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

Structural and Biochemical Characterization of the *C. elegans* SMG8-SMG9 Core Complex

Liang Li

aus

Huanggang, Hubei in China

2016

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, den 12. Dezember 2016

.....

Liang Li

Dissertation eingereicht am 20. 12. 2016

1. Gutachterin: Prof. Dr. Elena Conti

2. Gutachter: Prof. Dr. Karl-Peter Hopfner

Mündliche Prüfung am 24. 01. 2017

Abstract

Nonsense mediated mRNA decay (NMD) is an important mRNA quality control pathway conserved in eukaryotes. NMD targets aberrant mRNAs carrying premature stop codons (PTCs) for rapid degradation, preventing the accumulation of C-terminally truncated protein products that would otherwise be toxic to cells. NMD involves the concerted action of several trans-acting factors and it is a highly regulated process. A decisive event to trigger NMD in metazoans is the phosphorylation of the RNA helicase UPF1 by the SMG1 kinase. SMG8 and SMG9 form a heterodimer that interacts with SMG1 and inhibits its kinase activity. In recent years, electron microscopy studies of the human SMG1-SMG8-SMG9 complex provided low-resolution structural information that revealed the overall architecture of this complex. Still not much is known about the structure and function of SMG8 and SMG9 and how they interact with each other as well as with SMG1. In this thesis, I used biochemical approaches to identify the core of a SMG8-SMG9 complex amenable to crystallization and determined its three-dimensional structure at the resolution of 2.5 Å. I found that the *C. elegans* SMG8-SMG9 core complex resembles a G-domain heterodimer with a potentially active subunit (SMG9) and an inactive subunit (SMG8). Following this result, I characterized the nucleotide-binding properties of SMG-SMG9 using biophysical and structural methods. Fitting the atomic model in a previously published low-resolution EM map of a SMG1-SMG8-SMG9 complex raises interesting possibility that the nucleotide-binding state of SMG8-SMG9 might impact on the function of the kinase.

Contents

List of figures

List of Tables

Abbreviation

ATP	adenosine triphosphate
ADP	adenosine diphosphate
Amp	ampicillin
bp	base pair
β-ME	β-mercaptoethanol
dNTP	deoxynucleotide triphosphate
dTTP	2'-Deoxythymidine 5'-triphosphate
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EJC	exon junction complex
EBM	EJC-binding motif
EM	Electron Microscopy
E. coli	Escherichia coli
FL	Full length
GTP	Guanosine triphosphate
GDP	Guanosine diphosphate
GST	glutathione S-transferase
IP	Immunoprecipitation
IPTG	isopropyl β-D-1-thiogalactopyranoside
Kan	kanamycin
Kd	dissociation constant
LIC	ligation independent cloning
LB	Luria-Bertani
mRNA	messenger RNA
mRNP	messenger ribonucleoprotein particle
Mant	methylanthraniloyl
MCS	Multiple cloning site
NEB	New England Biolabs
NMD	Nonsense mediated mRNA decay

PEG	polyethylene glycol
P-loop	phosphate-binding loop
PTC	premature termination codon
PIKK	Phosphatidylinositol 3-kinase-related kinase
PI3K	Phosphatidylinositol 3-kinase
PCR	Polymerase chain reaction
PDB	Protein data bank
RNA	Ribonucleic acid
RRM	RNA recognition motif
RNAi	RNA interference
SMG	suppressor with morphological effect on genitalia
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
Trx	Thioredoxin
TB	Terrific Broth
UPF	up frame-shift
uORF	upstream open reading frame
UTR	untranslated region

1 Introduction

Genetic information in protein encoding genes is transferred to messenger RNA (mRNA) through the process of transcription and ultimately decoded to protein via the process of translation. The fidelity of gene transcription and translation is vital for cellular activities. Errors could occur and accumulate during transcription or splicing that lead to mutations on mRNA. Proteins translated from aberrant mRNA templates be misfolded or be truncated in the case of nonsense mutations. Eukaryotic organisms have evolved multiple quality control systems to detect aberrant mRNAs and subject them to rapid degradation. Three predominant translation-coupled mRNA quality control pathways have been discovered in eukaryotes termed Nonsense mediated mRNA decay (NMD), No-go decay (NGD) and Nonstop decay (NSD). NMD targets mRNAs that harbor a premature translation stop codon (PTC) for degradation and thus prevent production of C-terminally truncated proteins that might be toxic to cells [1]. NGD recognizes and degrades mRNAs that contain structured features that block translating ribosomes [2,3,4]. NSD targets mRNAs that lack a stop codon causing ribosomes to run to the end of mRNAs and stall [5,6]. This thesis work is focused on NMD and the main progress in NMD research that has been reported over the last three decades will be discussed in this chapter.

1.1 Nonsense mediated mRNA decay (NMD)

It was first observed over 30 years ago in yeast and human cells that half-lives of affected mRNAs can be reduced by nonsense mutation that terminates translation prematurely[7,8]. The term “Nonsense mediated mRNA decay” was first used in 1993 by Peltz et al. to describe the phenomenon that nonsense mutations can accelerate the decay rate of mRNAs[9]. NMD has been found in all eukaryotes examined and the core NMD machinery is conserved from yeast to humans[10]. Since NMD was discovered, it has been defined based on its function as a eukaryotic surveillance mechanism targeting mRNAs that contain a premature translation termination codon (PTC). However, in the last 10 years evidences from genome-wide analysis in different eukaryotic organisms suggest that NMD not only can target aberrant mRNAs harboring PTCs but also affects the stability of normal mRNAs that encode complete protein products, thereby regulating expression levels of a significant amount of

endogenous mRNAs[11] [12-14]. Therefore our knowledge to the function of NMD has expanded from a classical eukaryotic mRNA quality control pathway to post-transcriptional regulation of gene expression. However, despite extensive researches have been conducted on NMD and a plethora of structural and biochemical data are available, the molecular mechanism of NMD is still elusive due to its complexity and difficulty to recapitulate the whole process in vitro.

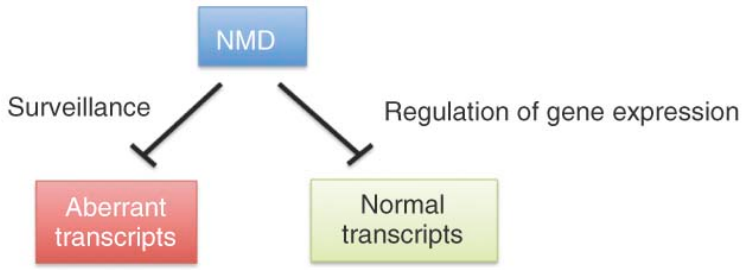


Figure 1.1: The dual role of NMD[15] (figure adapted from Ref. 15). NMD is not only a surveillance pathway that degrades aberrant transcripts but also involved in the post-transcriptional regulation of gene expression by influencing the stability of many normal transcripts.

1.1.1 NMD-inducing features

It is widely accepted that NMD targets aberrant PTC-containing and also apparently normal mRNAs. It has been estimated that depending on organisms or cell types, around 5%-20% of transcripts in a typical transcriptome are substrates of NMD[16]. NMD can be elicited only when a stop codon on mRNAs is in a context that can activate the NMD machinery. However, the NMD-inducing features of transcripts that have been discovered are rather diverse. The features of NMD substrates that have been reported can be described in the following two classes. The first class, which is also the classical one, includes typical substrates that contain a destabilizing stop codon inside the coding region. The premature stop codon can be generated by nonsense or frame shift mutations in endogenous genes, or by errors from transcription or alternative splicing events[17]. The second class contains physiologically relevant and apparently normal mRNAs such as mRNAs with upstream open reading frames (uORFs)[18,19] and mRNAs with long 3′ untranslated regions (3′-UTRs)[20,21]. Despite the fact that with current knowledge there is no unified NMD-inducing feature that can

define a NMD substrate and not all PTCs or uORFs can trigger NMD, one requirement for NMD to be initiated is that the NMD machinery needs to be recruited to the ribosome, which stalls on a stop codon and is not able to dissociate efficiently as in normal translation termination.

1.1.2 NMD machinery

NMD is a translation-dependent process conserved in eukaryotic organisms from yeast to humans. In order to activate efficient degradation of NMD substrates, multiple trans-acting factors have been employed by eukaryotic cells to assemble the NMD machinery to be recruited to the aberrantly terminated ribosome on a stop codon. Genetic screens in yeast identified the first three NMD factors called up-frameshift (UPF) proteins UPF1, UPF2 and UPF3[22,23] and genetic screens in *C. elegans* identified 7 NMD effectors that are named from SMG1-SMG7 (suppressor with morphological effect on genitalia)[24,25]. SMG2, SMG3 and SMG4 in *C. elegans* are the homologues for yeast UPF1, UPF2 and UPF3 respectively[26]. A few novel NMD factors including SMG8 and SMG9 have been identified in recent years with genetic screen or Co-IP with known NMD factors or homology search with bioinformatic methods[27,28], the mechanistic roles of which still need to be characterized. The NMD pathway in different organisms requires different factors that constitute its NMD machinery. Some factors such as the UPF proteins are conserved in all eukaryotes that have been studied that constitute the core NMD machinery while other factors such as homologues of SMG1, SMG5-7 proteins present only in higher eukaryotes and the exon junction complex (EJC) is required for NMD only in mammalian cells[1].

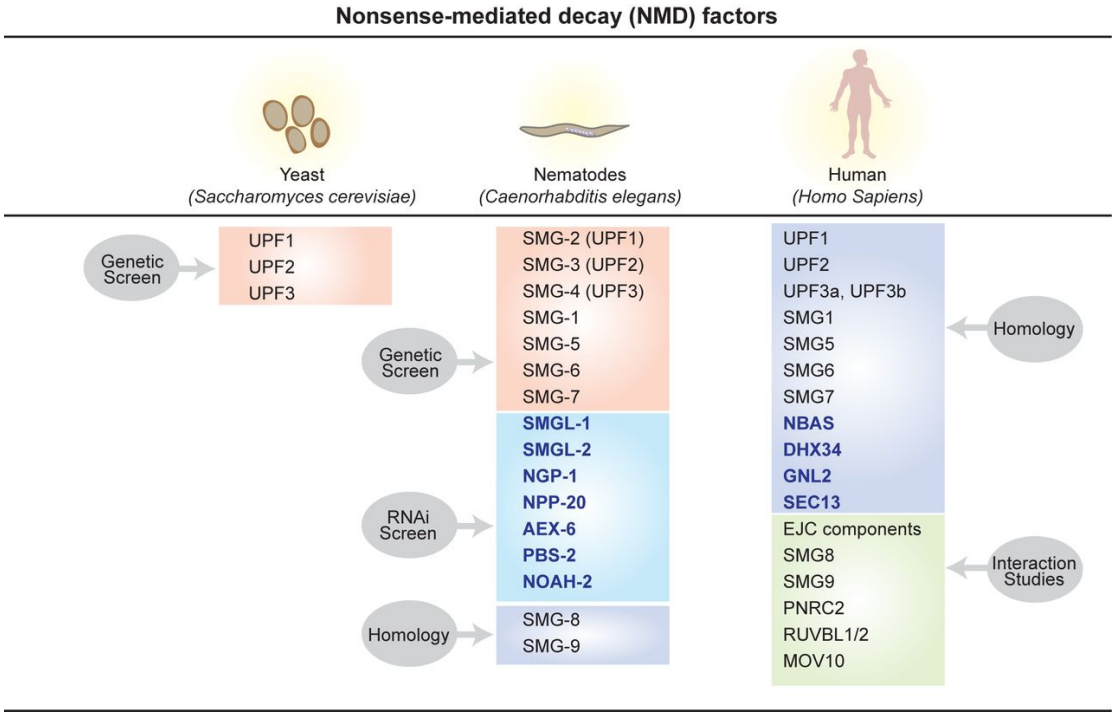


Figure 1.2: NMD factors (figure adapted from Ref. 28). NMD factors that have been identified in *Saccharomyces cerevisiae, Caenorhabditis elegans and Homo Sapiens* with different methods[28].

1.1.2.1 Core UPF proteins

UPF1 is a monomeric SF1 (superfamily 1) RNA helicase that plays the central role in NMD in all eukaryotes. UPF1 has a modular domain organization with a central helicase domain that is flanked by a conserved N-terminal cysteine-histidine-rich (CH) domain and a C-terminal serine-glutamine-rich (SQ) domain, which in metazoans gets phosphorylated by SMG1 at multiple SQ-motifs[29]. The UPF1 helicase and ATPase activities are essential for NMD in both yeast and humans[30,31] and are highly regulated by intra-molecular and intermolecular interactions[32]. The central helicase domain is comprised of two flexible RecA domains with the ATP binding site located in the cleft between these two domains[29]. The UPF1 binding affinity for RNA is reduced in the ATP-bound form[33,34]. The CH domain and SQ domain together suppress UPF1 helicase activity in vitro. Binding of UPF2 to the CH domain of UPF1 induces a large conformational change in UPF1 that promotes its ATPase and helicase activity, which is also a prerequisite for UPF1 phosphorylation[29,33].

UPF2 is the second core NMD factor that functions as a ring-like scaffold bridging UPF1 and UPF3[33,35]. UPF2 is comprised of three MIF4G (middle portion of eIF4G) domains, first two of which provide structural support[36] and the third domain interacts with UPF3B[37]. The C-terminal part of UPF2 called UBD interacts with UPF1[38].

