mM IPTG at 18˚C overnight. Cells were harvested by centrifugation, resuspended in buffer containing, 20 mM Tris (pH 7.5), 500 mM NaCl, 10% glycerol, 2 mM $MgCl_2$, 1 mM β-mercaptoethanol, and 0.1% NP-40, and then lysed by sonication. The lysate was clarified by centrifugation, loaded onto a HisTrap HP Ni column (GE Healthcare), washed, and eluted with 300-mM imidazole. Eluted protein was incubated with TEV protease while being dialyzed against buffer containing, 20 mM Tris (pH 7.5), 100 mM NaCl, 10% glycerol, 1 mM $MgCl_2$, and 3 mM DTT. Protein was reloaded onto a second HisTrap HP Ni column and collected from the flow through. The protein was then loaded onto a heparin column (GE Healthcare), equilibrated with buffer containing 20 mM Tris (pH 7.5), 85 mM KCl, 10% glycerol, 1 mM $MgCl_2$, and 1 mM DTT, and eluted with a linear gradient of 20 mM Tris (pH 7.5), 1000 mM KCl, 10% glycerol, 1 mM $MgCl_2$, and 1 mM DTT. The UPF1-containing peak was loaded onto a Superdex 200 Increase 10/300 GL column (GE Healthcare) for size-exclusion chromatography in buffer containing, 20 mM Tris (pH 7.5), 85 mM KCl, 10% glycerol, 1 mM DTT, and 1-mM $MgCl_2$.

For mTOR expression and purification, we chose a complex between mTOR (1376-2549) and LST8 ($mTOR^{\Delta N}$) that had been previously characterized to be stable and active[17]. Stable cell lines expressing Strep-tagged WT and mutant $mTOR^{\Delta N}$ were made using the PiggyBac transposon system[21,22] . Cells were grown to $1x10^6$ cells/ml and induced with 1 μg/ml doxycycline in the presence of 3.75 mM valproic acid for 72 hours. Cells were resuspended in PBS, 10% glycerol, 1 mM EGTA, 1 mM EDTA, 2 mM DTT and then lysed by sonication and clarified by centrifugation. The lysate was loaded onto a StrepTrap column (GE Healthcare, Chicago, IL, USA), and eluted with 5mM desthiobiotin in lysis buffer. The $mTOR^{\Delta N}$ complex was further purified by size exclusion in PBS 2 mM DTT on a Superdex200 Increase 10/300 column (GE Healthcare, Chicago, IL, USA). As mTOR substrate, we expressed a vector encoding GST-AKT1 (450-480) in E. coli BL-21 pLysS gold. Cells were lysed by sonication in 50 mM Tris pH 7.5, 500 mM NaCl, 1 mM DTT. After centrifugation, the supernatant was loaded on a glutathione Sepharose column and eluted

with 10 mM reduced glutathione. The protein was further purified over a Superdex75 16/60 (GE healthcare) in PBS, 1 mM DTT.

Cryo-EM data collection and processing

Cryogenic electron microscopy (cryo-EM) grids were prepared by applying 4 µl of SMG1-SMG8-SMG9 complex at 0.5 mg/ml, in the presence of 0.05% octyl-β-glucoside, to glow-discharged R2/1 200-mesh holey carbon grids (QUANTIFOIL, GroBlöbichau, Germany). Grids were blotted for 3.5 sec at approximately 95% humidity and 4°C and then plunge-frozen into liquid ethane and cooled by liquid nitrogen using a Vitrobot Mark IV (FEI, Hillsboro, OR, USA). A total of 10034 micrographs were collected on a Titan Krios electron microscope (FEI), operated at 300 kV, and equipped with a post-GIF K2 Summit direct electron detector (Gatan, Pleasanton, CA, USA). Images were acquired using the SerialEM software suite were used for automated acquisition, with defocus values ranging from 0.5 µm to 3.5 µm[24]. Images were acquired as 40-frame movies in electron counting mode (pixel size: 1.06 Å per pixel), using a total specimen dose of 52.8 $e^{-}/Å^2$. The dose-fractionated movies were gain-normalized, aligned, and dose-weighted using MotionCor2[25]. Defocus values were estimated using GCTF[26], and 1,569,697 particles were automatically chosen using Gautomatch software (https://www.mrc-lmb.cam.ac.uk/kzhang/Gautomatch/). Subsequent image processing steps were carried out in RELION 2.1[27] and RELION 3.0[28].

Particle sorting and reference-free 2D classification was performed to remove non-particle candidates and damaged particles, resulting in a total of 849,831 particles. An *ab initio* model was generated using the stochastic gradient descent (SGD) algorithm[28]. The resulting *ab initio* model was low-pass filtered to 60 Å, thereby preventing any model bias using a completely data driven starting model. This model was subjected to 3D classification using 5 classes. The two best aligning classes, a total of 214,254 particles, were combined and subsequently subjected to 3D refinement, yielding an overall resolution of 3.45 Å. Global resolution and B-factor (-144 $Å^2$) of the map were estimated by applying a soft mask around the protein density, using the gold standard Fourier shell correlation (FSC) = 0.143 criterion. This map was then used for *de novo* model building.

During processing, it was noted that there is a sub-set of particles carrying a C-terminal helical bundle on the SMG8 protein that was previously described in the *Caenorhabditis elegans* X-ray structure[14]. The final polished particles were classified into three classes using heterogeneous refinement in cryoSPARC[29]. A single class that carried this C-terminal extension was refined to 3.57 Å using the homogenous refinement routine, and this map was only used to display and model the C-terminus of SMG8.

Model building and validation

The atomic model of the SMG1-SMG8-SMG9 complex was refined against the cryo-EM map using the phenix.real_space_refine routine in the PHENIX software package[30]. The final model was probed for over fitting by randomly displacing the atoms of the structural model by 0.5 Å and subsequent refinement against the first of the two independent half-maps. The second half-map was not used for refinement, and the model vs. map FSC curve shows excellent agreement. Directional FSC curves and map anisotropy were assessed using the 3DFSC (Supplementary Fig. 2f)[31]. Statistical quality of the final model was assessed using the program Molprobity[32] (Table S1, Supplementary Fig. 4), and figures were prepared using PyMOL and UCSF ChimeraX.

***In-vitro* kinase assays**

Kinase activity assays were performed in 10-µl reactions, containing 1 µM human UPF1 and 50 nM of either WT-SMG1-SMG8-SMG9 complex or a version containing the K1530E mutation at the IP_6-binding site of SMG1. Reaction buffer consisted of PBS, 10 mM $MgCl_2$, and 1 mM DTT. Phosphorylation reactions were initiated by the addition of a radiolabeled ATP mix, consisting of 0.5 mM ATP and 0.8 µCi [γ-^{32}P]-ATP, and allowed to proceed for 30 min at 30˚C. Reactions were quenched by the addition of sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) loading buffer, and samples were analyzed by 10% SDS-PAGE. After Coomassie staining, phosphoproteins were visualized by autoradiography using a Typhoon FLA7000 imager. *In-vitro* kinase assays of mTOR$^{\Delta N}$ were essentially identical to those of SMG1-SMG8-SMG9, with the exception that we used 100 nM of each mTOR$^{\Delta N}$ variant and 1 µM GST-AKT1 450-480 as substrate.

Mass spectrometry analysis

The IP_6 molecule was resolved and detected using protocols adapted from a previous study[33]. In brief, 2 µM of the purified SMG1-SMG8-SMG9 complex were injected onto a C18 column (Agilent Zorbax Eclipse plus C18, 2.1x100 mm, pore size 3.5 µm), mounted on an Agilent 1290 High-Performance Liquid Chromatography (HPLC) apparatus. Bound ligands were eluted at 200 µl/min with the following gradient profile: 10% Buffer B for 3 min, 25% Buffer B for 4 min, and 100% Buffer B for 6 min (Buffer A: 20-mM triethylammonium acetate [pH 9.0], and Buffer B: 20 mM triethylammonium acetate [pH 9.0], 20% acetonitrile), in MS2-negative mode, a mass range of 200–1000 m/z, and collision cell energy of -40.0 eV. This experiment was conducted at pH 9.0, with the MS operating in negative ion mode to accommodate detection of the negatively charged IP_6, and therefore, the masses of both IP_6 and ATP (Fig. 5b) are off by -1 Da. MS analysis of $mTOR^{\Delta N}$ was essentially identical to that described for SMG1-SMG8-SMG9.

Supplementary references

21. Yusa, K., Zhou, L., Li, M. A., Bradley, A. & Craig, N. L. *Proc Natl Acad Sci USA* **108,** 1531–1536 (2011).
22. Li, X. *et al. Proc Natl Acad Sci USA* **110,** E2279–87 (2013).
23. Chakrabarti, S., Bonneau, F., Schüssler, S., Eppinger, E. & Conti, E. *Nucleic Acids Res* **42,** 9447–9460 (2014).
24. Mastronarde, D.N. *J. Struct. Biol.* **152,** 36-51 (2005).
25. Zheng, S. Q. *et al. Nat Methods* **14,** 331–332 (2017).
26. Zhang, K. *Journal of structural biology* **193,** 1–12 (2016).
27. Kimanius, D., Forsberg, B. O., Scheres, S. H. & Lindahl, E. *Elife* **5,** e18722 (2016).
28. Zivanov, J. *et al. Elife* **7,** (2018).
29. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. *Nat Methods* **14,** 290–296 (2017).
30. Adams, P. D. *et al. Acta Crystallogr D Biol Crystallogr* **66,** 213–221 (2010).
31. Tan, Y. Z. *et al. Nat Methods* **14,** 793–796 (2017).
32. Chen, V. B. *et al. Acta Crystallogr D Biol Crystallogr* **66,** 12–21 (2010).
33. Tur, F., Tur, E., Lentheric, E., Mendoza, P., Encabo, M., Isern, B., Grases, F., Maraschiello. C. & Perello, J. *J. Chromatography B* **928,** 146-154 (2013).

4.0 EXTENDED DISCUSSION

4.1 PART ONE – MTR4-EXOSOME ADAPTOR INTERACTIONS

4.1.1 SHORT LINEAR MOTIFS (SLiMS) AS STRUCTURALLY FLEXIBLE BINDING MODULES

The cell faces a unique challenge to be able to recognize and target a variety of substrates to the nuclear MTR4-exosome in a timely and accurate fashion. The MTR4-exosome, being a central component of the RNA degradation and quality control machinery, is an inflexible evolutionary target to modulate recognition of a variety of substrates. Thus, this machinery relies on adaptor proteins with a modular architecture to recognize RNA substrates. Research performed on MTR4-exosome adaptor interactions as part of this thesis and prior literature indicates that these interactions are mediated by short linear motifs (SLiMs).