UPF3 is the third core NMD factor. In humans, due to alternative splicing two isoforms of UPF3 co-exist, UPF3A and UPF3B[39,40]. However, UPF3B is the dominant isoform that functions in NMD in human cells. Only when UPF3B is depleted, UPF3A is stabilized and substitute UPF3B in NMD[41,42]. UPF3 shuttles between nucleus and cytoplasm but in steady state is found primarily in nucleus[38] where it is thought to associate with EJCs upon pre-mRNA splicing. UPF3 contains a N-terminal RNA recognition motif (RRM) that is the binding surface for UPF2 instead of RNA[37] and a short motif at the C-terminus called EJC binding motif (EBM) that interacts with EJC[33,43]. Therefore, UPF3 is the bridging molecule linking EJC and UPF2.

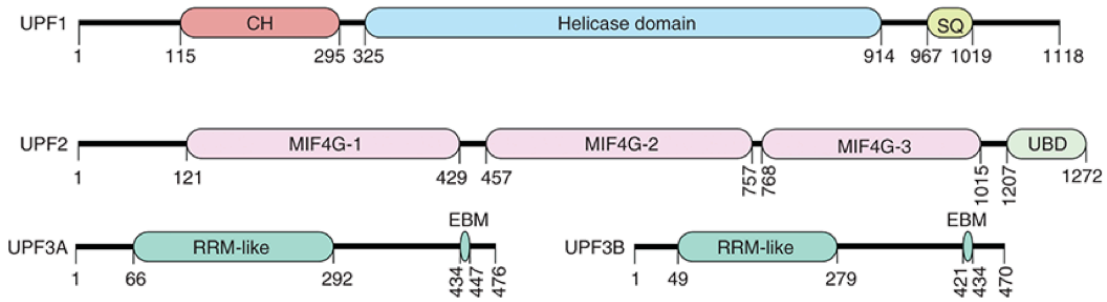


Figure 1.3: Domain architectures of human UPF proteins[1] (figure adapted from Ref. 1). CH: cysteine-histidine rich domain; SQ: serine-glutamine rich domain; MIF4G: middle of 4G-like domains; UBD: UPF1-binding domain; RRM: RNA recognition motif; EBM: exon junction binding motif.

1.1.2.2 Additional SMG proteins

SMG1 is one of essential players for triggering NMD response in metazoans[44,45]. SMG1 was first identified by genetic screen in *C. elegans*[24,25] and its human homologue was found by homology[28]. Human SMG1 is a large protein of ~410 kDa in molecular weight and belongs to the phosphatidylinositol 3-kinase-related protein kinase (PIKK) family of serine–threonine kinases. Similar to other members in the PIKK family, the primary structure of SMG1 contains a conserved C terminus preceded by a long stretch of helical repeats. The conserved C terminus is comprised of

a FAT (FRAP/TOR, ATM, and TRRAP) domain, a FRB (FKBP-rapamycin-binding) domain and a catalytic PI3K (phosphatidylinositol 3-kinase)-like kinase domain followed by a short C-terminal FATC (C-terminal FAT) domain. Besides these common structural features, a unique insertion region is located between the kinase domain and the FATC[46,47]. The SMG1-mediated UPF1 phosphorylation has been believed to be the definitive signal to trigger NMD[45].

The SMG8 and SMG9 factors were originally identified by interaction studies with human SMG1[27]. Homologues of SMG8 and SMG9 also exist in *C. elegans*. SMG8 and SMG9 have been reported as two regulators for SMG1 kinase[27,47]. The interactions between SMG8-SMG9 and SMG1 keep SMG1 kinase in the inactive state[27,47]. However, the mechanism for inhibition and activation of SMG1 kinase activity remains elusive due to the lack of high-resolution structural information. SMG8 was predicted to contain two conserved regions[27]. However, the boundary and function of each conserved region are unknown. SMG9 was predicted to contain a N-terminal unstructured region and a C-terminal NTPase domain[48]. Both the N-terminal region and the C-terminal domain of SMG9 are required for binding to SMG1[48]. SMG9 is the central molecule for the assembly of so-called SMG1C (SMG1-8-9 complex). SMG9 can bind SMG1 by itself or in the form of a preassembled SMG8-SMG9 complex. Binding of SMG8 to SMG1-SMG9 was reported to induce a large conformational change on SMG1, which is thought to cause the inhibition of the kinase activity of SMG1[47].

Finally, the NMD factors SMG5, SMG6 and SMG7 are recruited by phosphorylated UPF1. These proteins share a common phosphoserine-binding domain that adopts a similar fold to 14-3-3 proteins[49]. The 14-3-3-like domains of SMG5 and SMG7 interact with each other back to back forming a stable heterodimer and recognize the phosphorylated residues in the C-terminus of UPF1[50]. SMG5 and SMG7 can recruit protein phosphatase 2A (PP2A) to mediate dephosphorylation of UPF1[51]. SMG7 was recently shown to be involved in deadenylation by recruiting the CCR4-NOT deadenylase complex[52]. While SMG5 and SMG7 work together in a heterodimer, SMG6 functions as a monomer and contains distinct functional modules. It contains two N-terminal protein-protein interaction motifs (EBM)[53], a central TPR domain and a C-terminal PIN domain[54]. The SMG6 PIN domain adopts a similar fold to RNase H family endonuclease and mediates active endonucleolytic mRNA cleavage in the vicinity of NMD-inducing PTC[55-57]. In contrast, the C-terminal PIN-like domain of

SMG5 does not have endonuclease activity due to lacking the canonical motif of three aspartic acid residues[56]. Recent studies have shown that SMG6 is able to interact with UPF1 in a phosphorylation-dependent and in a phosphorylation-independent manner[54,58]. SMG6 and SMG5-SMG7 can bind UPF1 concomitantly[54]. SMG6 interacts with EJC through its N-terminal two conserved EJC binding motifs[53].

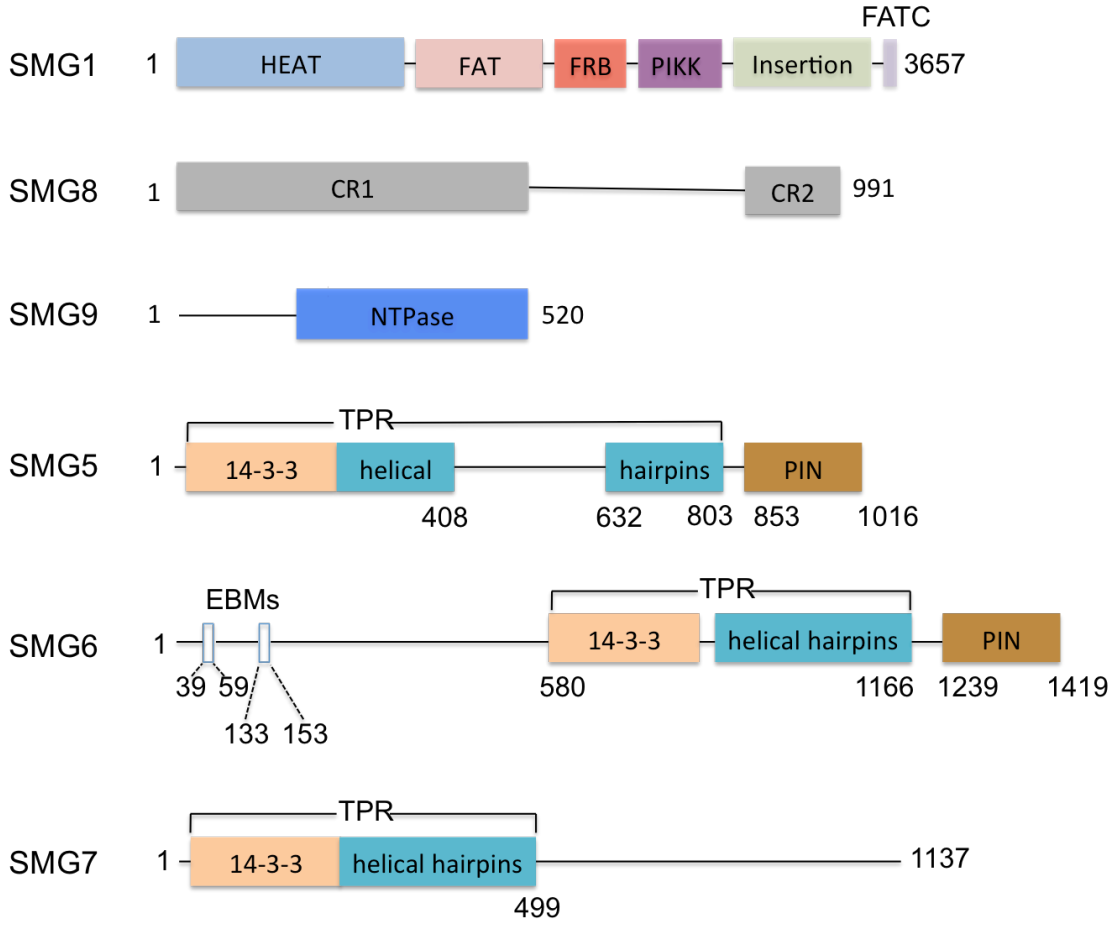


Figure 1.4: Domain organization of human SMG proteins[54]. HEAT: Huntingtin, elongation factor 3 (EF3), protein phosphatase 2A (PP2A), yeast kinase TOR1 domain; FAT: FRAP/TOR, ATM, and TRRAP; FRB: FKBP-rapamycin-binding domain; PIKK: phosphatidylinositol 3-kinase-related protein kinase domain; FATC: C-terminal FAT domain; CR: conserved region; PIN: PilT N-terminus domain; TPR: Tetratricopeptide repeat.

1.1.2.3 Exon junction complex (EJC)

In mammalian cells, the nuclear splicing machinery leaves a molecular mark on spliced mRNA[59]. This molecular mark is a multiprotein complex named exon junction

complex (EJC) deposited by spliceosomes during splicing onto mRNAs at a conserved position of 24nt upstream spliced junctions[59]. The EJC is a dynamic complex involved in various post-transcriptional processes, including splicing, transport, translation and NMD[60]. The EJC is transported together with its bound mRNA to cytoplasm, subsequently dissociates from mRNA during the first round of translation. The EJC is considered a strong enhancer of NMD in mammalian cells by providing a dynamic binding platform for the assembly of UPF complex, which is a part of the NMD complex when a ribosome stalls at the premature stop codon[60]. Despite the important role that EJC plays in NMD, the requirement for EJC in NMD is diverse in different eukaryotic organisms. There is no EJC in yeast while in *Caenorhabditis elegans*[61] and *Drosophila melanogaster*[62] the EJC is dispensable for NMD. The stable core of the EJC is composed of four proteins: eIF4AIII (eukaryotic initiation factor 4AIII), Mago, Y14 (also known as RNA-binding motif 8A) and Barentsz (Btz, also known as MLN51)[63,64]. Structural studies have revealed the assembly of EJC core and its interaction with RNA[65,66]. The EJC core binds RNA through eIF4AIII and Btz. eIF4AIII is a DEAD-box RNA helicase containing two RecA-like domains that adopt a closed conformation upon ATP binding and binds the sugar-phosphate backbones independently of bases. This conformation is stabilized through the interaction between two eIF4AIII RecA-like domains and Mago-Y14 heterodimer that prevents the conformational change induced by ATP hydrolysis thereby lock the helicase on the RNA. The EJC complex is further stabilized by Btz, which interacts with eIF4AIII through its two conserved regions.

1.1.2.4 Novel NMD factors

In addition to current well known NMD factors, a few novel NMD *trans*-acting factors have been identified by a variety of strategies including RNAi screen and Interactome-Mass spectrometry approach. However, the mechanisms underlying their functions remain to be investigated. With a GFP-reporter-based RNAi screen in *C. elegans*, two novel factors *smgl-1* and *smgl-2* were identified and they are conserved throughout evolution[61]. The human orthologues for *C. elegans smgl-1* and *smgl-2* are NBAS (Neuroblastoma amplified sequence) and DHX34 (DEAH box protein 34) respectively[61]. NBAS and DHX34 form part of an autoregulatory NMD circuit that regulates endogenous RNA targets in human cells as well as in *D. rerio* and *C.*

elegans[67]. DHX34 is a Superfamily 2 (SF2) DEAH-box RNA helicase. Previous work revealed that DHX34 directly interacts with SMG1 and UPF1 forming a complex that promotes UPF1 phosphorylation leading to functional NMD[68].

1.1.3 Current working models of NMD

NMD targets mRNAs containing a premature stop codon for rapid degradation. It has been recently reported that NMD not only can occur during the first round of translation but also can be triggered in the subsequent rounds of translation if the mRNA and mRNP features that define a stop codon as a PTC persist[69,70]. In the NMD research field, the critical questions have been "what are defining features that distinguish a PTC from normal translation termination codon?" and "when and how are NMD machinery assembled to initiate mRNA degradation? ". Despite lacking of a unified model that can explain NMD in all scenarios, currently two different working models of NMD have been proposed and widely accepted, the "EJC model" and the "faux 3'-UTR model"[1,28,71].

1.1.3.1 The exon junction complex (EJC) model

The EJC model is principally applicable to the NMD mechanism in mammalian cells. In mammals, NMD is a pathway that intimately linked with pre-mRNA splicing, which deposits the EJC complex 20 to 24 nucleotides upstream of an exon-exon junction. In the EJC model, the PTC that elicits efficient NMD needs to situate at least 50-55 nucleotides upstream the final exon-exon junction, which is the "50 nt rule"[72]. Normally all EJCs will be removed from mRNA by translation machinery after one round of translation while in the presence of a PTC that situates more than 50nt upstream of exon-exon junction, the EJC stays on mRNA and initiates successive events that leads to rapid mRNA degradation. The stalling ribosome on a PTC recruits NMD factors to form a so-called SURF complex containing UPF1, a RNA helicase and central molecule of NMD and SMG1, a PIKK kinase in its inactive form and its two associated factors SMG8 and SMG9, as well as eukaryotic release factor eRF1 and eRF3. Subsequently, the RNA helicase DHX34 functions as a bridging molecule and promotes the interaction between UPF1 and UPF2-UPF3-EJC complex to form a decay-inducing (DECID) complex in the vicinity of PTC, which leads to the

phosphorylation of UPF1 by SMG1 kinase and dissociation of eRF1 and eRF3. The cycle of phosphorylation and dephosphorylation of UPF1 is critical for NMD progression. SMG1 recognizes and phosphorylates serine and threonine residues that are next to a glutamine residue (S/TQ motifs), which are enriched at the C-terminal end of human UPF1. The phosphorylated UPF1 recruits phospho-binding proteins SMG6 and SMG5-SMG7 complex for endonucleolytic RNA cleavage and recruiting general RNA degradation factors, respectively. SMG6 is an active endonuclease and functions as a monomer that cleaves NMD targets in the vicinity of PTC and can interact with UPF1 in phospho-dependent and phospho-independent manner. The SMG5 and SMG7 form a stable heterodimer and bind to phosphorylated UPF1. SMG5-SMG7 complex also recruits protein phosphatase 2A (PP2A) and thus play a role in the dephosphorylation of UPF1. In addition, previous work shows that SMG5-SMG7 heterodimer directly recruits the CCR4-NOT deadenylase complex to NMD targets to induce the deadenylation-dependent decapping and subsequent RNA degradation. Upon the formation of DECID complex, UPF2 interacts with the inhibitory N-terminal CH domain of UPF1, induces a large conformational change on UPF1, thereby activates UPF1 helicase activity and leads to mRNP remodeling that allows the access of nucleases for RNA degradation[60].

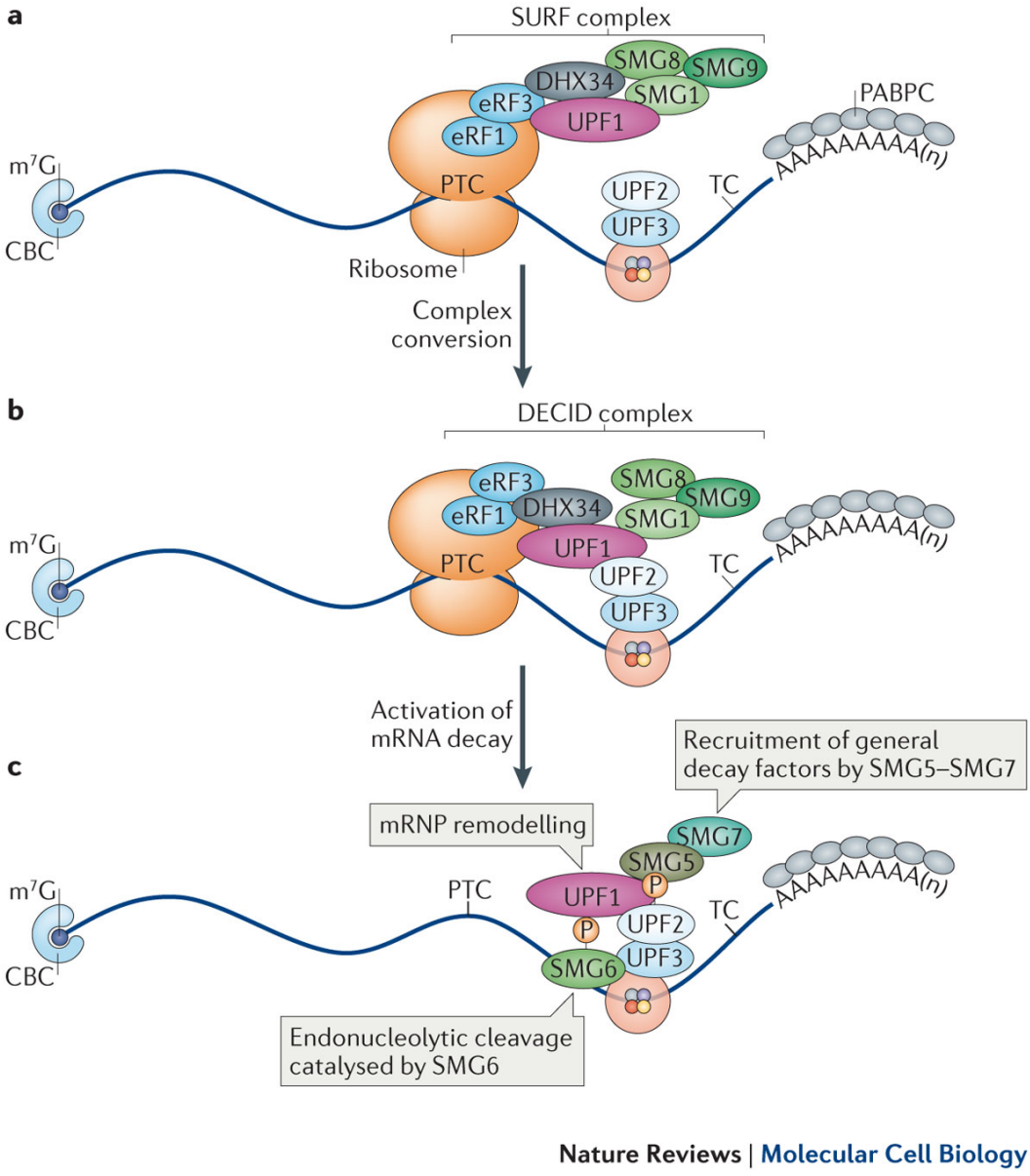


Figure 1.5: The functions of EJC in NMD[60] (figure adapted from Ref. 60). a. Ribosome stalling at a premature stop codon (PTC) promotes the formation of a so-called SURF complex containing SMG1-8-9, UPF1, eRF1 and eRF3. An EJC together with UPF3 and UPF2 are located downstream of SURF and upstream of a normal termination codon. b. Subsequent interaction between UPF1 and UPF2 promotes the formation of decay inducing complex (DECID) and leads to the phosphorylation of UPF1 by kinase SMG1. c. UPF2 activates UPF1 and leads to mRNP remodeling and recruitment of SMG6 for endonucleolytic cleavage of mRNA near the PTC and SMG5-SMG7 complex for recruitment of general decay factors.

1.1.3.2 Faux 3'-UTR model

Although the EJC-marker model is generally applicable to NMD in mammalian cells, the EJC is dispensable for PTC recognition in other eukaryotic organisms such as *S. cerevisiae, C. elegans,* and *D. melanogaster*[45]. It was first observed in yeast cells that the long 3'-UTR or faux 3'-UTR leads to less efficient ribosome dissociation as compared to normal 3'-UTR and triggers NMD[73]. The faux 3'-UTR model is subsequently supported by results in other systems including fly, plant and human cells[21,74 75]. The 3'-UTR is defined by the translation termination codon, which is recognized by the ribosome and release factors and the poly (A) tail which is coated by poly (A) binding proteins PABPs (PABPC1, PABPC4, and PABPN1). The C-terminal part of PABPC1 (Pab1p in *S. cerevisiae*) can interact with eRF3[76 77]. NMD can be inhibited by PABPC1 tethered in proximity downstream of a premature stop codon[78 75 79] suggesting that PABPC1 promotes correct translation termination. In addition, UPF1 has also been reported to interact with eRFs to form a so-called SURF complex[80]. The competition for binding to eRF3 between PABPC1 and UPF1 is thought to be the trigger for long 3'-UTR mediated PTC recognition[78 79 81 75]. In the case of long 3'-UTR, the dissociation of ribosome from a PTC might be less efficient caused by UPF1-mediated inhibition of interaction between eRF3 and PABP.

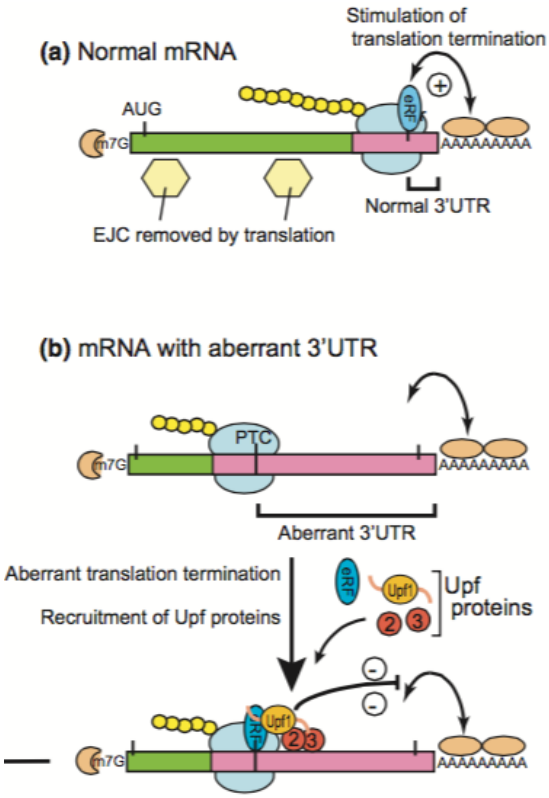


Figure 1.6: PTC recognition in faux 3'-UTR model[45] (figure adapted from Ref. 45). (a). Normal translation termination is stimulated by the interaction between eRF3 and PABP. (b). Aberrant 3' UTR decreases the efficiency of eRF3 binding to PABP and causes the recruitment of UPF proteins to the stalled ribosome to initiate NMD.

1.1.3.3 Redefinition of NMD

The classical definition of NMD has been changed during the last decade given the fact that NMD not only targets PTC-containing mRNAs for degradation but also influences the steady-state levels of a significant amount of normal mRNAs or PTC-free mRNAs. Thus, based on the current mechanistic models, the emerging point is that (reviewed in Ref. 1) NMD can be defined as an mRNA degradation pathway that requires a number of well-characterized NMD factors and targets transcripts that fail to terminate translation efficiently at their stop codons. The kinetic competition between efficient translation termination and the assembly of a degradation-triggering NMD complex determines the fate of mRNA[1].

1.1.4 Regulation of SMG1 kinase activity by SMG8 and SMG9

It is known that phosphorylation is the most common protein post-translation modification that plays essential roles in almost all cellular functions. The phosphorylation of UPF1 is believed to be the key event and the triggering signal in NMD to initiate mRNA degradation. SMG1 is a PIKK kinase that phosphorylates UPF1 and its two regulators, SMG8 and SMG9, regulate the kinase activity of SMG1. SMG8 and SMG9 interact with SMG1 to form the so-called SMG1C complex[27]. Previous studies using Electron Microscopy (EM) has revealed the 3D molecular architecture of human SMG1C complex[47]. The N-terminal region of SMG1 containing HEAT repeats provides the binding platform for SMG8 and SMG9. SMG9 can form a complex with SMG1, which requires both N-terminal unstructured region and C-terminal domain of SMG9[48]. SMG8 is recruited by SMG9 to SMG1 and binding of SMG8 to SMG1-SMG9 complex induces large conformational changes, which is thought to down-regulate SMG1 kinase activity[47].

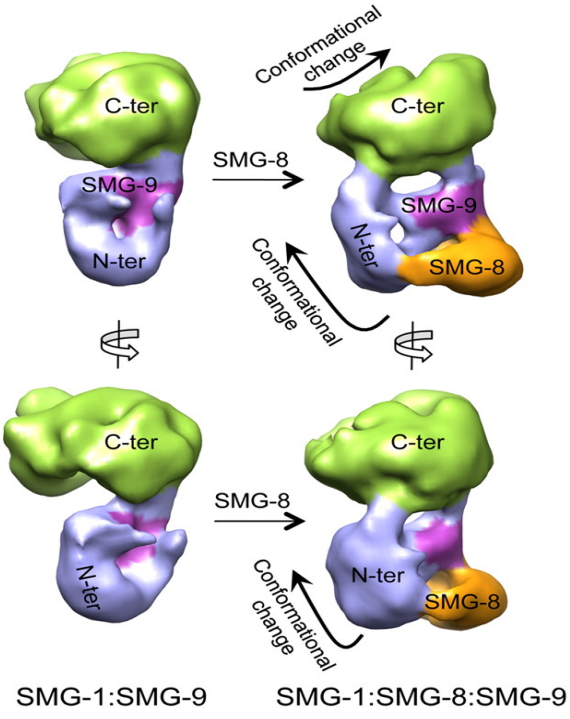


Figure 1.7: Views of the 3D structure of SMG1C compared with the compatible view in SMG1-SMG9[47] (figure adapted from Ref. 47). Binding of SMG8 to SMG1-SMG9 complex induces large conformational changes in SMG1.

2 The aim of the thesis

Nonsense mediated mRNA decay is a conserved and critical cellular surveillance pathway for eukaryotes for the elimination of PTC containing mRNAs to prevent the production of C-terminally truncated proteins, the accumulation of which is harmful for normal cellular functions and will cause diseases. The mechanism of NMD has been extensively investigated for over three decades. Many *trans*-acting factors involved in NMD have been identified by various approaches. In order to understand how these NMD players work together to execute and regulate this critical pathway, structural characterization of individual NMD proteins as well as complexes formed by interacting proteins has been played an important role.

UPF1 is the most conserved NMD factor in eukaryotes from yeast to humans and it is also the central molecule in NMD. The phosphorylation of UPF1 at its C-terminal low-complexity region is believed to be the definitive step to trigger NMD. SMG1 is a PIKK kinase that phosphorylates UPF1 in NMD. It was reported that SMG8 and SMG9 are two regulators of SMG1 kinase. Although the electron microscopy studies of SMG1-SMG8-SMG9 complex in recent years provided the information for the architecture of this complex, the structural basis for the regulation of SMG1 kinase by SMG8 and SMG9 is still elusive due to the low resolution structure obtained from electron microscope that have been published.