SLiMs are short stretches of adjacent amino acids located usually in the disordered regions of proteins allowing them to interact with a partner protein or a ligand (Dinkel et al., 2014). Examination of SLiMs at a sequence level along with their partner proteins indicates that SLiMs evolve in a convergent manner, which is distinct from how globular protein interaction modules evolve (Davey et al., 2012; Diella et al., 2008). Furthermore, SLiMs are comparatively short, allowing sampling of the binding space in a relatively short amount of evolutionary time (Davey et al., 2012). The motifs bind with relatively low affinities making them indispensible in mediating transient interactions. These evolutionarily flexible features of SLiMs make them ideal for being selected as 'access points' that link the MTR4-exosome to its substrates (Fig. 4.1).

SLiMs are tailored to interact with their respective binding surfaces and thus are very difficult to predict with standard computational and bioinformatics approaches (van der Lee et al., 2014). However, they can be identified based on the common features of several validated SLiMs. Some of the common features of SLiMs include distinct lack of bulky hydrophobic residues, disorder to order molecular recognition features and specific prediction flavors such as their tendency to be subjected to post translational modifications (van der Lee et al., 2014). SLiMs and intrinsically disordered regions also possess distinct biophysical properties like their tendency to form a phase transition *in vitro* because of the unique

distribution of charges across the sequence. Despite the wealth of research regarding SLiMs, they still remain very difficult to identify with appreciable accuracy.

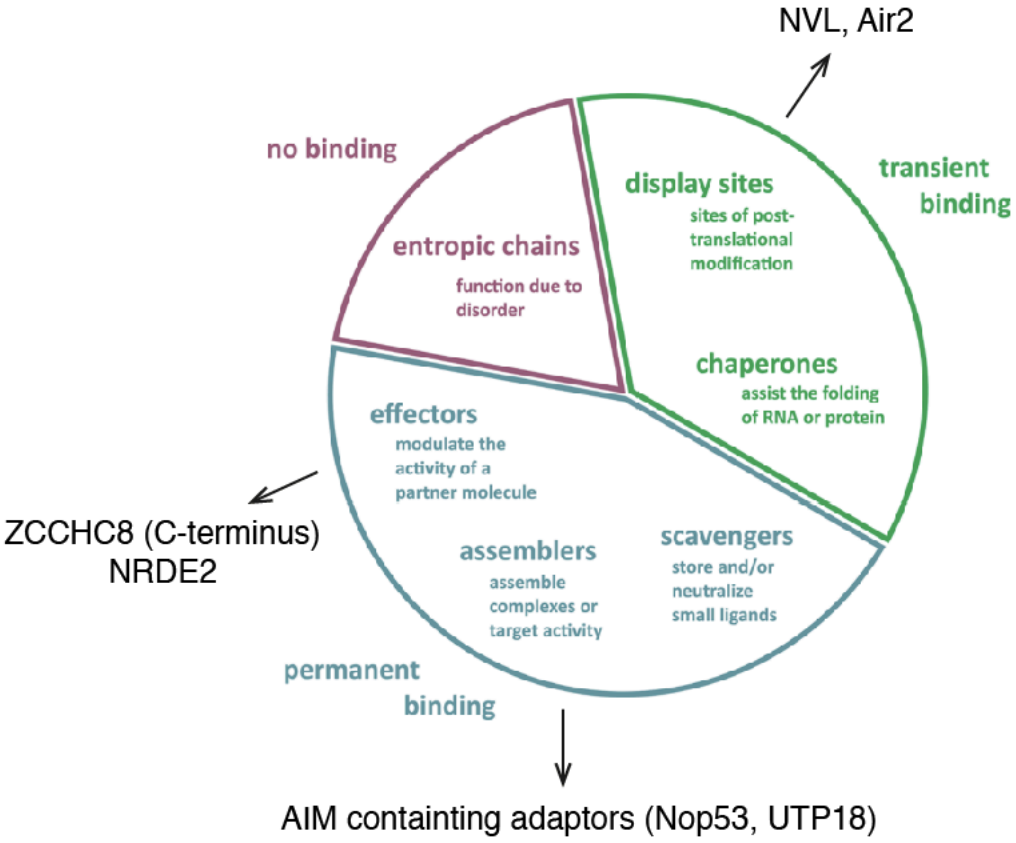


Figure 4.1: Functional characteristics of short linear motifs and intrinsically disordered regions.

The figure shows the functional characteristics of short linear motifs and intrinsically disordered regions. The area of each sector represents the relative abundance of the binding mode in all of the motifs identified until 2014. The MTR4 adaptors are listed and are associated with their properties indicating the versatility of the factors. The figure has been adapted from (van der Lee et al., 2014).

The key features of the SLiM required to target the MTR4 KOW region can be identified based on the work performed as part of this thesis and prior literature. Structure based mutagenesis on the identified regions allowed us to refine the consensus of the SLiM to xωxxD(x)$_{1/2}$G/P. Although the motif is quite flexible in terms of the sequence, the combination of an aspartate residue and a predominantly hydrophobic nature of the motif seems indispensible. Examination of the interaction surface indicates that a network of

hydrophobic stacking interactions and a single salt bridge between MTR4 Arg743 and the SLiM Asp residue stabilize the complex. Mutagenesis indicates that both the charged and the hydrophobic residues are individually important, and mutations that drastically change the properties of the individual amino acids severely impact the ability of the motif to bind the interacting surface. The structural data also indicate that the sequence tends to organize itself as a β−sheet on the surface of the MTR4-KOW region making the presence of a glycine or proline important for the positioning of the motif. The presence of key hydrophobic residues mediating the interaction makes the motif difficult to predict computationally as bulky hydrophobic residues are a rare feature among SLiMs.

4.1.2 MODULATION OF BINDING AFFINITIES IN SLiMS

Since MTR4 needs to interact with several substrate-presenting adaptors, the modulation of binding affinity becomes paramount to allow MTR4 to act in a target specific manner. MTR4 is localized throughout the nucleus, whereas the exosome adaptors are localized to specific compartments where their substrates are likely present. The adaptors examined in this study, Nop53 and NVL, are localized in the nucleolus, whereas ZCCHC8 is localized to the nucleoplasm (Ogami et al., 2018). Even though, MTR4 is likely present as a component in several distinct complexes, it would be beneficial for the cell to utilize the available pool of MTR4 protein molecules as efficiently as possible.

Examination of the structures of several AIMs bound to MTR4 reveals that the affinities are possibly modulated by a hydrophobic effect achieved by modulation of the hydrophobicity of the motif. Biophysical analysis of the NVL and Nop53 interactions reveals nearly an order of magnitude of difference in binding affinities rationalized by the characteristics of the hydrophobic residues. However, as in the case of ZCCHC8 (Puno & Lima, 2018) and NRDE-2 (J. Wang et al., 2019), the binding affinities can also be modulated by the presence of auxiliary features that aid in the interaction of the adaptor protein along with the broadly conserved SLiM. Taken together, it is likely that MTR4 is directed passively toward the substrates as needed due to the motifs on different exosome adaptors competing with each other to bind MTR4. A competition pull-down experiment examining the MTR4 binding factors ZCCHC8 and ZFC3H1, which are localized in the nucleoplasm, suggests indeed that there might be functionally relevant competition (Fig. 4.2).

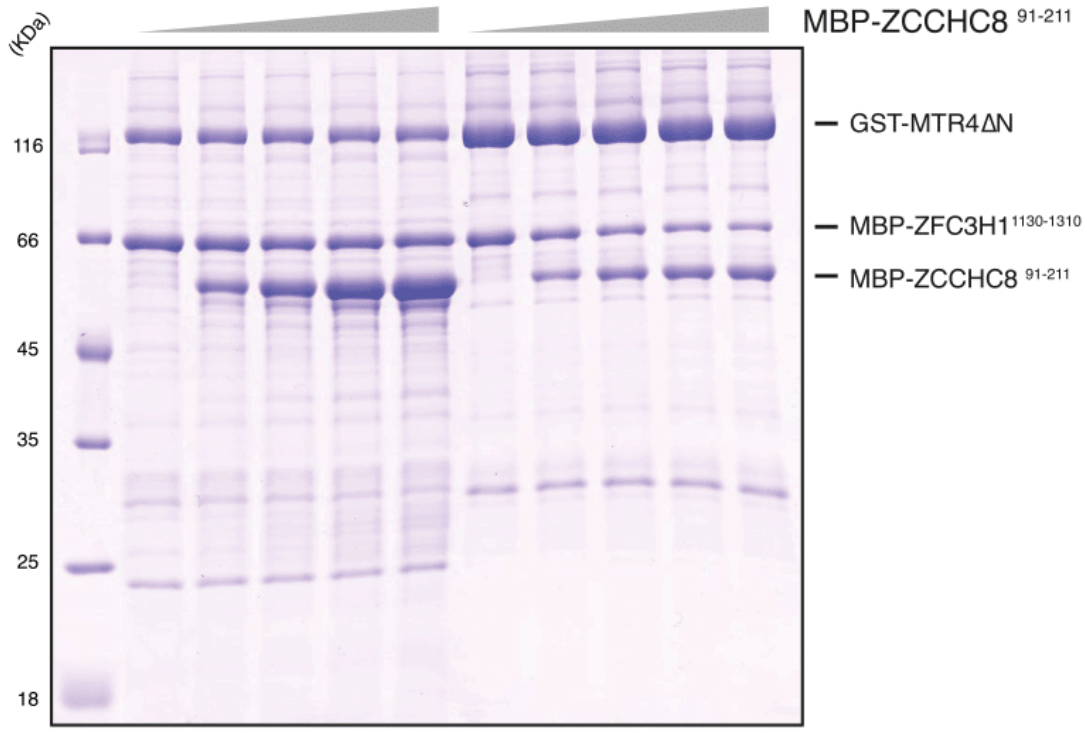


Figure 4.2: A GST pulldown competition assay showing the competition between the arch interacting ZCCHC8 and ZFC3H1.

A preformed GST-MTR4ΔN-MBP- ZFC3H1$^{1130-1310}$ was incubated with increasing concentrations of MBP- ZCCHC8$^{91-211}$ before co-precipitation with glutathione sepharose beads. A total of 3% of the input (lanes on the left hand side) and 30% of the eluates (lanes on the right hand side) were analyzed on 15% SDS-PAGE gels and visualized by staining with Coomassie brilliant blue (unpublished results).

Post translational modifications (PTMs) are yet another way SLiM-based interactions are regulated. Although there is no direct evidence of PTMs regulating the interactions centered around MTR4, examination of the Trf4-Air2-Mtr4 (TRAMP) (Falk et al., 2014) and NVL-MTR4 structures indicate that PTMs might very well play a role in modulating these interactions. In the TRAMP complex, the conserved arginine of Mtr4 interacts with a sulfate ion coordinated by a threonine from Air2 mimicking a phosphorylation event. We observe a similar situation for the interface between NVL and the MTR4 KOW domain (Fig. 4.3). Interestingly, the residues coordinating the sulfate ion are conserved and have been reported

to be phosphorylated indicating that interactions centered on MTR4 are possibly modulated by PTMs.

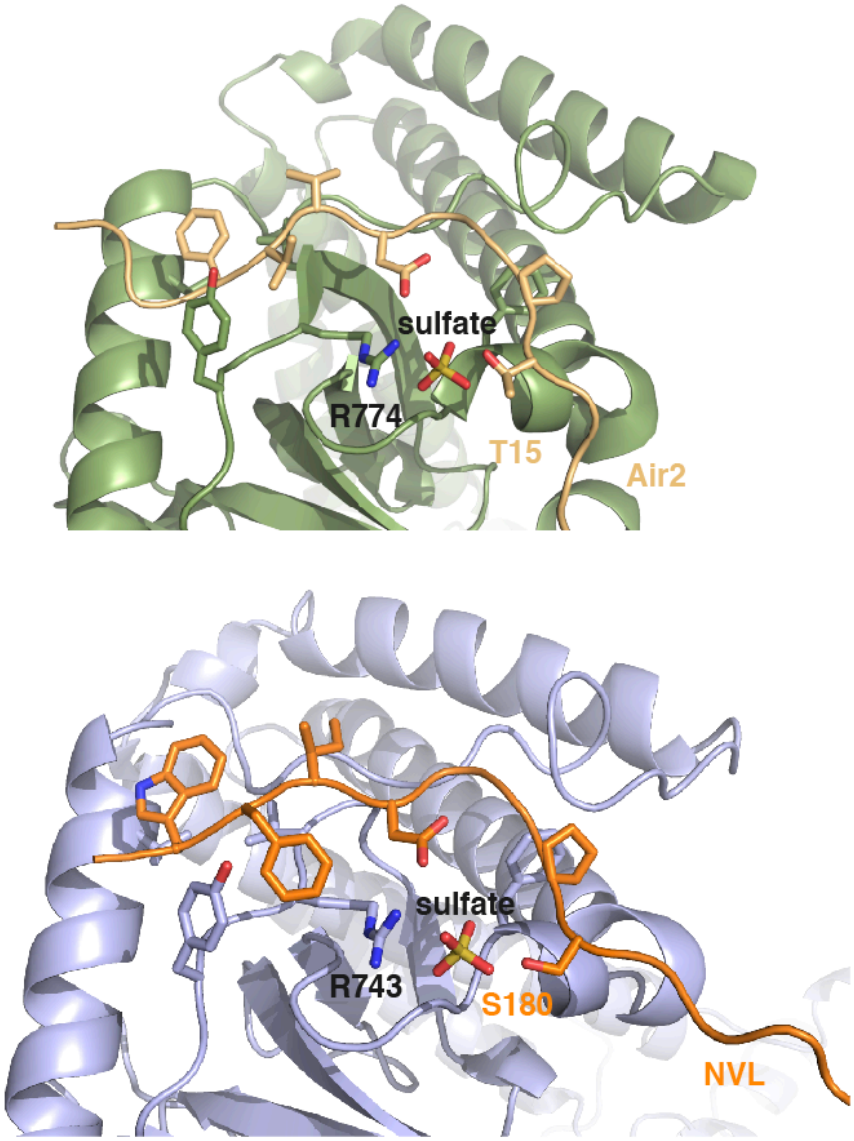


Figure 4.3: Zoom-in view of KOW-AIM interfaces in NVL-MTR4 (PDB 6RO1) and Air2-Mtr4 (PDB 4U4C) structures.

KOW domains of yeast Mtr4 and human MTR4 are colored in light green (upper panel) and light blue (lower panel), respectively. Air2 and NVL are colored in pale orange and orange, respectively. The representation shows a possible role of phosphorylation in modulating the interaction.

Thus, AIMs in the exosome adaptors seem to employ a combination of features that could result in both active and passive regulation of binding affinities.

4.2 PART TWO – STRUCTURAL ANALYSIS OF SMG1-SMG8-SMG9 COMPLEX

4.2.1 EXAMINATION OF THE STRUCTURE ACTIVITY HYPOTHESES OF SMG1 IN THE LITERATURE

Initial *in vitro* functional analysis of SMG1 indicated that the N-terminal HEAT repeat region and the C-terminal FATC domain are crucial for the kinase activity of the protein (Morita et al., 2007). This finding is in line with biochemical analysis performed on other PIKK proteins where minor changes or deletion of either the FATC domain or HEAT repeat region lead to drastic changes in the kinase activity (Beamish et al., 2000; Priestley et al., 1998). The structure of SMG1 rationalizes these hypotheses to a certain extent. The structure reveals that the hydrophobic C-terminal part of the FATC domain that is distinctly absent in other PI3K structures (Miller et al., 2010; Walker et al., 1999), packs into a loop of the kinase domain, indicating that it might be crucial for the structural integrity of the kinase domain. Upon structural alignment with mTOR, organization of the C-terminal portion of FATC and the loop regions of the kinase domain resemble how FATC of mTOR stabilizes the mTOR activation loop. This stabilization effect might be one of the possible allosteric mechanisms by which the FATC domain influences the activity of SMG1. However, the structure does not rationalize why the L3646A mutation (Morita et al., 2007) in the FATC domain dramatically affects the kinase activity, as this residue is present at the surface and does not appear to be involved in any van der Waals contacts. The structure also indicates that a portion of the FATC domain of SMG1 is solvent exposed and is therefore available to mediate protein-protein interactions like the FATC domains of ATM, ATR and TRRAP kinase where it is believed to mediate interactions with Tip60 histone acetyltransferase (Jiang et al., 2006). Given that several residues of the FATC domain are conserved among several PIKKs, it can be envisioned that the FATC of SMG1 might also be involved in a similar protein-protein interaction role, although the interaction might not be necessary for activation of the kinase.

Activity assays performed to examine the effects of N-terminal deletions indicated that the region is indispensable for the activity of SMG1 (Morita et al., 2007). Low-resolution cryo-EM reconstructions of the SMG1-SMG8-SMG9 complex revealed the role of the N-terminal HEAT repeat region of SMG1 as a scaffold to recruit SMG8-SMG9 (Arias-Palomo et al., 2011). The N-terminal HEAT repeat deletions that compromise the kinase activity of SMG1

coincide with the regions that form surfaces for SMG8-SMG9 interaction. This finding along with the fact that SMG1 does not require SMG8-SMG9 for kinase activity indicate that the HEAT repeat region might be involved in inducing and stabilizing some long range conformational changes, priming the kinase domain for efficient activity. A more detailed understanding would require high-resolution structural analysis of SMG1 in isolation in order to examine changes in the overall architecture of the protein with a specific focus on the conformation of the kinase domain.

While previous studies attempting to understand SMG1 activity are valuable, it is well known in the field that sample quality has been the limiting factor in structural and biochemical characterization of SMG1. Therefore, it is crucial to interpret the structural-activity data in the literature with care. In our structural study, we overcame the limitation by establishing a mammalian expression system stably expresses SMG1-SMG8-SMG9 complex, which enabled us to purify a well-behaved sample of SMG1-SMG8-SMG9 (See Methods in 3.3).

4.2.2 Inositol hexaphosphate (IP_6) as a structural co-factor of PIKKs

IP_6 is an abundant metabolite in eukaryotic cells (Monserrate & York, 2010). IP_6 and its precursors have been shown to be structural co-factors of other enzymes involved in RNA regulation like ADAR2 (Macbeth et al., 2005) and the Dbp5 complex (Montpetit et al., 2011). In all the previous reported structures, IP_6 functions as a tether bringing two individual domains or proteins together to help stabilize a particular conformation. In SMG1 and mTOR, however, IP_6 helps package the spine region of the FAT domain onto the kinase domain, perhaps stabilizing a certain kinase domain conformation. The structural role played by IP_6 in SMG1 and mTOR in packaging the helical repeats is also observed in other proteins. In the cohesion dynamics controlling factor Pds5, for example, IP_6 restricts the conformational freedom of the helical repeat region (Fig. 4.4), thus enabling it to bind cohesin (Ouyang et al., 2016).

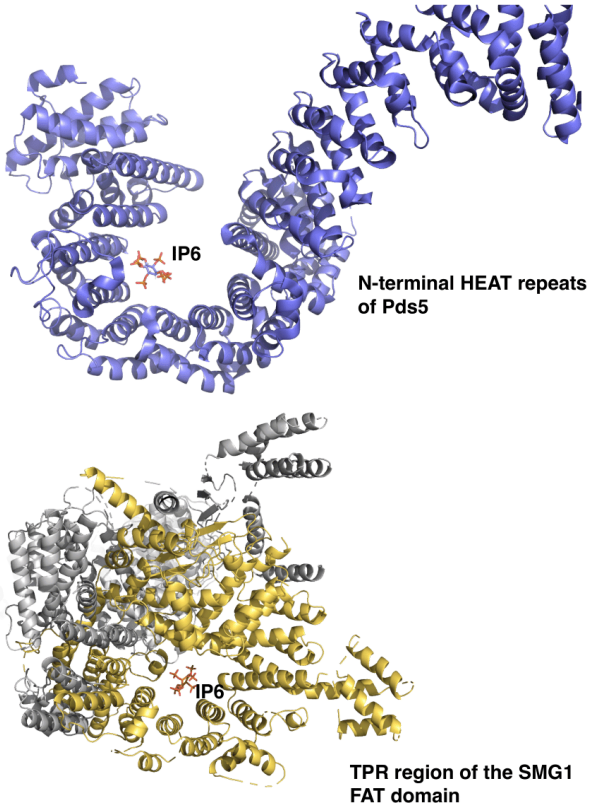


Figure 4.4: Structures of Pds5 and SMG1 showing the similar ways in which IP_6 restricts the helical regions.