At the moment, not much is known about SMG8 and SMG9 in terms of their structure and functions. Many questions still can be asked such as which family of proteins does SMG8 and SMG9 belong to and how SMG8 interacts with SMG9 and how their function as regulators of SMG1 kinase is achieved by their structures.

The aim of this thesis is to characterize the structure and function of SMG8-SMG9 complex with combined approaches of crystallography, biochemistry and biophysics. With the structure of SMG8-SMG9 complex, we could expect to gain insights on the mechanism of regulation of SMG1, the kinase that is responsible for UPF1 phosphorylation.

3 Materials

3.1 Cloning and expression strains

Strain/Cell lines	Description
XL1-Blue	Electrocompetent bacteria cell for plasmid reproduction.
BL21 (DE3) Gold pLysS	Electrocompetent bacteria cell for protein expression.
DH10MultiBac	Electrocompetent bacteria cell for recombinant bacmid production.
Sf21	Insect cell for recombinant baculovirus production and protein expression.
High Five	Insect cell for recombinant protein expression.

3.2 Constructs

Construct	Description
CeSMG8 (1-423)	*C.elegans* SMG8 N-terminal fragment (1-423) in pEC-K-3C-GST vector
CeSMG9 (1-385)	*C.elegans* full-length SMG9 in pEC-A-3C-Trx vector
CeSMG9 (39-385)	*C.elegans* SMG9 C-terminal fragment (39-385) in pEC-A-3C-Trx vector
CeSMG9 (60-385)	*C.elegans* SMG9 C-terminal fragment (60-385) in pEC-A-3C-Trx vector
CeSMG8 (1-423)/ CeSMG9 (39-385)	*C.elegans* SMG8 N-terminal fragment (1-423) and *C.elegans* SMG9 C-terminal fragment (39-385) were subcloned in MCS1 and MCS2 respectively in pFL vector, SMG8 is C-terminally his-tagged.

CeSMG8 (1-423)/ CeSMG9 (59-385)	*C.elegans* SMG8 N-terminal fragment (1-423) and *C.elegans* SMG9 C-terminal fragment (59-385) were subcloned in MCS1 and MCS2 respectively in pFL vector, SMG8 is C-terminally his-tagged.
CeSMG8 (1-423)/ CeSMG9 (59-375)	*C.elegans* SMG8 N-terminal fragment (1-423) and *C.elegans* SMG9 C-terminal fragment (59-375) were subcloned in MCS1 and MCS2 respectively in pFL vector, SMG8 is C-terminally his-tagged.

3.3 DNA oligonucleotides

All DNA oligonucleotides were purchased from Sigma-Aldrich and treated as recommended by manufacturer. All sequences are given in 5′ -3′ direction.

Oligo name	Sequence	Oligo ID
CeSMG9 1FW 3C	CCAGGGGCCCGACTCGATGATGAAAAAAGTGGAAATTCT	oLL40
CeSMG9 385RV 3C	CAGACCGCCACCGACTGCTTAGCTAAAAAATTTGTTCGCGT	oLL41
CeSMG9 385RV	ATGCGTCGACTCAGCTAAAAAATTTGTTCGCGT	oLL48
CeSMG8 1FW	ATGCCCCGGGATGGACATAGCTAAATGGGT	oLL49
CeSMG8 1FW 3C	CCAGGGGCCCGACTCGATGATGGACATAGCTAAATGGGT	oLL55
CeSMG8 423RV 3C	CAGACCGCCACCGACTGCTTAACGCATATCGGATTGCCAC	oLL61
CeSMG9 39FW 3C	CCAGGGGCCCGACTCGATGCCGGTGGCCGATGACGTGGC	oLL71
CeSMG9 60FW 3C	CCAGGGGCCCGACTCGATGAAGGAGTCTGTGCGATTTTT	oLL72
CeSMG8 423RV	ATGCGCTAGCTCAGTGGTGGTGGTGGTGGTGACGCATATCGGATTGCCAC	oLL74
CeSMG9 39FW	ATGCGGATCCATGCCGGTGGCCGATGACGT	oLL75

CeSMG9 59FW	ATGCGGATCCATGAAGGAGTCTGTGCGATT	oLL81
CeSMG9 375RV	ATGCGTCGACTCAATGGAATCCATTGTCAA AATG	oLL82

3.4 Vectors

Vector name	Description
pEC-K-3C-GST	House made, Kanamycin resistance, 3C cleavage site, N-terminal His-GST tag
pEC-A-3C-Trx	House made, Ampicillin resistance, 3C cleavage site, N-terminal His-Trx tag
pFL-deltaSpeI	From Imre Berger lab

3.5 Enzymes

Enzyme	Source
Phusion Polymerase	Finnzymes
Taq polymerase	MPIB core facility
T4 DNA Polymerase	NEB
Restriction endonucleases	NEB
T4 DNA Ligase	MPIB core facility
Trypsin, Chymotrypsin, Elastase, Glu C, Substilisin	Roche

3.6 Chemicals and reagents

All common chemicals and reagents were purchased from Sigma-Aldrich, Fluka or Hampton unless otherwise stated.

3.7 Kits

Kit	Supplier
Qiaquick Gel Extraction	Qiagen
Qiaquick Spin Miniprep	Qiagen
Crystallization screen Kits	Hampton Research/Qiagen

3.8 Buffers and Media

Media	Component
Luria-Bertani (LB) (Miller, 1972)	1% (w/v) bacto tryptone 0.5% (w/v) bacto yeast extract 170 mM NaCl Adjust pH to 7.6 with NaOH
LB agar plate	1.5 % (w/v) bacto agar in LB Antibiotics in respective appropriate concentrations
SOC Medium	2% (w/v) bacto tryptone 0.5% (w/v) bacto yeast extract 10 mM NaCl 1 mM $MgCl_2$ 2.5 mM KCl 10 mM $MgSO_4$ 0.4% glucose Adjust pH to 7.2
Terrific Broth (TB)	1.2 % bacto tryptone 2.4 % bacto yeast extract 0.4 % glycerol ddH_2O to 900 ml 0.017 M KH2PO4 0.072 M K2HPO4

X-Gal/IPTG LB Agar Plates	1.5 % (w/v) bacto agar in LB 50 µg/ml Kanamycin, 7 µg/ml Gentamycin, 10 µg/ml Tetracyclin 40 µg/ml X-gal 0.15 mM IPTG
Sf-900™ II SFM insect cell culture medium	Purchased from company life technologies
ESF 921 insect cell Culture medium, protein free	Purchased from company Expression Systems

Buffers used for purification and assays are indicated in the Methods in Chapter 4.

Buffer	Component
TBE (20 ×)	1 M Tris 0.89 M Boric acid 20 mM EDTA pH 8.0
SDS-PAGE running buffer (10 ×)	0.25 M Trizma Base 1.92 M Glycine 1% SDS
SDS loading buffer (2 ×)	100 mM Tris pH 6.8 10% β-Mercaptoethanol 4% SDS 0.2% Bromophenol Blue 20% Glycerol

3.9 Equipment

Equipment	Model	Manufacturer
PCR machine	Mastercycler	Eppendorf

Electroporator	Gene Pulser/Micro Pulser	Bio-Rad
Electro-cuvette	Gene pulser 0.1 cm electrode gap	Bio-Rad
Bacteria shaker	KS-15/Climo-Shaker ISF1X	Kühner
Cell lysis sonicator	Sonifyer VS70T/VS72T	Bandelin Electronics
Insect cell shaker	Kuehner shaker	Kuehner
Laminar flow hood	Holten LaminAir	Thermo electron corporation
Cell viability analyzer	Vi-Cell-XR	Beckman Coulter
Dounce homogenizer	Borosilicate Glass	Fisher Scientific
Centrifuge	Avanti J-20 XP Micro centrifuge 5417C & 5810	Beckman Coulter Eppendorf
Chromatography columns	HisTrap 5 mL HiTrap Heparin 5mL Mono Q 1mL Superdex 200 HiLoad 200	GE Healthcare
Chromatography FPLC	ÄKTA Purifier	GE Healthcare
Nanodrop spectrophotometer	Nanodrop	PeqLab
Crystallization pipetting robot	Phoenix	Art Robbins Instruments
Crystallization visualization system	Xtal-Focus	ExploraNova La Rochelle
X-ray diffractometer	PX scanner D8 venture	Rigaku Bruker
pH meter	Lab860	Schott
Pipettes	Eppendorf Research	Eppendorf

Gel imaging	Gel visualization	Roth
Thermo shaker	Thermomixer comfort	Eppendorf
Vortex mixer	Vortex-Genie	Scientific Industries
Fluorescence spectrometer	Infinite M1000 Pro	Tecan

3.10 X-ray sources and synchrotron facility

Crystals were tested with in-house X-ray diffractometer PX-scanner (Rigaku) and D8 venture (Bruker). Diffraction data sets were collected at Swiss Light Source (SLS) at the Paul Scherrer Institute (PSI), Villigen, Switzerland and the Deutsches Elektronen-Synchrotron (DESY) PETRA III in Hamburg, Germany.

3.11 Software and web servers

The following software and web servers were used for the data processing and analysis, figure generation and thesis writing.

CLC Sequence Viewer (http://www.clcbio.com/products/clc-sequence-viewer/)
CodonCode Aligner (http://www.codoncode.com/aligner/)
Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2)
PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/)
UNIPROT (http://www.uniprot.org)
ProtParam (http://web.expasy.org/cgi-bin/protparam)
Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/)
Jalview (http://www.jalview.org/)
Hampton Research Make-tray tool (http://hamptonresearch.com/make_tray.aspx)
XDS (http://xds.mpimf-heidelberg.mpg.de)
SHELX (http://shelx.uni-ac.gwdg.de/SHELX/)
Phenix (http://www.phenix-online.org)
Coot (http://www2.mrc-lmb.cam.ac.uk/Personal/pemsley/coot/)

Molprobity (http://molprobity.biochem.duke.edu)
PISA (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html)
Pymol (http://pymol.org)
Chimera (http://www.cgl.ucsf.edu/chimera/)
Origin (www.originlab.com)
Adobe Illustrator (www.adobe.com/products/illustrator)
EndNote X7 (http://endnote.com)
Microsoft Office (www.microsoft.com)

4 Methods

4.1 Cloning

4.1.1 PCR

DNA sequences of interested were amplified by standard PCR procedure as follows.

Reaction mix (50 µl)
30 ng template DNA
1.5 µl Forward primer (10 µM)
1.5 µl Reverse primer (10 µM)
1 µl dNTPs (10 mM stock)
1 µl Phusion polymerase (0.5 u/µl)
10 µl reaction buffer (5 ×)
ddH_2O

The following cycling conditions were used:

Initial denaturation,	95 °C,	5 min,	1 cycle
Denaturation	95 °C,	30 s,	
Annealing	55 °C,	30 s,	30 cycles
Extension	72 °C,	up to DNA length and polymerase	
Final extension	72 °C,	10 min	1 cycle

In some cases, touch down PCR was used, which decreases annealing temperature of 1 °C per cycle.

4.1.2 Agarose gel electrophoresis

Agarose gel was used to examine PCR amplification and extract the product of interest. 1% agarose gel was prepared in 1 × TBE. SYBR Safe stock (Invitrogen) was added in 1: 10000 dilutions as dye to the agarose solution for visualization. DNA samples were mixed with 6 × loading buffer (Orange Dye, Fermentas) with the volume

ratio of 5:1. The electropheris was performed in 1 × TBE buffer. DNA was visualized with an UV transilluminator.

4.1.3 Purification of DNA fragments

PCR products were examined on agarose gels. Bands of interest were cut from the gel and purified with the gel extraction kit (Qiagen) following the manufacturer's protocol. The DNA fragments produced by restricted digestion were directly purified with the gel extraction kit with the manufacturer's protocol. All DNA fragments were eluted in ddH_2O.

4.1.4 LIC cloning

The ligation independent cloning (LIC) was used for making *E. coli* expression constructs.

4.1.4.1 Principle

The LIC system utilizes the 3' to 5' exonuclease activity of T4 DNA polymerase to generate 12-15 bp overhangs on both the vector and a PCR product. These overhangs are long enough to stick together during transformation such that no ligation step beforehand is needed. The *E. coli* machinery repairs nicks during transformation.

4.1.4.2 Insert processing

PCR products were processed with the following system in a total volume of 20 µl. The reaction mix was incubated at room temperature for 30 minutes and the enzyme was inactivated at 75 °C for 20 minutes.

Reaction mix	
Gel purified PCR product	600 ng
T4 DNA Polymerase buffer (10 ×)	2 µl
dATP (25 mM)	2 µl
DTT (100 mM)	1 µl
T4 DNA Polymerase (Novagen) (3u/µl)	0.4 µl
H_2O	to 20 µl

4.1.4.3 Vector processing

Vectors were first linearized by digestion or by PCR. Digest 2 µg vector with 60 u SacII (or 20 u ZraI for 3C LIC vectors) in 100 µl reaction volume, load 250 ng per lane of a 0.8% Agarose gel, run at least until slow dye has passed the first half of the gel and cut the bands carefully avoiding uncut vector and purify them with the Qiaquick purification kit. The linearized vector was further processed by T4 DNA polymerase in the presence of dTTP in the following reaction mix. The mix was incubated for 30 minutes at room temperature and the enzyme was inactivated at 75 °C for 20 minutes.