A comparison of the structures of Pds5 (PDB 5HDT) with helical repeats colored in blue (upper panel) and SMG1 with regions stabilized by IP6 colored in yellow and the rest of the protein in gray. In both structures, the helical region wraps around IP6 thus stabilizing the domain which seems to be crucial for protein function.

5.0 OUTLOOK

5.1 MTR4-EXOSOME ADAPTOR INTERACTIONS

Recent studies regarding MTR4 interactions along with the results presented in thesis shed significant light in our journey toward a comprehensive mechanistic understanding of the nuclear exosome. Yet, a lot remains to be understood regarding the individual mechanisms of the complexes, both structurally and biochemically. Structural studies on larger MTR4 containing assemblies provide the basis for how a substrate itself might be targeted to the exosome as demonstrated by the cryo-EM analysis of the pre-ribosome bound to the exosome (Schuller et al., 2018). Functional analysis of the complexes sheds new light on the interplay between MTR4 containing complexes and other related cellular pathways. Functional analysis and knock-down studies could also pave the way to define new exosome adaptors in a manner akin to the experiments that led to the identification of the poly(A) exosome targeting connection (Meola & Jensen, 2017).

An interesting question that arises as a result of the identification of several MTR4 centered interactions is how these complexes are regulated. Findings related to the MTR4-NRDE-2 interaction point to the fact that a portion of unbound MTR4 might be sequestered from the exosome-interacting pool of MTR4 thereby limiting the productive substrate-MTR4-exosome interaction (J. Wang et al., 2019). However, it is unclear whether this phenomenon truly represents a sequestering effect because NRDE-2 is capable of stimulating MTR4 and improving its RNA binding abilities (J. Wang et al., 2019). These findings also lead to several questions centered around the activity of MTR4 itself – why is it that some exosome adaptors, like ZCCHC8 and NRDE-2 stimulate MTR4 while factors like NVL and NOP53 do not seem to. Differential regulation of MTR4 activity and RNA binding ability indicate that some adaptors might have evolved to provide MTR4 with abilities to recognize and unwind far more complex substrates and to allow an additional layer of regulation of MTR4 activity.

5.2 STRUCTURAL CHARACTERIZATION OF THE SMG1-SMG8-SMG9 COMPLEX

The structure of SMG1-SMG8-SMG9 shows remarkable similarity to other phosphoinositide 3-kinase related kinases, especially mTOR. At the same time, the complex is sufficiently unique and leaves several unanswered questions and unverified hypotheses. Although, the structure revealed the basis for the SMG8-SMG9 interaction with SMG1, it is unclear how they impact the activity of SMG1. Given that SMG1 seems to be a constitutively active kinase without the need for an activator, it was hypothesized and verified via *in vitro* kinase assays that SMG8-SMG9 could play an inhibitory role. This regulation seems to be in contrast with how Rheb regulates mTOR by interacting with the PIKK regulatory domain at the C-terminus of mTOR (H. Yang et al., 2017). While SMG1 has its own unique insertion domain toward the C-terminus, the fact that it is not ordered in the SMG1-SMG8-SMG9 complex, points to a more elusive role of this domain rather than a regulatory role involving SMG8 and SMG9. Furthermore, it is unclear how SMG1 recruits UPF1 and how the activity of the kinase varies based on the absence or presence of SMG8-SMG9. Further high-resolution structural studies would also shed light on the role of the DHX34 helicase in targeting unphosphorylated UPF1 to SMG1 and the role of UPF2 in inducing substrate release (Deniaud et al., 2015; Melero et al., 2016). High-resolution structural studies in conjunction with biochemical and functional analyses could contribute to obtain a mechanistic understanding of the initial steps of the nonsense-mediated decay pathway.

6.0 BIBLIOGRAPHY

Allmang, C., Kufel, J., Chanfreau, G., Mitchell, P., Petfalski, E., & Tollervey, D. (1999). Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *EMBO J, 18*(19), 5399-5410.

An, W., Du, Y., & Ye, K. (2018). Structural and functional analysis of Utp24, an endonuclease for processing 18S ribosomal RNA. *PLoS One, 13*(4), e0195723.

Anderson, J. S., & Parker, R. P. (1998). The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex. *EMBO J, 17*(5), 1497-1506.

Arias-Palomo, E., Yamashita, A., Fernandez, I. S., Nunez-Ramirez, R., Bamba, Y., Izumi, N., . . . Llorca, O. (2011). The nonsense-mediated mRNA decay SMG-1 kinase is regulated by large-scale conformational changes controlled by SMG-8. *Genes Dev, 25*(2), 153-164.

Arribere, J. A., & Fire, A. Z. (2018). Nonsense mRNA suppression via nonstop decay. *Elife, 7*.

Beamish, H. J., Jessberger, R., Riballo, E., Priestley, A., Blunt, T., Kysela, B., & Jeggo, P. A. (2000). The C-terminal conserved domain of DNA-PKcs, missing in the SCID mouse, is required for kinase activity. *Nucleic Acids Res, 28*(7), 1506-1513.

Becker, T., Armache, J. P., Jarasch, A., Anger, A. M., Villa, E., Sieber, H., . . . Beckmann, R. (2011). Structure of the no-go mRNA decay complex Dom34-Hbs1 bound to a stalled 80S ribosome. *Nat Struct Mol Biol, 18*(6), 715-720.

Bousquet-Antonelli, C., Presutti, C., & Tollervey, D. (2000). Identification of a regulated pathway for nuclear pre-mRNA turnover. *Cell, 102*(6), 765-775.

Bresson, S., & Tollervey, D. (2018). Surveillance-ready transcription: nuclear RNA decay as a default fate. *Open Biol, 8*(3).

Briggs, M. W., Burkard, K. T., & Butler, J. S. (1998). Rrp6p, the yeast homologue of the human PM-Scl 100-kDa autoantigen, is essential for efficient 5.8 S rRNA 3' end formation. *J Biol Chem, 273*(21), 13255-13263.

Buchwald, G., Ebert, J., Basquin, C., Sauliere, J., Jayachandran, U., Bono, F., . . . Conti, E. (2010). Insights into the recruitment of the NMD machinery from the crystal structure of a core EJC-UPF3b complex. *Proc Natl Acad Sci U S A, 107*(22), 10050-10055.

Buttner, K., Nehring, S., & Hopfner, K. P. (2007). Structural basis for DNA duplex separation by a superfamily-2 helicase. *Nat Struct Mol Biol, 14*(7), 647-652.

Carneiro, T., Carvalho, C., Braga, J., Rino, J., Milligan, L., Tollervey, D., & Carmo-Fonseca, M. (2007). Depletion of the yeast nuclear exosome subunit Rrp6 results in accumulation of polyadenylated RNAs in a discrete domain within the nucleolus. *Mol Cell Biol, 27*(11), 4157-4165.

Chakrabarti, S., Bonneau, F., Schussler, S., Eppinger, E., & Conti, E. (2014). Phospho-dependent and phospho-independent interactions of the helicase UPF1 with the NMD factors SMG5-SMG7 and SMG6. *Nucleic Acids Res, 42*(14), 9447-9460.

Chakrabarti, S., Jayachandran, U., Bonneau, F., Fiorini, F., Basquin, C., Domcke, S., . . . Conti, E. (2011). Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. *Mol Cell, 41*(6), 693-703.

Chanfreau, G., Rotondo, G., Legrain, P., & Jacquier, A. (1998). Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1. *EMBO J, 17*(13), 3726-3737.

Chang, Y. F., Imam, J. S., & Wilkinson, M. F. (2007). The nonsense-mediated decay RNA surveillance pathway. *Annu Rev Biochem, 76*, 51-74.

Chiu, S. Y., Serin, G., Ohara, O., & Maquat, L. E. (2003). Characterization of human Smg5/7a: a protein with similarities to Caenorhabditis elegans SMG5 and SMG7 that functions in the dephosphorylation of Upf1. *RNA, 9*(1), 77-87.

Clerici, M., Mourao, A., Gutsche, I., Gehring, N. H., Hentze, M. W., Kulozik, A., . . . Cusack, S. (2009). Unusual bipartite mode of interaction between the nonsense-mediated decay factors, UPF1 and UPF2. *EMBO J, 28*(15), 2293-2306.

Coller, J., & Parker, R. (2004). Eukaryotic mRNA decapping. *Annu Rev Biochem, 73*, 861-890.

Couvillion, M. T., Bounova, G., Purdom, E., Speed, T. P., & Collins, K. (2012). A Tetrahymena Piwi bound to mature tRNA 3' fragments activates the exonuclease Xrn2 for RNA processing in the nucleus. *Mol Cell, 48*(4), 509-520.

D'Orazio, K. N., Wu, C. C., Sinha, N., Loll-Krippleber, R., Brown, G. W., & Green, R. (2019). The endonuclease Cue2 cleaves mRNAs at stalled ribosomes during No Go Decay. *Elife, 8*.

Danin-Kreiselman, M., Lee, C. Y., & Chanfreau, G. (2003). RNAse III-mediated degradation of unspliced pre-mRNAs and lariat introns. *Mol Cell, 11*(5), 1279-1289.

Davey, N. E., Van Roey, K., Weatheritt, R. J., Toedt, G., Uyar, B., Altenberg, B., . . . Gibson, T. J. (2012). Attributes of short linear motifs. *Mol Biosyst, 8*(1), 268-281.

de la Cruz, J., Kressler, D., Tollervey, D., & Linder, P. (1998). Dob1p (Mtr4p) is a putative ATP-dependent RNA helicase required for the 3' end formation of 5.8S rRNA in Saccharomyces cerevisiae. *EMBO J, 17*(4), 1128-1140.

Dengl, S., & Cramer, P. (2009). Torpedo nuclease Rat1 is insufficient to terminate RNA polymerase II in vitro. *J Biol Chem, 284*(32), 21270-21279.

Deniaud, A., Karuppasamy, M., Bock, T., Masiulis, S., Huard, K., Garzoni, F., . . . Schaffitzel, C. (2015). A network of SMG-8, SMG-9 and SMG-1 C-terminal insertion domain regulates UPF1 substrate recruitment and phosphorylation. *Nucleic Acids Res, 43*(15), 7600-7611.

Diella, F., Haslam, N., Chica, C., Budd, A., Michael, S., Brown, N. P., . . . Gibson, T. J. (2008). Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Front Biosci, 13*, 6580-6603.