Reaction mix	
linearized vector	450 ng
T4 DNA Polymerase buffer (10 ×)	3 µl
dTTP (25 mM)	3 µl
DTT (100 mM)	1.5 µl
T4 DNA Polymerase (Novagen) (3u/µl)	0.6 µl
H_2O	to 30 µl

4.1.4.4 Annealing reaction

2 µl of processed insert and 1 µl of processed vector were mixed and incubated for 10 minutes at room temperature. In next step, 1 µl EDTA (25 mM) was added to the mix and incubated for 10 minutes at room temperature. 2 µl of mix was transformed in competent *E. coli.*

4.1.5 Restriction digest and ligation

Cloning of constructs for insect cell expression was done by using classical restriction digest and ligation. DNA fragments from PCR and vectors were digested by appropriate restriction endonucleases (NEB) in a system of 20 µl. 1 µg of PCR product or vector and 0.5 µl of each enzyme (20 u/µl) together with 2 µl of appropriate NEB reaction buffer (10 ×) were added to the reaction mix. The digestion was carried out at 37 °C for 5 hours. The products of digestion were purified with Qiaquick Gel Extraction kit (Qiagen). The insert and linearized vector were further ligated in 15-µl

reaction system containing 1 µl of T4 DNA ligase (400 u/µl), 1.5 µl of T4 DNA ligase buffer (10 ×), 11 µl of insert and 1.5 µl of linearized vector. The ligation reaction was performed at room temperature for 2 hours.

4.1.6 Transformation

Appropriate amount of plasmid or ligation product were added to the thawed electro-competent cells (50 µl) on ice and transformed using the electroporation method. Cells were transferred to 0.1 cm electroporation cuvette and an electrical pulse of 1.8 kV was applied. Cell suspension was mixed with 200 µl SOC medium and then transferred to a 1.5 ml-eppendorf tube and incubated in a thermomixer at 37 °C for 50 min, 1000 rpm. Cells were then plated on a LB agar plate with appropriate antibiotics and incubated overnight at 37 °C.

4.1.7 Plasmid purification

In order to extract recombinant plasmids, single colony was picked from LB agar plate and inoculated into 5 ml LB culture with appropriate antibiotics. The culture was shaken in a shaker at 37 °C overnight with the speed of 220 rpm. Bacteria cells were pelleted by centrifugation with 4000 rpm for 10 minutes. Plasmids were purified with Qiaquick Miniprep kit and eluted in ddH_2O in a volume of 50 µl. The concentration of plasmid was measured with the Nanodrop spectrophotometer.

4.1.8 DNA sequencing

DNA sequencing was done in the core facility of Max Planck Institute of Biochemistry or in the company Eurofins.

4.2 Expression of recombinant proteins

Recombinant proteins were expressed in both *E. coli* and insect cells with different procedures.

4.2.1 Expression of recombinant proteins in *E. coli*

Proteins were either singly expressed or co-expressed by the means of co-transformation in BL21 (DE3) Gold pLysS under the control of T7 bacteriophage transcription. Cells were grown in TB medium with appropriate antibiotics at 37 °C with constant shaking at the speed of 220 rpm. After the cell density OD_{600} reached between 1 and 2, the temperature was reduced to 18 °C and 0.3 mM IPTG was added to the culture to induce protein expression. Cells were harvested approximately 16 hours post-induction by centrifugation (6000 rpm, 10 minutes). The cell pellets were either immediately lysed for purification or frozen with liquid nitrogen and stored at -80 °C.

4.2.2 Expression of recombinant proteins in insect cell

C. elegans SMG8 and SMG9 were subcloned in MCS1 (multiple cloning site 1) and MCS2 (multiple cloning site 2) respectively in the same pFL vector for co-expression. The recombinant plasmid was then transformed into DH10MultiBac competent cell and the recombinant Bacmid was selected with blue-white screening method. Insect cells sf21 and High Five were used for virus production and protein expression respectively.

4.2.2.1 Blue-white screen

Generally 300 ng of the recombinant plasmid was transformed in 50 µl of DH10MultiBac competent cell with electroporation method. The transformed cell suspension was mixed with 800 µl of SOC medium and incubated at 37 °C with constant shaking at 1000 rpm in a thermomixer for 4 hours. Appropriate volume of cell suspension (generally 100 µl) was plated on a X-Gal/IPTG LB Agar plates. The plates were placed at 37 °C for 48 hours. Single white colony containing composite Multibac DNA was picked and inoculated into 2 ml of LB culture containing Kanamycin and Gentamycin for overnight growth at 37 °C.

4.2.2.2 Bacmid isolation

Bacmid DNA was isolated by alkaline lysis, using the Qiaprep Spin Miniprep kit with the following protocol:

a. Resuspend cell pellet in 250 µl Buffer P1 and transfer to a microcentrifuge tube.
b. Add 250 µl Buffer P2 and mix thoroughly by inverting the tube 4-6 times.
c. Add 350 µl Buffer P3 and mix by inverting.
d. Centrifuge for 10 min at maximum speed (e.g. 16.100 × g). Transfer supernatant to a fresh tube.
e. Repeat step d to remove all residual precipitation.
f. Add 800 µl iso-propanol, invert tube a few times to mix and place on ice for 10 min to precipitate DNA.
g. Centrifuge sample for 15 min at maximum speed at room temperature.
h. Locate the DNA pellet and carefully remove supernatant without disturbing the pellet. Add 70% ethanol and invert the tube several times to wash the pellet.
i. Centrifuge for 5 min at maximum speed at room temperature.
j. Remove as much supernatant as possible. Air-dry pellet for 5-10 min in a sterile environment (e.g. in a Petri dish).
k. Dissolve DNA pellet in 40 µl sterile H_2O by gently tapping the bottom of the centrifuge tube. Measure concentration after 10 min.

4.2.2.3 Transfection

The isolated bacmid DNA was then transfected into Sf21 cells to generate P1 virus (initial virus) with the following protocol:
a. For every bacmid DNA, put 2 ml 0.8×10^6 freshly diluted Sf21 cells in two wells each of a 6-well tissue culture plate and incubate for 30 min at 27 °C.
b. Gently mix 1 µg bacmid DNA with 8 µl PEI transfection agent (1 mg/ml aqueous solution, sterile filtered) in 200 µl of serum free medium (e.g. Sf-900 II SFM). Incubate for 15 min at 27 °C.
c. Add bacmid mix drop by drop to wells with adherent cells in 6-well plate. Seal plate with lid or parafilm and incubate at 27 °C.
d. After 3-5 days, harvest virus-containing supernatant by aspirating with a pipet and store in 15 ml Falcon tubes at 4 °C, protected from light.

4.2.2.4 Virus amplification

P1 virus produced from transfection was amplified to higher titer P2 and P3 virus for protein expression with the following protocol.

a. Add 1 ml P1 virus to 25 ml freshly diluted Sf21 cells at 0.5 × 10^6 cells/ml in a 250 ml shaker flask. Incubate cells on the shaker (80 rpm) at 27 °C.
b. After 48 hours, monitor cell count. Cells should have doubled approximately and started to swell. Split to below 1 × 10^6.
c. After additional 48 hours, collect supernatant and store at 4 °C in the dark. This is P2 virus.
d. Add 0.1 % (v/v) of P2 virus to 200 – 500 ml freshly diluted Sf21 cells at 0.5 × 10^6 cells/ml in a 2 L shaker flask (Fernbach type). Incubate cells on the shaker (80 rpm) at 27 °C and continue as in step b and c. This is P3 virus.

4.2.2.5 Large-scale expression

Appropriate amount of P3 virus (based on small scale expression test) was added to freshly diluted 500 ml High-Five cells at 1 × 10^6 cell/ml in each 2 L shaker flask. Incubate cells on the shaker (80 rpm) at 27 °C. After 72 hours, harvest cells by centrifugation (2000 rpm, 20 min).

4.2.2.6 Expression of selenium - methionine derivatized protein

In order to express selenium-methionine derivatized protein, normal medium and ESF 921 medium (purchased from Expression System) were mixed at the ratio of 2: 1 in a total volume of 300 ml to adapt High-Five cells. After 72 hours, cells were diluted to 0.5 × 10^6 cell/ml in 600 ml ESF 921 medium (300 ml in each flask). After additional 72 hours, combine two flasks and count cell density. Cells were diluted to 1 × 10^6 cell/ml in 500 ml ESF 921 medium in each of 6 flasks and cells were infected with P3 virus. L-selenium-methionine was added to cell culture at a final concentration of 0.1 mg/ml at three time points (18, 25, 45 hours) after infection. Cells were harvested 60 hours post-infection.

4.3 Purification of recombinant proteins

4.3.1 Purification of *C. elegans* full-length SMG9 (1-385)

Bacterial pellets from 1 L TB cell culture were resuspended in lysis buffer containing 50 mM Tris PH 8.0, 300 mM NaCl, 10% Glycerol and 25mM imidazole. Cells were lysed by sonication for 10 min (40% amplitude, 0.5s on/0.5s off). Supernatant was collected after centrifugation at 25,000 rpm for 45 min at 10 °C and filtered using 5 μm filter (Merck Millipore). The supernatant was loaded into a 5 ml HisTrap Ni-NTA column pre-equilibrated with 5-column volume of lysis buffer. The column was then washed with 250 ml washing buffer containing same ingredients as in lysis buffer. Protein was eluted with 45 ml elution buffer containing 50 mM Tris PH 8.0, 300 mM NaCl, 10% Glycerol, and 500 mM imidazole. The eluate was added with his-tagged 3C protease (1 ml, 1 mg/ml in stock) and subjected to dialysis overnight in 2 L dialysis buffer containing 50 mM Tris PH 8.0, 100 mM NaCl, 10% Glycerol. The protein solution after dialysis was reloaded into Ni-NTA column pre-equilibrated with dialysis buffer to remove cleaved his-Trx tag. The flow-through from this step was collected and loaded into mono Q column on ÄKTA purifier system and run ion-exchange chromatography with Buffer A containing 50 mM Tris PH 8.0, 100 mM NaCl, 10% Glycerol and buffer B containing 50 mM Tris PH 8.0, 1 M NaCl, 10% Glycerol. The protein of interest was purer in flow-through than in eluted fractions. The flow-through was concentrated and subjected to gel-filtration chromatography with gel-filtration buffer containing 50 mM Tris PH 8.0, 150 mM NaCl, 10% Glycerol on ÄKTA purifier system.

4.3.2 Purification of *C. elegans* SMG9 C-terminal fragment (39-385)

The expression and purification of *C. elegans* SMG9 C-terminal fragment (39-385) followed same protocol as for *C. elegans* full-length SMG9 (1-385), except in the purification of CeSMG9 (39-385), the flow-through from mono Q was reloaded to the same mono Q column (equilibrated with buffer A) again to remove impurities. The flow-through from second mono Q run was concentrated and subjected to gel-filtration chromatography with gel-filtration buffer containing 50 mM Tris PH 8.0, 300 mM NaCl, 10% Glycerol.

4.3.3 Purification of *C. elegans* SMG9 C-terminal fragment (59-385)

The expression and purification of CeSMG9 (59-385) followed same protocol as for *C. elegans* full-length SMG9 (1-385), except in the purification of CeSMG9 (59-385), instead of the flow-through, the fractions of peak from mono Q were concentrated and subjected to gel-filtration chromatography.

4.3.4 Purification of *C. elegans* SMG8 N-terminal fragment (1-423) in complex with *C. elegans* SMG9-FL

N-terminal his-GST tagged CeSMG8 (1-423) and N-terminal his-Trx tagged CeSMG9-FL were co-transformed in BL21 (DE3) Gold pLysS and expressed in 4 liters of TB cell culture. Cells were lysed with previously described protocol in lysis buffer containing 50 mM Tris PH 8.0, 300 mM NaCl, 10% Glycerol. The supernatant was loaded on a gravity-flow column manually packed with 5ml Glutathione Sepharose resins. The column was then washed with 20-column-volume lysis buffer and eluted with 45 ml elution buffer containing 50 mM Tris PH 8.0, 300 mM NaCl, 10% Glycerol, 20 mM GSH. The eluate was added with his-tagged 3C protease (1ml, 1 mg/ml in stock) and subjected to dialysis overnight in 2 L dialysis buffer containing 20 mM Tris PH 8.0, 100 mM NaCl, 10% Glycerol. The protein solution was then loaded on mono Q column to run ion-exchange chromatography with buffer A containing 50 mM Tris PH 8.0, 100 mM NaCl, 10% Glycerol and buffer B containing 50 mM Tris PH 8.0, 1 M NaCl, 10% Glycerol. The fractions of peak were then concentrated and subjected to gel-filtration chromatography with gel-filtration buffer containing 20 mM Tris PH 8.0, 150 mM NaCl.

4.3.5 Purification of *C. elegans* SMG8 N-terminal fragment (1-423) in complex with *C. elegans* SMG9 C-terminal fragment (39-385)

Insect cell pellets from 1 L cell culture were resuspended with lysis buffer containing 20 mM Tris PH 8.0, 150 mM NaCl and 20 mM imidazole. Cells were lysed with Dounce homogenizer on ice. The lysate was centrifuged at 4000 rpm for 15 minutes. The supernatant was then centrifuged at 25,000 rpm for 1 hour. The supernatant from this step was filtered with a 5 μm filter and then loaded to Ni-NTA column pre-equilibrated with 5-column-volumn washing buffer containing 50 mM Tris PH 8.0,

300 mM NaCl, 10% glycerol and 25 mM imidazole. The column was then washed with 200 ml washing buffer. Proteins were eluted with 40 ml elution buffer containing 20 mM Tris PH 8.5, 100 mM NaCl and 400 mM imidazole. The eluate was loaded to mono Q to run ion-exchange chromatography with buffer A containing 20 mM Tris PH 8.5, 100 mM NaCl and buffer B containing 20 mM Tris PH 8.5, 1 M NaCl. The flow-through and eluted fractions from mono Q were collected and subjected to gel-filtration chromatography with gel-filtration buffer containing 10 mM Hepes pH 7.2, 150 mM NaCl.

4.3.6 Purification of *C. elegans* SMG8 N-terminal fragment (1-423) in complex with *C. elegans* SMG9 C-terminal fragment (59-385)

The expression and purification of *C. elegans* SMG8 (1-423) in complex with *C. elegans* SMG9 (59-385) followed the same protocol as for the complex of *C. elegans* SMG8 (1-423) and *C. elegans* SMG9 (39-385) as described in 4.3.5.

4.3.7 Purification of *C. elegans* SMG8 N-terminal fragment (1-423) in complex with *C. elegans* SMG9 C-terminal fragment (59-375)

Insect cell pellets from 3 liters of cell culture were resuspended with lysis buffer containing 25 mM Tris PH 8.0, 300 mM NaCl and 20 mM imidazole. Cells were lysed with Dounce homogenizer on ice. The lysate was centrifuged at 4000 rpm for 15 minutes. The supernatant was then centrifuged at 25,000 rpm for 1 hour. The supernatant from this step was filtered with a 5 μm filter and then loaded to Ni-NTA column pre-equilibrated with 5-column-volumn washing buffer containing 25 mM Tris PH 8.0, 300 mM NaCl and 25 mM imidazole. The column was then washed with 200 ml washing buffer. Proteins were eluted with 45 ml elution buffer containing 25 mM Tris PH 8.0, 300 mM NaCl and 400 mM imidazole. The eluate was dialyzed in buffer containing 20 mM Tris PH 8.0, 100 mM NaCl overnight at 4 °C. The protein solution was then loaded on a 5 mL HiTrap Heparin column for ion-exchange chromatography with buffer A containing 20 mM Tris PH 7.5, 100 mM NaCl and buffer B containing 20 mM Tris PH 7.5, 1 M NaCl. The flow-through and pure fractions from Heparin run were collected and subjected to gel-filtration chromatography with buffer containing 25 mM Tris PH 8.0, 300 mM NaCl and 2 mM DTT.

4.3.8 Purification of selenium-methionine derivatized *C. elegans* SMG8 (1-423) in complex with *C. elegans* SMG9 (59-375)

The purification protocol for selenium-methionine derivatized *C. elegans* SMG8 (1-423) in complex with *C. elegans* SMG9 (59-375) was same as described in 4.3.7, except 4 mM β-ME was added in lysis buffer and buffers for Ni-affinity purification and 2 mM DTT was added to buffers for subsequent purification steps. All buffers were degased.

4.4 SDS-PAGE

SDS-PAGE was used to examine the purity of protein of interest. 12.5% and 15% SDS-gel were made with following recipe.

Reagent	5% Stacking gel (ml)	12.5% Resolving gel (ml)	15% Resolving gel (ml)
0.5 M Tris pH 6.8	1.26		
1.5 M Tris pH 8.8		2.5	2.5
30% Acrylamide/ methylene-bisacrylamide (37.5:1)	0.83	4.17	5
H_2O	2.75	3.12	2.29
10% SDS	0.05	0.1	0.1
10% APS	0.1	0.1	0.1
TEMED	0.01	0.01	0.01

The SDS-PAGE was run in 1 × SDS running buffer. Samples were mixed with loading buffer and loaded to gel wells. 100 V was applied until the running front reached resolving gel and 220 V was applied until the running front reached the bottom of gel. Gels were stained by Coomassie staining.

4.5 Measurement of protein concentration

Nanodrop spectrophotometer was used to measure the UV absorbance of protein at 280 nm. Extinction coefficient was calculated with ProtParam. Given the path length as 1 cm, the protein concentration was calculated as: protein concentration (mg/ml) = A280 (mg/ml)/extinction coefficient.

4.6 Limited proteolysis

Proteases Trypsin, Elastase, Chymotypsin, GluC and Subtilisin were prepared in 1: 10, 1: 100, 1: 1000 dilutions (stocks at 1 mg/ml) with buffer containing 20 mM Hepes pH 7.5, 50 mM NaCl, 10 mM $MgSO_4$. For each reaction, 10 µl of protein (0.6 mg/ml) was mixed with 3 µl of diluted protease and incubated on ice for 30 minutes. 1 µl AEBSF (100 mM) was used to stop the reaction. The sample was added with 5 µl 3 × SDS-loading buffer and boiled for 5 minutes at 95 °C and loaded on SDS-gel. For time course proteolysis, 10 µl of protein substrate (0.6 mg/ml) was incubated with 3 µl of a particular protease at a fixed concentration for different amount of time.

4.7 Mass spectrometry

The protein total mass measurement and peptide fingerprint analysis were done in the MPIB core facility.

4.8 Protein storage

Proteins for long-term storage were flash frozen in liquid nitrogen and stored at -80 °C in freezer.

4.9 Crystallization and structure determination

4.9.1 Crystallization

Optimized native crystals of CeSMG8 (1-423)/CeSMG9 (59-375) complex were obtained at 10 °C by hanging drop vapor diffusion within 1 day in drops formed by equal volumes (1.5 µl) of protein (6.8 mg/ml in gel filtration buffer 25 mM Tris, 300 mM NaCl, pH 8.0 mixed with 0.11 mM YCl_3) and crystallization buffer (10%

PEG3350, 0.1 M Tris pH 8.5). Crystals were cryoprotected with crystallization buffer supplemented with 25% ethylene glycol prior to cryo-cooling and data collection.
Selenium-methionine derivatized crystals of CeSMG8 (1-423)/CeSMG9 (59-375) were obtained at 10 °C by hanging drop vapor diffusion within 1 day in drops formed by equal volumes (1.5 µl) of protein (6.7 mg/ml in gel filtration buffer 25 mM Tris, 300 mM NaCl, pH 8.0, 2 mM DTT mixed with 0.11 mM YCl_3 and 2 mM TCEP) and crystallization buffer (11% PEG3350, 0.1 M Tris pH 8.5). In order to increase the size of selenium-methionine derivatized crystal and boost the anomalous signal, 1 mL paraffin oil was added on top of the reservoir buffer (0.7 mL). Crystals were cryoproteced with crystallization buffer supplemented with 25% ethylene glycol and 5 mM TCEP prior to cryo-cooling and data collection.

4.9.2 Crystal soaking with nucleotides

Native crystals were soaked for 10 minutes in cryoprotectant supplemented with 10 mM GDP and for 30 minutes in cryoprotectant supplemented with 10 mM ADP prior to cryo-cooling and data collection.

4.9.3 Data collection

All diffraction data were collected at 100K at the Swiss Light Source (SLS) beamline PXII, except diffraction data for ADP-soaked crystals were collected at the Deutsches Elektronen-Synchrotron (DESY) PETRA III beamline P11. The selenium-methionine derivatized crystals were measured at the peak wavelength of the selenium at 0.979 Å. The GDP and ADP-soaked crystals were measured at the wavelength of 0.9785 Å and 1.2547 Å, respectively.

4.9.4 Structure determination and refinement

The data were processed and scaled with XDS[82]. The crystals belong to trigonal space group P3221 with 3 copies of the complex in 3 fold NCS per asymmetric unit. SHELX[83] was used for phasing and phenix.autobuild[84] was used for initial automatic model building. The best built copies of SMG8 and SMG9 were chosen from initial selenium-methionine derivatized model and used as search models for molecular replacement with Phaser[85]. The model was completed with iterative rounds of manual building in Coot[86] and refinement with phenix.refine[87].

The GDP bound SMG8-9 structure was solved by molecular replacement with Phaser using selenium-methionine derivatized SMG8 and SMG9 structures as search models. The ADP bound SMG8-9 structure was solved by molecular replacement with Phaser using SMG8 and SMG9 structures from SMG8-9-GDP model as search models. The models were completed with Coot and refined with phenix.refine. The data processing and refinement statistics are summarized in Table 1 in supplementary material.

4.10 Nucleotides binding experiment

The affinities for nucleotides 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 nucleotides for 30 minutes at room temperature. Fluorescence of mant-nucleotides 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 fitting and dissociation constant (Kd) were done using Origin. Curves were done in triplicate.

4.11 Sequence alignments

The multiple sequence alignments of SMG8 and SMG9 were generated using program Clustal Omega[88] and edited in Jalview[89].

5 Results

5.1 Domain organization of SMG8 and SMG9

SMG8 and SMG9 from both humans and *C. elegans* have been investigated in this project. The domain architectures of SMG8 and SMG9 have not been well characterized since they were identified as two interacting partners of SMG1. Sequence alignments and disordered region prediction (supplementary materials) revealed that SMG8 in both human and *C. elegans* contain two conserved and structured regions, a N-terminal large region and a C-terminal small region connected by a large disordered region. SMG9 in both human and *C. elegans* were predicted to contain a N-terminal disordered region and a C-terminal putative NTPase domain. The accurate domain boundaries have not been defined in the absence of detailed structural information. In this thesis, *C. elegans* SMG8 and SMG9 were used as subjects for structural characterization with X-ray crystallography as they are smaller and contain less disordered regions compared to their human orthologues, which are advantages for crystallization. Various constructs of CeSMG9 were either expressed and purified singly or co-expressed and purified with N-terminal fragment of CeSMG8 for crystallization.

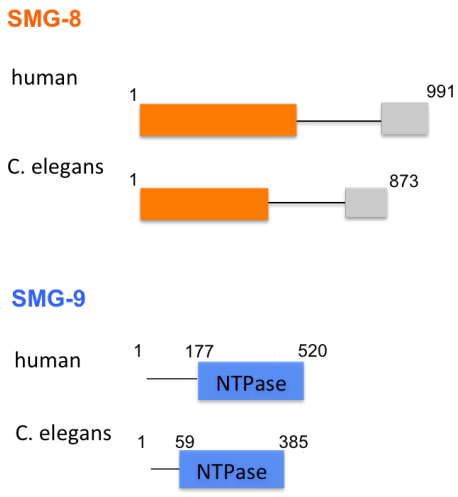


Figure 5.1: Domain organization of SMG8 and SMG9 in humans and *C. elegans*. The SMG8 N-terminal domain is shown in orange and C-terminal domain is shown in grey. The SMG9 C-terminal NTPase domain is shown in blue. The disordered regions are shown as lines.

5.2 Purification of full-length *C. elegans* SMG9

The full-length (FL) CeSMG9 (1-385) was the first construct purified. Expression and purification procedures were described in Methods 4.3.1. CeSMG9-FL was eluted as a single peak on a Superdex 200 gel filtration column. The elution volume of peak suggests that CeSMG9-FL exists as a monomer in the solution. Fractions of peak were collected and examined with SDS-PAGE. Although the protein was pure, apparent degradation was observed on the gel. The degradation probably occurred at the N-terminal disordered region of CeSMG9.

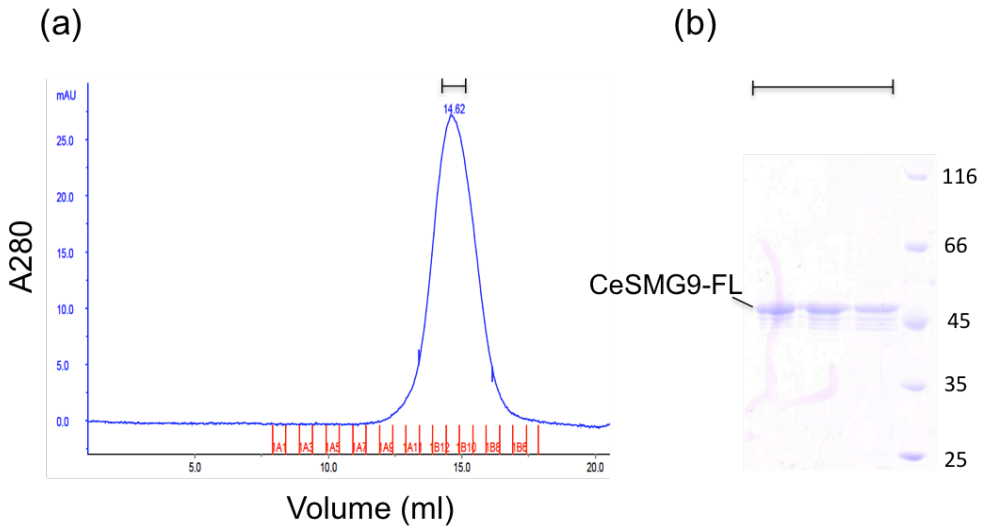


Figure 5.2: Gel filtration chromatography of CeSMG9-FL and SDS-PAGE. (a). *C. elegans* full-length SMG9 was eluted as a single peak from a superdex 200 column. The elution volume of peak was 14.62 ml as indicated on top of the peak. (b). 12.5% SDS-PAGE of selected fractions of gel filtration chromatography shown in (a).

5.3 Limited proteolysis of CeSMG9-FL

In order to obtain a structural core of CeSMG9 for crystallization, limited proteolysis of CeSMG9-FL was performed with Trypsin, Elastase, Chymotrypsin, GluC and Subtilisin with protocol described in Methods 4.6. With Trypsin digestion, significant bands were observed on the SDS-gel. A time course proteolysis with trypsin was subsequently performed to obtain stable fragments. The bands of interest as well as a negative control were extracted from the gel and analyzed by peptide mass

fingerprinting to identify the boundary of each fragment. The results are shown in the figure below. The band No.4 is the full-length CeSMG9 used as the negative control, the band No.3 is the fragment consisting of residues from 39 to 385, the band No.2 is the fragment consisting of residues from 171 to 385 and the band No.1 can't be identified due to weak intensity.