Dinkel, H., Van Roey, K., Michael, S., Davey, N. E., Weatheritt, R. J., Born, D., . . . Gibson, T. J. (2014). The eukaryotic linear motif resource ELM: 10 years and counting. *Nucleic Acids Res, 42*(Database issue), D259-266.

Doma, M. K., & Parker, R. (2006). Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature, 440*(7083), 561-564.

Dziembowski, A., Lorentzen, E., Conti, E., & Seraphin, B. (2007). A single subunit, Dis3, is essentially responsible for yeast exosome core activity. *Nat Struct Mol Biol, 14*(1), 15-22.

Eaton, J. D., Davidson, L., Bauer, D. L. V., Natsume, T., Kanemaki, M. T., & West, S. (2018). Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. *Genes Dev, 32*(2), 127-139.

Eberle, A. B., Lykke-Andersen, S., Muhlemann, O., & Jensen, T. H. (2009). SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. *Nat Struct Mol Biol, 16*(1), 49-55.

El Hage, A., Koper, M., Kufel, J., & Tollervey, D. (2008). Efficient termination of transcription by RNA polymerase I requires the 5' exonuclease Rat1 in yeast. *Genes Dev, 22*(8), 1069-1081.

Falk, S., Bonneau, F., Ebert, J., Kogel, A., & Conti, E. (2017). Mpp6 Incorporation in the Nuclear Exosome Contributes to RNA Channeling through the Mtr4 Helicase. *Cell Rep, 20*(10), 2279-2286.

Falk, S., Finogenova, K., Melko, M., Benda, C., Lykke-Andersen, S., Jensen, T. H., & Conti, E. (2016). Structure of the RBM7-ZCCHC8 core of the NEXT complex reveals connections to splicing factors. *Nat Commun, 7*, 13573.

Falk, S., Tants, J. N., Basquin, J., Thoms, M., Hurt, E., Sattler, M., & Conti, E. (2017). Structural insights into the interaction of the nuclear exosome helicase Mtr4 with the preribosomal protein Nop53. *RNA, 23*(12), 1780-1787.

Falk, S., Weir, J. R., Hentschel, J., Reichelt, P., Bonneau, F., & Conti, E. (2014). The molecular architecture of the TRAMP complex reveals the organization and interplay of its two catalytic activities. *Mol Cell, 55*(6), 856-867.

Fong, N., Brannan, K., Erickson, B., Kim, H., Cortazar, M. A., Sheridan, R. M., . . . Bentley, D. L. (2015). Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. *Mol Cell, 60*(2), 256-267.

Franks, T. M., Singh, G., & Lykke-Andersen, J. (2010). Upf1 ATPase-dependent mRNP disassembly is required for completion of nonsense- mediated mRNA decay. *Cell, 143*(6), 938-950.

Fukuhara, N., Ebert, J., Unterholzner, L., Lindner, D., Izaurralde, E., & Conti, E. (2005). SMG7 is a 14-3-3-like adaptor in the nonsense-mediated mRNA decay pathway. *Mol Cell, 17*(4), 537-547.

Gagnon, K. T., Li, L., Chu, Y., Janowski, B. A., & Corey, D. R. (2014). RNAi factors are present and active in human cell nuclei. *Cell Rep, 6*(1), 211-221.

Garneau, N. L., Wilusz, J., & Wilusz, C. J. (2007). The highways and byways of mRNA decay. *Nat Rev Mol Cell Biol, 8*(2), 113-126.

Gatfield, D., Unterholzner, L., Ciccarelli, F. D., Bork, P., & Izaurralde, E. (2003). Nonsense-mediated mRNA decay in Drosophila: at the intersection of the yeast and mammalian pathways. *EMBO J, 22*(15), 3960-3970.

Gerlach, P., Schuller, J. M., Bonneau, F., Basquin, J., Reichelt, P., Falk, S., & Conti, E. (2018). Distinct and evolutionary conserved structural features of the human nuclear exosome complex. *Elife, 7*.

Glavan, F., Behm-Ansmant, I., Izaurralde, E., & Conti, E. (2006). Structures of the PIN domains of SMG6 and SMG5 reveal a nuclease within the mRNA surveillance complex. *EMBO J, 25*(21), 5117-5125.

Glover, Marissa L., Burroughs, A. Max., Egelhofer, Thea A., Pule, Makena N., Aravind, L., & Arribere, Joshua A. (2019). NONU-1 encodes a conserved endonuclease required for mRNA translation surveillance. *bioRxiv*, 674358.

Goldstrohm, A. C., & Wickens, M. (2008). Multifunctional deadenylase complexes diversify mRNA control. *Nat Rev Mol Cell Biol, 9*(4), 337-344.

Graille, M., & Seraphin, B. (2012). Surveillance pathways rescuing eukaryotic ribosomes lost in translation. *Nat Rev Mol Cell Biol, 13*(11), 727-735.

Grimson, A., O'Connor, S., Newman, C. L., & Anderson, P. (2004). SMG-1 is a phosphatidylinositol kinase-related protein kinase required for nonsense-mediated mRNA Decay in Caenorhabditis elegans. *Mol Cell Biol, 24*(17), 7483-7490.

Grzechnik, P., & Kufel, J. (2008). Polyadenylation linked to transcription termination directs the processing of snoRNA precursors in yeast. *Mol Cell, 32*(2), 247-258.

Grzechnik, P., Szczepaniak, S. A., Dhir, S., Pastucha, A., Parslow, H., Matuszek, Z., . . . Proudfoot, N. J. (2018). Nuclear fate of yeast snoRNA is determined by co-transcriptional Rnt1 cleavage. *Nat Commun, 9*(1), 1783.

Gu, S. Q., Bakthavachalu, B., Han, J., Patil, D. P., Otsuka, Y., Guda, C., & Schoenberg, D. R. (2012). Identification of the human PMR1 mRNA endonuclease as an alternatively processed product of the gene for peroxidasin-like protein. *RNA, 18*(6), 1186-1196.

Gu, S. Q., Gallego-Perez, D., McClory, S. P., Shi, J., Han, J., Lee, L. J., & Schoenberg, D. R. (2016). The human PMR1 endonuclease stimulates cell motility by down regulating miR-200 family microRNAs. *Nucleic Acids Res, 44*(12), 5811-5819.

Haile, S., Estevez, A. M., & Clayton, C. (2003). A role for the exosome in the in vivo degradation of unstable mRNAs. *RNA, 9*(12), 1491-1501.

Halbach, F., Reichelt, P., Rode, M., & Conti, E. (2013). The yeast ski complex: crystal structure and RNA channeling to the exosome complex. *Cell, 154*(4), 814-826.

Halbach, F., Rode, M., & Conti, E. (2012). The crystal structure of S. cerevisiae Ski2, a DExH helicase associated with the cytoplasmic functions of the exosome. *RNA, 18*(1), 124-134.

Harigaya, Y., Tanaka, H., Yamanaka, S., Tanaka, K., Watanabe, Y., Tsutsumi, C., . . . Yamamoto, M. (2006). Selective elimination of messenger RNA prevents an incidence of untimely meiosis. *Nature, 442*(7098), 45-50.

Hilal, T., Yamamoto, H., Loerke, J., Burger, J., Mielke, T., & Spahn, C. M. (2016). Structural insights into ribosomal rescue by Dom34 and Hbs1 at near-atomic resolution. *Nat Commun, 7*, 13521.

Hiraishi, N., Ishida, Y. I., Sudo, H., & Nagahama, M. (2018). WDR74 participates in an early cleavage of the pre-rRNA processing pathway in cooperation with the nucleolar AAA-ATPase NVL2. *Biochem Biophys Res Commun, 495*(1), 116-123.

Hiraishi, N., Ishida, Y., & Nagahama, M. (2015). AAA-ATPase NVL2 acts on MTR4-exosome complex to dissociate the nucleolar protein WDR74. *Biochem Biophys Res Commun, 467*(3), 534-540.

Hodgkin, J., Papp, A., Pulak, R., Ambros, V., & Anderson, P. (1989). A new kind of informational suppression in the nematode Caenorhabditis elegans. *Genetics, 123*(2), 301-313.

Holub, P., Lalakova, J., Cerna, H., Pasulka, J., Sarazova, M., Hrazdilova, K., . . . Vanacova, S. (2012). Air2p is critical for the assembly and RNA-binding of the TRAMP complex and the KOW domain of Mtr4p is crucial for exosome activation. *Nucleic Acids Res, 40*(12), 5679-5693.

Holub, P., & Vanacova, S. (2012). TRAMP Stimulation of Exosome. *Enzymes, 31*, 77-95.

Hrossova, D., Sikorsky, T., Potesil, D., Bartosovic, M., Pasulka, J., Zdrahal, Z., . . . Vanacova, S. (2015). RBM7 subunit of the NEXT complex binds U-rich sequences and targets 3'-end extended forms of snRNAs. *Nucleic Acids Res, 43*(8), 4236-4248.

Hug, N., & Caceres, J. F. (2014). The RNA Helicase DHX34 Activates NMD by Promoting a Transition from the Surveillance to the Decay-Inducing Complex. *Cell Rep, 8*(6), 1845-1856.

Huntzinger, E., Kashima, I., Fauser, M., Sauliere, J., & Izaurralde, E. (2008). SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. *RNA, 14*(12), 2609-2617.

Ibrahim, H., Wilusz, J., & Wilusz, C. J. (2008). RNA recognition by 3'-to-5' exonucleases: the substrate perspective. *Biochim Biophys Acta, 1779*(4), 256-265.

Jankowsky, A., Guenther, U. P., & Jankowsky, E. (2011). The RNA helicase database. *Nucleic Acids Res, 39*(Database issue), D338-341.

Jia, H., Wang, X., Anderson, J. T., & Jankowsky, E. (2012). RNA unwinding by the Trf4/Air2/Mtr4 polyadenylation (TRAMP) complex. *Proc Natl Acad Sci U S A, 109*(19), 7292-7297.

Jiang, X., Sun, Y., Chen, S., Roy, K., & Price, B. D. (2006). The FATC domains of PIKK proteins are functionally equivalent and participate in the Tip60-dependent activation of DNA-PKcs and ATM. *J Biol Chem, 281*(23), 15741-15746.

Johnson, A. W. (1997). Rat1p and Xrn1p are functionally interchangeable exoribonucleases that are restricted to and required in the nucleus and cytoplasm, respectively. *Mol Cell Biol, 17*(10), 6122-6130.

Johnson, S. J., & Jackson, R. N. (2013). Ski2-like RNA helicase structures: common themes and complex assemblies. *RNA Biol, 10*(1), 33-43.

Jonas, S., Weichenrieder, O., & Izaurralde, E. (2013). An unusual arrangement of two 14-3-3-like domains in the SMG5-SMG7 heterodimer is required for efficient nonsense-mediated mRNA decay. *Genes Dev, 27*(2), 211-225.

Jones, C. I., Zabolotskaya, M. V., & Newbury, S. F. (2012). The 5' --> 3' exoribonuclease XRN1/Pacman and its functions in cellular processes and development. *Wiley Interdiscip Rev RNA, 3*(4), 455-468.

Kadaba, S., Wang, X., & Anderson, J. T. (2006). Nuclear RNA surveillance in Saccharomyces cerevisiae: Trf4p-dependent polyadenylation of nascent hypomethylated tRNA and an aberrant form of 5S rRNA. *RNA, 12*(3), 508-521.

Kadlec, J., Izaurralde, E., & Cusack, S. (2004). The structural basis for the interaction between nonsense-mediated mRNA decay factors UPF2 and UPF3. *Nat Struct Mol Biol, 11*(4), 330-337.

Karam, R., & Wilkinson, M. (2012). A conserved microRNA/NMD regulatory circuit controls gene expression. *RNA Biol, 9*(1), 22-26.

Karginov, F. V., Cheloufi, S., Chong, M. M., Stark, A., Smith, A. D., & Hannon, G. J. (2010). Diverse endonucleolytic cleavage sites in the mammalian transcriptome depend upon microRNAs, Drosha, and additional nucleases. *Mol Cell, 38*(6), 781-788.

Kashima, I., Jonas, S., Jayachandran, U., Buchwald, G., Conti, E., Lupas, A. N., & Izaurralde, E. (2010). SMG6 interacts with the exon junction complex via two conserved EJC-binding motifs (EBMs) required for nonsense-mediated mRNA decay. *Genes Dev, 24*(21), 2440-2450.

Keller, C., Woolcock, K., Hess, D., & Buhler, M. (2010). Proteomic and functional analysis of the noncanonical poly(A) polymerase Cid14. *RNA, 16*(6), 1124-1129.

Khemici, V., & Linder, P. (2018). RNA helicases in RNA decay. *Biochem Soc Trans, 46*(1), 163-172.

Kim, K., Heo, D. H., Kim, I., Suh, J. Y., & Kim, M. (2016). Exosome Cofactors Connect Transcription Termination to RNA Processing by Guiding Terminated Transcripts to the Appropriate Exonuclease within the Nuclear Exosome. *J Biol Chem, 291*(25), 13229-13242.

Kim, W. C., King, D., & Lee, C. H. (2010). RNA-cleaving properties of human apurinic/apyrimidinic endonuclease 1 (APE1). *Int J Biochem Mol Biol, 1*(1), 12-25.

Kowalinski, E., Kogel, A., Ebert, J., Reichelt, P., Stegmann, E., Habermann, B., & Conti, E. (2016). Structure of a Cytoplasmic 11-Subunit RNA Exosome Complex. *Mol Cell, 63*(1), 125-134.

Krzyszton, M., Zakrzewska-Placzek, M., Koper, M., & Kufel, J. (2012). Rat1 and Xrn2: The Diverse Functions of the Nuclear Rat1/Xrn2 Exonuclease. *Enzymes, 31*, 131-163.

Kulkarni, M., Ozgur, S., & Stoecklin, G. (2010). On track with P-bodies. *Biochem Soc Trans, 38*(Pt 1), 242-251.

Kurosaki, T., Popp, M. W., & Maquat, L. E. (2019). Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat Rev Mol Cell Biol, 20*(7), 406-420.

Labno, A., Tomecki, R., & Dziembowski, A. (2016). Cytoplasmic RNA decay pathways - Enzymes and mechanisms. *Biochim Biophys Acta, 1863*(12), 3125-3147.

LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A., & Tollervey, D. (2005). RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell, 121*(5), 713-724.

Lamanna, A. C., & Karbstein, K. (2009). Nob1 binds the single-stranded cleavage site D at the 3'-end of 18S rRNA with its PIN domain. *Proc Natl Acad Sci U S A, 106*(34), 14259-14264.

Larochelle, M., Lemay, J. F., & Bachand, F. (2012). The THO complex cooperates with the nuclear RNA surveillance machinery to control small nucleolar RNA expression. *Nucleic Acids Res, 40*(20), 10240-10253.

Le Hir, H., Moore, M. J., & Maquat, L. E. (2000). Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon-exon junctions. *Genes Dev, 14*(9), 1098-1108.

Lee, C. Y., Lee, A., & Chanfreau, G. (2003). The roles of endonucleolytic cleavage and exonucleolytic digestion in the 5'-end processing of S. cerevisiae box C/D snoRNAs. *RNA, 9*(11), 1362-1370.

Lee, K. P., Dey, M., Neculai, D., Cao, C., Dever, T. E., & Sicheri, F. (2008). Structure of the dual enzyme Ire1 reveals the basis for catalysis and regulation in nonconventional RNA splicing. *Cell, 132*(1), 89-100.

Li, X., & Manley, J. L. (2006). Cotranscriptional processes and their influence on genome stability. *Genes Dev, 20*(14), 1838-1847.

Li, Y., Yamane, D., & Lemon, S. M. (2015). Dissecting the roles of the 5' exoribonucleases Xrn1 and Xrn2 in restricting hepatitis C virus replication. *J Virol, 89*(9), 4857-4865.

Liu, Q., Greimann, J. C., & Lima, C. D. (2006). Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell, 127*(6), 1223-1237.

Liu, Z., Lee, A., & Gilbert, W. (1995). Gene disruption of a G4-DNA-dependent nuclease in yeast leads to cellular senescence and telomere shortening. *Proc Natl Acad Sci U S A, 92*(13), 6002-6006.

Longman, D., Plasterk, R. H., Johnstone, I. L., & Caceres, J. F. (2007). Mechanistic insights and identification of two novel factors in the C. elegans NMD pathway. *Genes Dev, 21*(9), 1075-1085.

Lorentzen, E., Basquin, J., Tomecki, R., Dziembowski, A., & Conti, E. (2008). Structure of the active subunit of the yeast exosome core, Rrp44: diverse modes of substrate recruitment in the RNase II nuclease family. *Mol Cell, 29*(6), 717-728.

Losh, J. S., & van Hoof, A. (2015). Gateway Arch to the RNA Exosome. *Cell, 162*(5), 940-941.

Lubas, M., Andersen, P. R., Schein, A., Dziembowski, A., Kudla, G., & Jensen, T. H. (2015). The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. *Cell Rep, 10*(2), 178-192.

Lubas, M., Christensen, M. S., Kristiansen, M. S., Domanski, M., Falkenby, L. G., Lykke-Andersen, S., . . . Jensen, T. H. (2011). Interaction profiling identifies the human nuclear exosome targeting complex. *Mol Cell, 43*(4), 624-637.

Luke, B., Panza, A., Redon, S., Iglesias, N., Li, Z., & Lingner, J. (2008). The Rat1p 5' to 3' exonuclease degrades telomeric repeat-containing RNA and promotes telomere elongation in Saccharomyces cerevisiae. *Mol Cell, 32*(4), 465-477.

Lykke-Andersen, J., & Bennett, E. J. (2014). Protecting the proteome: Eukaryotic cotranslational quality control pathways. *J Cell Biol, 204*(4), 467-476.

Macbeth, M. R., Schubert, H. L., Vandemark, A. P., Lingam, A. T., Hill, C. P., & Bass, B. L. (2005). Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. *Science, 309*(5740), 1534-1539.

Makino, D. L., & Conti, E. (2013). Structure determination of an 11-subunit exosome in complex with RNA by molecular replacement. *Acta Crystallogr D Biol Crystallogr, 69*(Pt 11), 2226-2235.

Makino, D. L., Schuch, B., Stegmann, E., Baumgartner, M., Basquin, C., & Conti, E. (2015). RNA degradation paths in a 12-subunit nuclear exosome complex. *Nature, 524*(7563), 54-58.

Matsushita, K., Takeuchi, O., Standley, D. M., Kumagai, Y., Kawagoe, T., Miyake, T., . . . Akira, S. (2009). Zc3h12a is an RNase essential for controlling immune responses by regulating mRNA decay. *Nature, 458*(7242), 1185-1190.

Melero, R., Hug, N., Lopez-Perrote, A., Yamashita, A., Caceres, J. F., & Llorca, O. (2016). The RNA helicase DHX34 functions as a scaffold for SMG1-mediated UPF1 phosphorylation. *Nat Commun, 7*, 10585.

Melero, R., Uchiyama, A., Castano, R., Kataoka, N., Kurosawa, H., Ohno, S., . . . Llorca, O. (2014). Structures of SMG1-UPFs complexes: SMG1 contributes to regulate UPF2-dependent activation of UPF1 in NMD. *Structure, 22*(8), 1105-1119.

Meola, N., & Jensen, T. H. (2017). Targeting the nuclear RNA exosome: Poly(A) binding proteins enter the stage. *RNA Biol, 14*(7), 820-826.

Miki, T. S., Richter, H., Ruegger, S., & Grosshans, H. (2014). PAXT-1 promotes XRN2 activity by stabilizing it through a conserved domain. *Mol Cell, 53*(2), 351-360.

Miller, S., Tavshanjian, B., Oleksy, A., Perisic, O., Houseman, B. T., Shokat, K. M., & Williams, R. L. (2010). Shaping development of autophagy inhibitors with the structure of the lipid kinase Vps34. *Science, 327*(5973), 1638-1642.

Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M., & Tollervey, D. (1997). The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-->5' exoribonucleases. *Cell, 91*(4), 457-466.

Molleston, J. M., Sabin, L. R., Moy, R. H., Menghani, S. V., Rausch, K., Gordesky-Gold, B., . . . Cherry, S. (2016). A conserved virus-induced cytoplasmic TRAMP-like complex recruits the exosome to target viral RNA for degradation. *Genes Dev, 30*(14), 1658-1670.

Monserrate, J. P., & York, J. D. (2010). Inositol phosphate synthesis and the nuclear processes they affect. *Curr Opin Cell Biol, 22*(3), 365-373.

Montpetit, B., Thomsen, N. D., Helmke, K. J., Seeliger, M. A., Berger, J. M., & Weis, K. (2011). A conserved mechanism of DEAD-box ATPase activation by nucleoporins and InsP6 in mRNA export. *Nature, 472*(7342), 238-242.