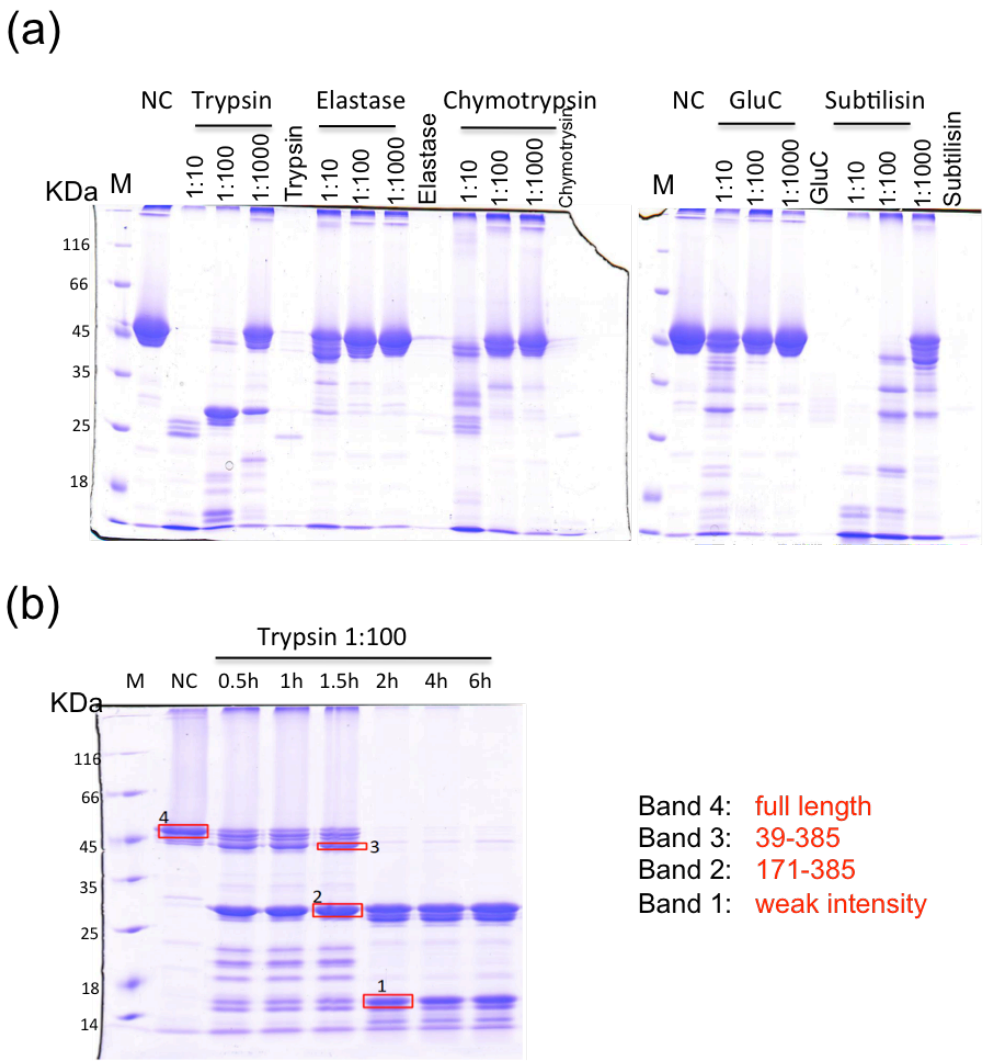


Figure 5.3: Limited proteolysis of CeSMG9-FL. (a). Trypsin, Elastase, Chymotrypsin, GluC and Subtilisin were used to digest protein CeSMG9-FL. Each enzyme was diluted to three different working concentrations and incubated with protein substrate for the same time (30 minutes). M: Marker; NC: negative control. (b). Protein CeSMG9-FL was incubated with Trypsin (0.01 mg/ml) for 0.5 hour, 1 hour, 1.5 hours, 2 hours, 4 hours and 6 hours. Samples were examined with 12.5% SDS-PAGE. Bands extracted for peptide mass spectrometry analysis were labeled with

numbered red rectangles. Results are shown at the right of the gel.

5.4 Purification of CeSMG9 C-terminal fragment (39-385)

Based on the result of limited proteolysis of CeSMG9-FL, a new construct CeSMG9 (39-385) was expressed and purified with the protocol described in Methods 4.3.2. CeSMG9 (39-385) was eluted in a homogenous peak in gel filtration chromatography. Fractions of peak were examined on SDS-PAGE gel and no degradation was observed on the gel. However, although the protein was pure and homogenous, no crystal hits were obtained from crystallization screening.

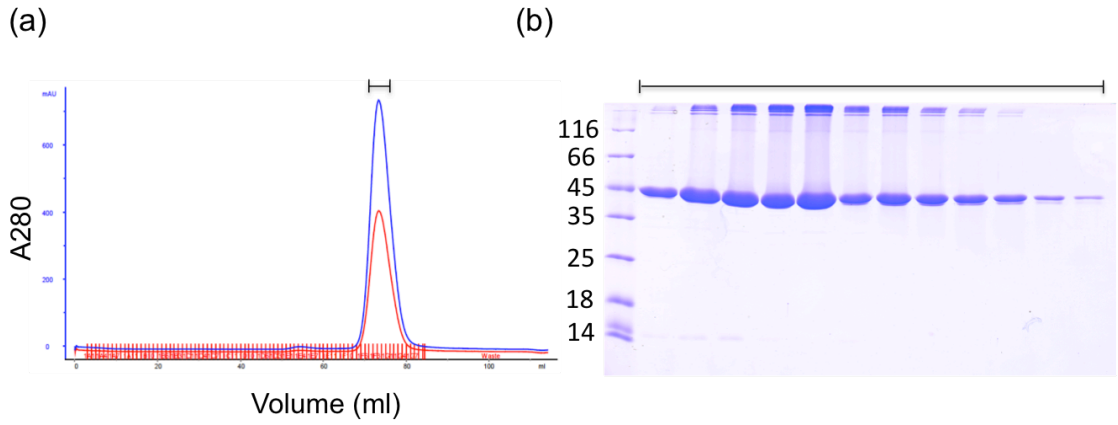


Figure 5.4: Purification of CeSMG9 (39-385) and SDS-PAGE. (a). Gel filtration chromatography on a Hiload 75 column. Blue line represents the absorption at 280 nm; red line represents the absorption at 260 nm. (b). 12.5% SDS-PAGE of selected fractions of gel filtration chromatography shown in (a).

5.5 Purification of CeSMG9 C-terminal fragment (59-385)

The CeSMG9 was further N-terminally truncated to a shorter construct containing residues from 59 to 385 based on secondary structure prediction and sequence conservation. CeSMG9 (59-385) was expressed and purified with the protocol as described in Methods 4.3.3. The protein was pure as shown on the SDS-gel, however, it's not homogenous as indicated by the profile of peak in gel filtration chromatography. No crystal hits were found by crystallization screening.

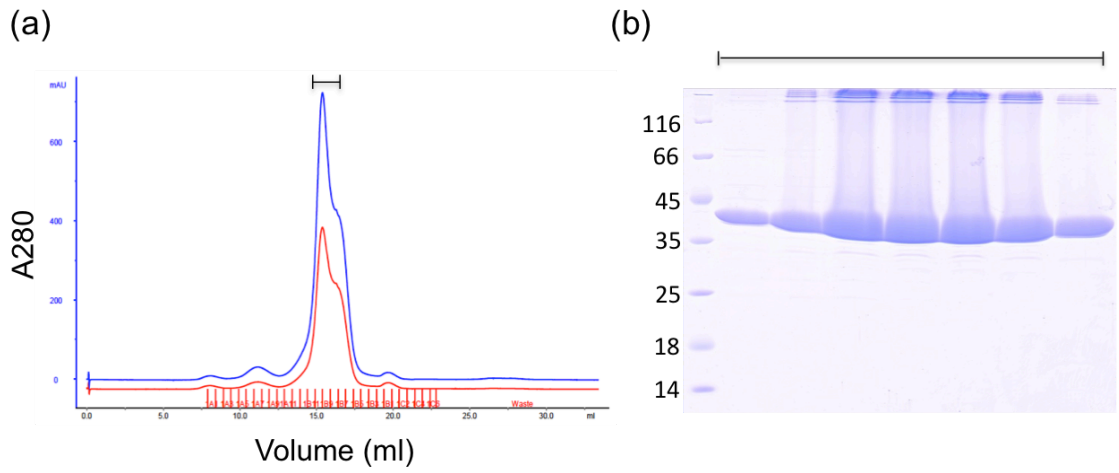


Figure 5.5: Purification of CeSMG9 (59-385). (a). Gel filtration chromatography on a superdex 200 column. Blue line represents the absorption at 280 nm; red line represents the absorption at 260 nm. (b). 12.5% SDS-PAGE of fractions of peak of gel filtration chromatography shown in (a).

5.6 Co-expression and purification of CeSMG8 N-terminal fragment (1-423) and CeSMG9-FL

Based on secondary structure prediction, the N-terminal 423 amino acids of CeSMG8 is likely the region that corresponds to a folded domain. Unlike CeSMG9, CeSMG8 (1-423) is not soluble when it is singly expressed in E. coli. Therefore, to test whether this N-terminal region of CeSMG8 is sufficient to bind CeSMG9, the plasmids of constructs of CeSMG8 (1-423) and CeSMG9-FL were co-transformed to the competent cell and co-expressed in E. coli. The expression and purification procedures were described previously in Methods 4.3.4. The results of gel filtration chromatography and SDS-PAGE gel shown below indicate that CeSMG8 (1-423) can be stabilized by CeSMG9 during co-expression and is sufficient to bind the full-length CeSMG9.

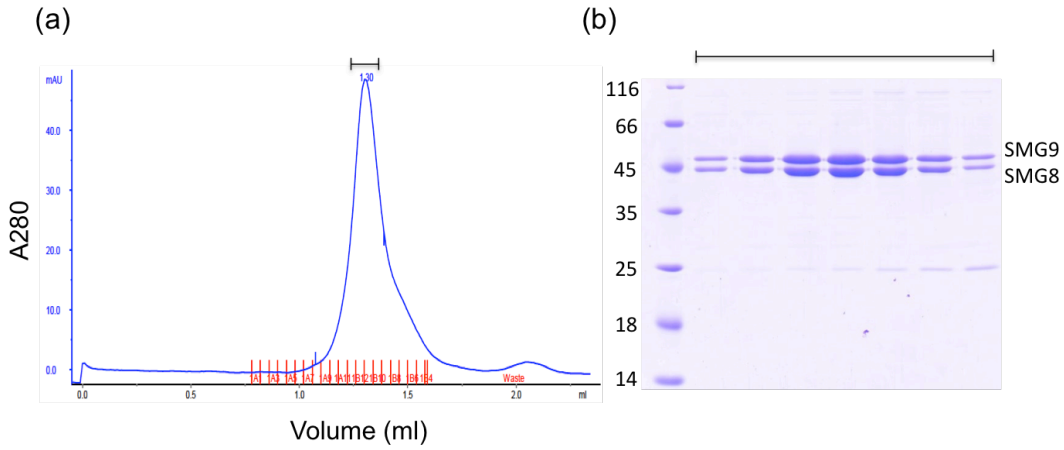


Figure 5.6: Purification of CeSMG8 (1-423) in complex with CeSMG9-FL. (a). Gel filtration chromatography on an analytical superdex 200 column. (b). 12.5% SDS-PAGE of fractions of peak from gel filtration chromatography shown in (a).

5.7 Co-expression and purification of CeSMG8 N-terminal fragment (1-423) and CeSMG9 C-terminal fragment (39-385)

Although the complex of CeSMG8 (1-423) and CeSMG9-FL can be expressed and purified from E. coli, the yield was very low (approximately 30 µg protein in total can be purified from 6 liters of E. coli culture). Therefore, expression of this complex in insect cells was necessary to obtain higher yield of protein for crystallization. The second problem was that CeSMG9-FL degraded also in complex with CeSMG8 (1-423) after staying in fridge at 4 °C for overnight. Based on previous results that the degradation of CeSMG9 occurs at its N-terminal unstructured region, a shorter construct CeSMG9 (39-385) was co-expressed with CeSMG8 (1-423) in insect cells. The expression and purification procedures were described in Methods 4.3.5. The protein yield from insect cells expression was much higher than expression in E. coli (shown in Figure 5.7 (a) and (b)) and the N-terminally truncated CeSMG9 is still sufficient to bind CeSMG8 (1-423) (shown in Figure 5.7 (c) and (d)).