Morita, T., Yamashita, A., Kashima, I., Ogata, K., Ishiura, S., & Ohno, S. (2007). Distant N- and C-terminal domains are required for intrinsic kinase activity of SMG-1, a critical component of nonsense-mediated mRNA decay. *J Biol Chem, 282*(11), 7799-7808.

Nagarajan, V. K., Jones, C. I., Newbury, S. F., & Green, P. J. (2013). XRN 5'-->3' exoribonucleases: structure, mechanisms and functions. *Biochim Biophys Acta, 1829*(6-7), 590-603.

Oeffinger, M., Zenklusen, D., Ferguson, A., Wei, K. E., El Hage, A., Tollervey, D., . . . Rout, M. P. (2009). Rrp17p is a eukaryotic exonuclease required for 5' end processing of Pre-60S ribosomal RNA. *Mol Cell, 36*(5), 768-781.

Ogami, K., Chen, Y., & Manley, J. L. (2018). RNA surveillance by the nuclear RNA exosome: mechanisms and significance. *Noncoding RNA, 4*(1).

Ogami, K., Richard, P., Chen, Y., Hoque, M., Li, W., Moresco, J. J., . . . Manley, J. L. (2017). An Mtr4/ZFC3H1 complex facilitates turnover of unstable nuclear RNAs to prevent their cytoplasmic transport and global translational repression. *Genes Dev, 31*(12), 1257-1271.

Okada-Katsuhata, Y., Yamashita, A., Kutsuzawa, K., Izumi, N., Hirahara, F., & Ohno, S. (2012). N- and C-terminal Upf1 phosphorylations create binding platforms for SMG-6 and SMG-5:SMG-7 during NMD. *Nucleic Acids Res, 40*(3), 1251-1266.

Ouyang, Z., Zheng, G., Tomchick, D. R., Luo, X., & Yu, H. (2016). Structural Basis and IP6 Requirement for Pds5-Dependent Cohesin Dynamics. *Mol Cell, 62*(2), 248-259.

Park, E., & Maquat, L. E. (2013). Staufen-mediated mRNA decay. *Wiley Interdiscip Rev RNA, 4*(4), 423-435.

Patrick, E. M., Srinivasan, S., Jankowsky, E., & Comstock, M. J. (2017). The RNA helicase Mtr4p is a duplex-sensing translocase. *Nat Chem Biol, 13*(1), 99-104.

Pefanis, E., Wang, J., Rothschild, G., Lim, J., Chao, J., Rabadan, R., . . . Basu, U. (2014). Noncoding RNA transcription targets AID to divergently transcribed loci in B cells. *Nature, 514*(7522), 389-393.

Pisareva, V. P., Skabkin, M. A., Hellen, C. U., Pestova, T. V., & Pisarev, A. V. (2011). Dissociation by Pelota, Hbs1 and ABCE1 of mammalian vacant 80S ribosomes and stalled elongation complexes. *EMBO J, 30*(9), 1804-1817.

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M. S., Mapendano, C. K., . . . Jensen, T. H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. *Science, 322*(5909), 1851-1854.

Preti, M., O'Donohue, M. F., Montel-Lehry, N., Bortolin-Cavaille, M. L., Choesmel, V., & Gleizes, P. E. (2013). Gradual processing of the ITS1 from the nucleolus to the cytoplasm during synthesis of the human 18S rRNA. *Nucleic Acids Res, 41*(8), 4709-4723.

Priestley, A., Beamish, H. J., Gell, D., Amatucci, A. G., Muhlmann-Diaz, M. C., Singleton, B. K., . . . Taccioli, G. E. (1998). Molecular and biochemical characterisation of DNA-dependent protein kinase-defective rodent mutant irs-20. *Nucleic Acids Res, 26*(8), 1965-1973.

Pulak, R., & Anderson, P. (1993). mRNA surveillance by the Caenorhabditis elegans smg genes. *Genes Dev, 7*(10), 1885-1897.

Puno, M. R., & Lima, C. D. (2018). Structural basis for MTR4-ZCCHC8 interactions that stimulate the MTR4 helicase in the nuclear exosome-targeting complex. *Proc Natl Acad Sci U S A, 115*(24), E5506-E5515.

Razew, M., Warkocki, Z., Taube, M., Kolondra, A., Czarnocki-Cieciura, M., Nowak, E., . . . Nowotny, M. (2018). Structural analysis of mtEXO mitochondrial RNA degradosome reveals tight coupling of nuclease and helicase components. *Nat Commun, 9*(1), 97.

Richter, H., Katic, I., Gut, H., & Grosshans, H. (2016). Structural basis and function of XRN2 binding by XTB domains. *Nat Struct Mol Biol, 23*(2), 164-171.

Robinson, S. R., Oliver, A. W., Chevassut, T. J., & Newbury, S. F. (2015). The 3' to 5' Exoribonuclease DIS3: From Structure and Mechanisms to Biological Functions and Role in Human Disease. *Biomolecules, 5*(3), 1515-1539.

Rougemaille, M., Gudipati, R. K., Olesen, J. R., Thomsen, R., Seraphin, B., Libri, D., & Jensen, T. H. (2007). Dissecting mechanisms of nuclear mRNA surveillance in THO/sub2 complex mutants. *EMBO J, 26*(9), 2317-2326.

Rowley, P. A., Ho, B., Bushong, S., Johnson, A., & Sawyer, S. L. (2016). XRN1 Is a Species-Specific Virus Restriction Factor in Yeasts. *PLoS Pathog, 12*(10), e1005890.

Ryan, K., Calvo, O., & Manley, J. L. (2004). Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *RNA, 10*(4), 565-573.

Saito, S., Hosoda, N., & Hoshino, S. (2013). The Hbs1-Dom34 protein complex functions in non-stop mRNA decay in mammalian cells. *J Biol Chem, 288*(24), 17832-17843.

Schaeffer, D., Clark, A., Klauer, A. A., Tsanova, B., & van Hoof, A. (2011). Functions of the cytoplasmic exosome. *Adv Exp Med Biol, 702*, 79-90.

Schaeffer, D., Tsanova, B., Barbas, A., Reis, F. P., Dastidar, E. G., Sanchez-Rotunno, M., . . . van Hoof, A. (2009). The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities. *Nat Struct Mol Biol, 16*(1), 56-62.

Schaeffer, D., & van Hoof, A. (2011). Different nuclease requirements for exosome-mediated degradation of normal and nonstop mRNAs. *Proc Natl Acad Sci U S A, 108*(6), 2366-2371.

Schmidt, K., & Butler, J. S. (2013). Nuclear RNA surveillance: role of TRAMP in controlling exosome specificity. *Wiley Interdiscip Rev RNA, 4*(2), 217-231.

Schneider, C., Kudla, G., Wlotzka, W., Tuck, A., & Tollervey, D. (2012). Transcriptome-wide analysis of exosome targets. *Mol Cell, 48*(3), 422-433.

Schneider, C., Leung, E., Brown, J., & Tollervey, D. (2009). The N-terminal PIN domain of the exosome subunit Rrp44 harbors endonuclease activity and tethers Rrp44 to the yeast core exosome. *Nucleic Acids Res, 37*(4), 1127-1140.

Schuch, B., Feigenbutz, M., Makino, D. L., Falk, S., Basquin, C., Mitchell, P., & Conti, E. (2014). The exosome-binding factors Rrp6 and Rrp47 form a composite surface for recruiting the Mtr4 helicase. *EMBO J, 33*(23), 2829-2846.

Schuller, J. M., Falk, S., Fromm, L., Hurt, E., & Conti, E. (2018). Structure of the nuclear exosome captured on a maturing preribosome. *Science, 360*(6385), 219-222.

Sheth, U., & Parker, R. (2003). Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. *Science, 300*(5620), 805-808.

Shoemaker, C. J., Eyler, D. E., & Green, R. (2010). Dom34:Hbs1 promotes subunit dissociation and peptidyl-tRNA drop-off to initiate no-go decay. *Science, 330*(6002), 369-372.

Shoemaker, C. J., & Green, R. (2011). Kinetic analysis reveals the ordered coupling of translation termination and ribosome recycling in yeast. *Proc Natl Acad Sci U S A, 108*(51), E1392-1398.

Shoemaker, C. J., & Green, R. (2012). Translation drives mRNA quality control. *Nat Struct Mol Biol, 19*(6), 594-601.

Silla, T., Karadoulama, E., Makosa, D., Lubas, M., & Jensen, T. H. (2018). The RNA Exosome Adaptor ZFC3H1 Functionally Competes with Nuclear Export Activity to Retain Target Transcripts. *Cell Rep, 23*(7), 2199-2210.

Simms, C. L., Thomas, E. N., & Zaher, H. S. (2017). Ribosome-based quality control of mRNA and nascent peptides. *Wiley Interdiscip Rev RNA, 8*(1).

Sloan, K. E., Mattijssen, S., Lebaron, S., Tollervey, D., Pruijn, G. J., & Watkins, N. J. (2013). Both endonucleolytic and exonucleolytic cleavage mediate ITS1 removal during human ribosomal RNA processing. *J Cell Biol, 200*(5), 577-588.

Stevens, A. (1980). Purification and characterization of a Saccharomyces cerevisiae exoribonuclease which yields 5'-mononucleotides by a 5' leads to 3' mode of hydrolysis. *J Biol Chem, 255*(7), 3080-3085.

Szczepinska, T., Kalisiak, K., Tomecki, R., Labno, A., Borowski, L. S., Kulinski, T. M., . . . Dziembowski, A. (2015). DIS3 shapes the RNA polymerase II transcriptome in humans by degrading a variety of unwanted transcripts. *Genome Res, 25*(11), 1622-1633.

Tanner, N. K., & Linder, P. (2001). DExD/H box RNA helicases: from generic motors to specific dissociation functions. *Mol Cell, 8*(2), 251-262.

Taylor, L. L., Jackson, R. N., Rexhepaj, M., King, A. K., Lott, L. K., van Hoof, A., & Johnson, S. J. (2014). The Mtr4 ratchet helix and arch domain both function to promote RNA unwinding. *Nucleic Acids Res, 42*(22), 13861-13872.

Thoms, M., Thomson, E., Bassler, J., Gnadig, M., Griesel, S., & Hurt, E. (2015). The Exosome Is Recruited to RNA Substrates through Specific Adaptor Proteins. *Cell, 162*(5), 1029-1038.

Tiedje, C., Lubas, M., Tehrani, M., Menon, M. B., Ronkina, N., Rousseau, S., . . . Gaestel, M. (2015). p38MAPK/MK2-mediated phosphorylation of RBM7 regulates the human nuclear exosome targeting complex. *RNA, 21*(2), 262-278.

Tomecki, R., Kristiansen, M. S., Lykke-Andersen, S., Chlebowski, A., Larsen, K. M., Szczesny, R. J., . . . Jensen, T. H. (2010). The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L. *EMBO J, 29*(14), 2342-2357.

Tudek, A., Porrua, O., Kabzinski, T., Lidschreiber, M., Kubicek, K., Fortova, A., . . . Libri, D. (2014). Molecular basis for coordinating transcription termination with noncoding RNA degradation. *Mol Cell, 55*(3), 467-481.

Unterholzner, L., & Izaurralde, E. (2004). SMG7 acts as a molecular link between mRNA surveillance and mRNA decay. *Mol Cell, 16*(4), 587-596.

Valencia-Sanchez, M. A., Liu, J., Hannon, G. J., & Parker, R. (2006). Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev, 20*(5), 515-524.

van der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., . . . Babu, M. M. (2014). Classification of intrinsically disordered regions and proteins. *Chem Rev, 114*(13), 6589-6631.

van Hoof, A., Frischmeyer, P. A., Dietz, H. C., & Parker, R. (2002). Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science, 295*(5563), 2262-2264.

Vanacova, S., Wolf, J., Martin, G., Blank, D., Dettwiler, S., Friedlein, A., . . . Keller, W. (2005). A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS Biol, 3*(6), e189.

Vasiljeva, L., & Buratowski, S. (2006). Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Mol Cell, 21*(2), 239-248.

Vasudevan, S., Peltz, S. W., & Wilusz, C. J. (2002). Non-stop decay--a new mRNA surveillance pathway. *Bioessays, 24*(9), 785-788.

Volanakis, A., Passoni, M., Hector, R. D., Shah, S., Kilchert, C., Granneman, S., & Vasiljeva, L. (2013). Spliceosome-mediated decay (SMD) regulates expression of nonintronic genes in budding yeast. *Genes Dev, 27*(18), 2025-2038.

Walker, E. H., Perisic, O., Ried, C., Stephens, L., & Williams, R. L. (1999). Structural insights into phosphoinositide 3-kinase catalysis and signalling. *Nature, 402*(6759), 313-320.

Wang, J., Chen, J., Wu, G., Zhang, H., Du, X., Chen, S., . . . Cheng, H. (2019). NRDE2 negatively regulates exosome functions by inhibiting MTR4 recruitment and exosome interaction. *Genes Dev, 33*(9-10), 536-549.

Wang, X., Arai, S., Song, X., Reichart, D., Du, K., Pascual, G., . . . Kurokawa, R. (2008). Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature, 454*(7200), 126-130.

Wasmuth, E. V., Januszyk, K., & Lima, C. D. (2014). Structure of an Rrp6-RNA exosome complex bound to poly(A) RNA. *Nature, 511*(7510), 435-439.

Weick, E. M., Puno, M. R., Januszyk, K., Zinder, J. C., DiMattia, M. A., & Lima, C. D. (2018). Helicase-Dependent RNA Decay Illuminated by a Cryo-EM Structure of a Human Nuclear RNA Exosome-MTR4 Complex. *Cell, 173*(7), 1663-1677 e1621.

Weir, J. R., Bonneau, F., Hentschel, J., & Conti, E. (2010). Structural analysis reveals the characteristic features of Mtr4, a DExH helicase involved in nuclear RNA processing and surveillance. *Proc Natl Acad Sci U S A, 107*(27), 12139-12144.

Wells, G. R., Weichmann, F., Sloan, K. E., Colvin, D., Watkins, N. J., & Schneider, C. (2017). The ribosome biogenesis factor yUtp23/hUTP23 coordinates key interactions in the yeast and human pre-40S particle and hUTP23 contains an essential PIN domain. *Nucleic Acids Res, 45*(8), 4796-4809.

West, S., Gromak, N., & Proudfoot, N. J. (2004). Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature, 432*(7016), 522-525.

Widner, W. R., & Wickner, R. B. (1993). Evidence that the SKI antiviral system of Saccharomyces cerevisiae acts by blocking expression of viral mRNA. *Mol Cell Biol, 13*(7), 4331-4341.

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J. C., Dufour, M. E., Boulay, J., . . . Jacquier, A. (2005). Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell, 121*(5), 725-737.

Xiang, S., Cooper-Morgan, A., Jiao, X., Kiledjian, M., Manley, J. L., & Tong, L. (2009). Structure and function of the 5'-->3' exoribonuclease Rat1 and its activating partner Rai1. *Nature, 458*(7239), 784-788.

Xue, Y., Bai, X., Lee, I., Kallstrom, G., Ho, J., Brown, J., . . . Johnson, A. W. (2000). Saccharomyces cerevisiae RAI1 (YGL246c) is homologous to human DOM3Z and encodes a protein that binds the nuclear exoribonuclease Rat1p. *Mol Cell Biol, 20*(11), 4006-4015.

Yamashita, A., Izumi, N., Kashima, I., Ohnishi, T., Saari, B., Katsuhata, Y., . . . Ohno, S. (2009). SMG-8 and SMG-9, two novel subunits of the SMG-1 complex, regulate remodeling of the mRNA surveillance complex during nonsense-mediated mRNA decay. *Genes Dev, 23*(9), 1091-1105.

Yang, H., Jiang, X., Li, B., Yang, H. J., Miller, M., Yang, A., . . . Pavletich, N. P. (2017). Mechanisms of mTORC1 activation by RHEB and inhibition by PRAS40. *Nature, 552*(7685), 368-373.

Yang, X. C., Sullivan, K. D., Marzluff, W. F., & Dominski, Z. (2009). Studies of the 5' exonuclease and endonuclease activities of CPSF-73 in histone pre-mRNA processing. *Mol Cell Biol, 29*(1), 31-42.

Zhang, H., Lu, Y., Chen, E., Li, X., Lv, B., Vikis, H. G., & Liu, P. (2017). XRN2 promotes EMT and metastasis through regulating maturation of miR-10a. *Oncogene, 36*(27), 3925-3933.

Zhou, Y., Zhu, J., Schermann, G., Ohle, C., Bendrin, K., Sugioka-Sugiyama, R., . . . Fischer, T. (2015). The fission yeast MTREC complex targets CUTs and unspliced pre-mRNAs to the nuclear exosome. *Nat Commun, 6*, 7050.

Zinder, J. C., & Lima, C. D. (2017). Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. *Genes Dev, 31*(2), 88-100.

7.0 ACKNOWLEDGEMENTS

None of what is presented in this document would be possible without the support of my adviser Prof. Elena Conti. I would like to express my sincerest thanks to her for hiring me, allowing me to work in her laboratory, for her ever-supportive guidance and for placing enough trust in me even when the winds were blowing against me.

I would like to thank the IMPRS-LS team for organizing and managing a wonderful graduate program without which it would have been very hard for someone like me to get noticed and get a PhD position at the Max-Planck Institute of Biochemistry. The Boehringer Ingelheim Fonds team has my gratitude for funding the initial stages of my PhD and for letting me take part in very useful workshops and retreats. Immense thanks are also owed to Prof. Andreas Ladurner and Prof. Esben Lorentzen for being part of my thesis committee, for always providing encouragement and constructive comments to help along different projects during the course of my PhD.

I would like to thank Sebastian Falk for his patient guidance and for teaching me a lot of valuable lessons that I hope to carry with me for the rest of my career. I owe a lot to Sebastian for his intellectual contribution and for being a very strong driving force behind anything and everything that is related to MTR4. I would also like to thank Jérôme Basquin for his enthusiastic support and guidance on the crystallography part of the project.

Despite not being blessed with good fortune in the SMG project, I learnt a lot from several people in the lab for which I feel immensely grateful. To start with, I would like to thank Fabien Bonneau and Ingmar Schaefer for teaching me a lot during the early days and for their continued support and encouragement. I would also like to thank Fabien Bonneau again, Jan Schuller, Yair Gat and Liang Li for persevering with their respective aspects of the project and finally providing some meaning to the time I spent on the project.

A very special thanks is also owed to Daniela who battled bravely to finally prepare stable cells that express SMG1, the protein purification trio – Steffen Schussler, Elisabeth Stegmann and Judith Ebert for performing countless purifications which all contributed in their own little ways to move projects along, Peter for his resourcefulness and column pacing skills, the crystallization facility team for always being open to work with my demands,

Claire Basquin for support with several biophysics experiments, Marc Baumgartner along with Elfriede Eppinger for insect cell culture support, Thomas Ebert for media prep that aided in growing hundreds of liters of MTR4 cultures, Andrea Spankova for the SDS-PAGE gels, Walter for his masterful troubleshooting and the kitchen staff for ensuring that the glassware is ready and available.

I felt very lucky to be surrounded by a very talented crop of Conti postdoctoral fellows and gained a lot by being around them. I'd like to thank Rajan Prabhu, Eva Kowalinski, Sutapa Chakrabarthi, Sevim Özgür and Piotr Gerlach for being very approachable and answering any and every question I might have had regarding tackling issues with the project.

A good PhD timeline is not just about working with amazingly talented people and on challenging projects, it is also about working in a very healthy supportive environment and I needed a lot of support during the course of my PhD. Firstly, I thank Petra Lee and Ulrike Goldschmitt for administrative support which allowed me to have a smooth transition into 'German' life. I owe immense gratitude to Jana Albrecht, Michaela Hartwig, Giulia Chiapparini and Ksenia Finogenova for their kindness and support during the darkest times in the last five years. On the same token, I would like to thank Christian Benda for always being open to listen to all of my emotional battles and at the same time providing very sound project related input. I would also like to thank Nicole Eisele, Achim Keidel, Alexander Koegel, Felix Sandmeir, Lukas Langer and Kenny Jungfer for always being open to my oversharing and complaining tendencies and I appreciate every single minute that they spared for me. I also owe thanks to the lunch club, the Conti volleyball team, Conti PhD dinner club and all the members of the Conti department for contributing in their own ways to make the life centered around MPIB enjoyable.

I'd also like to thank my friends beyond MPI especially Cristina for being a positive impact on me and for selflessly supporting me through the writing phase of my PhD. Finally, I would like to thank my family, specifically my mother, sister and my uncle who sacrificed a lot to provide for me and enable me to be where I am right now.