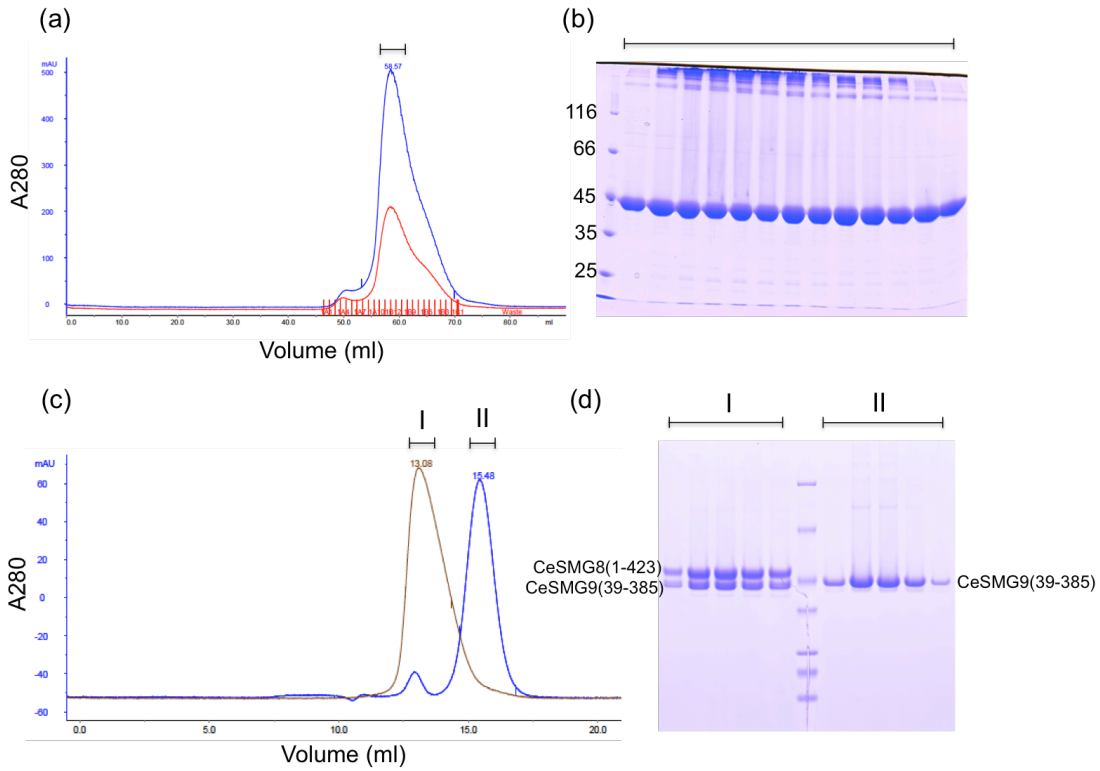


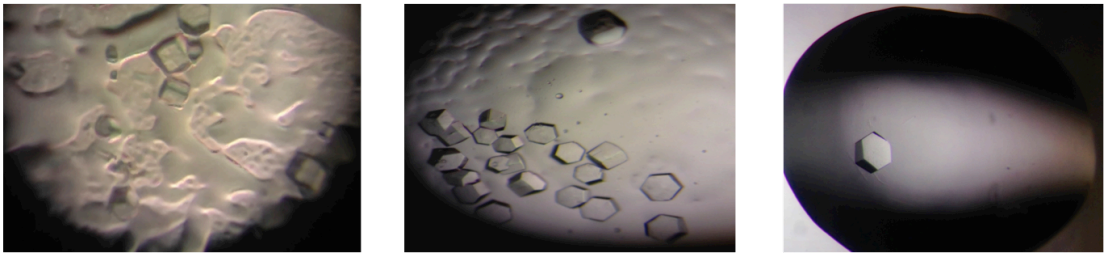
Figure 5.7: Purification of CeSMG8 (1-423) in complex with CeSMG9 (39-385). (a). Gel filtration chromatography on a superdex hiload-75 column. (b). 12.5% SDS-PAGE of fractions of peak from gel filtration chromatography shown in (a). CeSMG8 (1-423) and CeSMG9 (39-385) were not well separated on the gel. (c). Overlay of profiles of gel filtration chromatography of CeSMG9 (39-385) alone and in complex with CeSMG8 (1-423). Both gel filtration run were carried out on the same superdex 200 column. (d). Fractions of each peak were examined on a precast 4% -12% gradient SDS-PAGE gel. The molecular weight for each marker band from top to bottom is 116 kDa, 66 kDa, 45 kDa, 35 kDa, 25 kDa, 18 kDa and 14 kDa.

5.8 Crystallization of complex CeSMG8 (1-423)/CeSMG9 (39-385)

Purified CeSMG8 (1-423)/CeSMG9 (39-385) complex was used for crystallization screening. Small crystals shown in Figure 5.8 (a) were obtained within 1 day with sitting drop method in initial screening. The crystals could be reproduced shown in Figure 5.8 (b) with hanging drop method in a 24-well plate by changing the concentration of protein or precipitant. However, those crystals were very small, fragile and diffracted only to around 7 Å. In order to increase the crystal size and

improve the resolution, repeated seeding was carried out in sitting drops and the crystal after optimization shown in Figure 5.8 (c) can diffract to 3.1 Å. However, data analysis with phenix.xtriage[90] (Figure 5.8, d) indicated that the crystal had problem of perfect twinning, which became the main challenge for the structure determination.

(a) (b) (c)



(d)

Results of the twinning test with phenix.xtriage
For acentric data
$\langle I^2 \rangle / \langle I \rangle^2$: 1.470 (untwinned: 2.0, perfect twin: 1.5)
$\langle F \rangle^2 / \langle F^2 \rangle$: 0.902 (untwinned: 0.785, perfect twin: 0.885)
$\langle \lvert E^2-1 \rvert \rangle$: 0.514 (untwinned: 0.736, perfect twin: 0.541)
Multivariate Z score L-test: 34.175 (lower than 3.5 for good to reasonable data)

Figure 5.8: Crystals of complex CeSMG8 (1-423)/CeSMG9 (39-385). (a). Crystals from initial screening grew at 10 °C in the condition of 18% PEG 4000, 0.1M Tris-Cl, pH 8.0 in a sitting drop containing 150 nl protein (12 mg/ml) and 150 nl crystallization buffer. (b). Crystals were reproduced at 10°C in the condition of 10% PEG 4000, 0.1M Tris-Cl, pH 8.0 in a hanging drop containing 1 µl protein (12 mg/ml) and 1 µl crystallization buffer. (c). Crystal was optimized with repeated seeding and grew at 10°C in the condition of 18% PEG4000, 0.1M Tris-Cl, pH 8.0 in a sitting drop containing 200 nl protein (11 mg/ml), 200 nl crystallization buffer and 50 nl seeds. (d). Results of the twinning test with phenix.xtriage. Both Wilson ratios and moments are close to the values of perfect twin. The large Z score of L-test also indicates twinning. The methods of twinning test are well documented in the tutorial of phenix.xtriage.

5.9 Co-expression and purification of complex CeSMG8 (1-423)/CeSMG9 (59-385)

In order to improve the quality of crystals, the shorter construct CeSMG9 (59-385) was used to form the complex with CeSMG8 (1-423). Both proteins were co-expressed in insect cells and purified with procedures described previously in Methods 4.3.6. The results of gel filtration chromatography and SDS-PAGE shown in figure 5.9 indicate that this shorter construct CeSMG9 (59-385) is sufficient to interact with CeSMG8 (1-423). However, two bands were observed at the position of CeSMG9 on the gel due to possible degradation. The profile of peak of gel filtration chromatography also showed the inhomogeneity of the protein, which could be caused by the degradation.

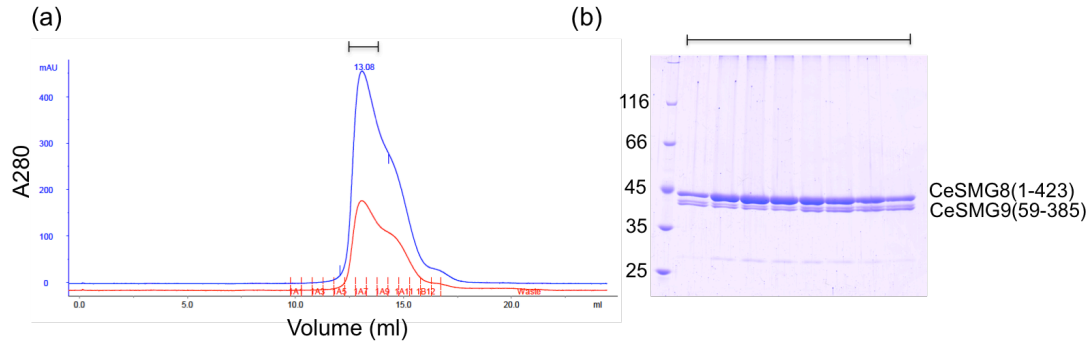


Figure 5.9: Purification of complex CeSMG8 (1-423/CeSMG9 (59-385). (a). Gel filtration chromatography on a superdex 200 column. (b). 12.5% SDS-PAGE of fractions of peak of gel filtration chromatography shown in (a).

5.9.1 Identification of degradation boundary by Mass Spectrometry

In order to know the degradation boundary of CeSMG9 (59-385), purified protein of complex CeSMG8 (1-423)/CeSMG9 (59-385) was sent to Mass Spectrometry analysis. The molecular weight of two CeSMG9 fragments shown in figure 5.10 indicates that the degradation probably occurred at the C-terminus of CeSMG9 and the shorter fragment is CeSMG9 (59-375) based on the calculation of its theoretical molecular weight.

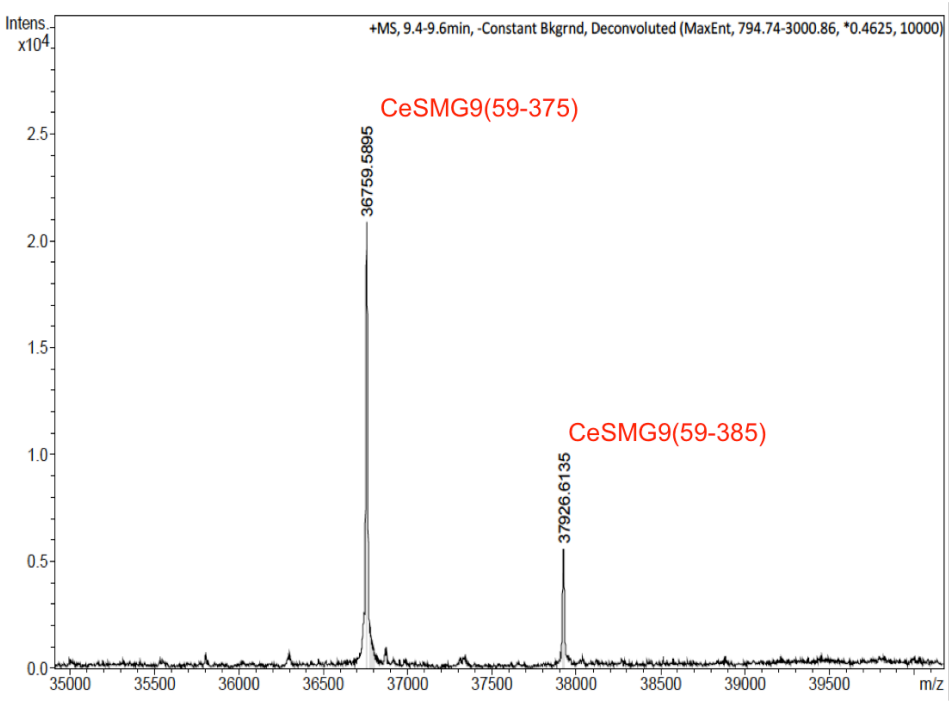


Figure 5.10: Molecular weight of CeSMG9 fragments measured by Mass Spectrometry. The boundaries of CeSMG9 fragments were marked on top of each peak.

5.10 Co-expression and purification of complex CeSMG8 (1-423)/CeSMG9 (59-375)

Based on the result from mass spectrometry, in order to obtain a homogeneous core complex of CeSMG8-9, CeSMG9 (59-375) was co-expressed with CeSMG8 (1-423) in insect cells and purified with procedures described in Methods 4.3.7. The results of gel-filtration chromatography and SDS-PAGE gel shown below indicate that the C-terminally truncated construct CeSMG9 (59-375) is sufficient to bind CeSMG8 (1-423) and no obvious protein degradation can be observed on the gel.

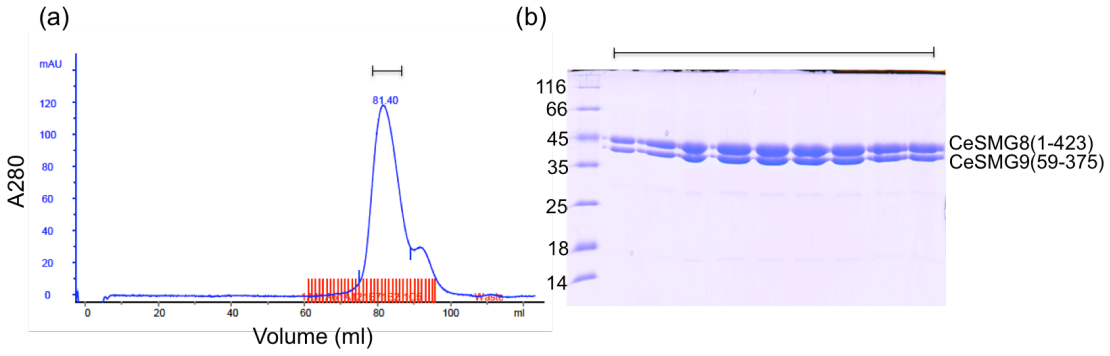


Figure 5.11: Purification of complex CeSMG8 (1-423)/CeSMG9 (59-375). (a). Gel filtration chromatography on a superdex hiload-200 column. (b). 12.5% SDS-PAGE of selected fractions of peak of gel filtration chromatography shown in (a).

5.11 Crystallization of complex CeSMG8 (1-423)/CeSMG9 (59-375)

Crystallization screening was set up with this core complex CeSMG8 (1-423)/CeSMG9 (59-375), crystals shown in figure 5.12 (a) were reproduced from initial screening with sitting drop method at room temperature but only diffracted to 7 Å and had problem of anisotropy. Unexpectedly this complex can also crystallize in the same condition in which the complex CeSMG8 (1-423)/CeSMG9 (39-385) has crystallized. Crystals of complex CeSMG8 (1-423)/CeSMG9 (59-375) shown in figure 5.12 (b) share the same morphology with crystals of complex CeSMG8 (1-423)/CeSMG9 (39-385) and diffracted to 3.5 Å but also had same twinning problem.

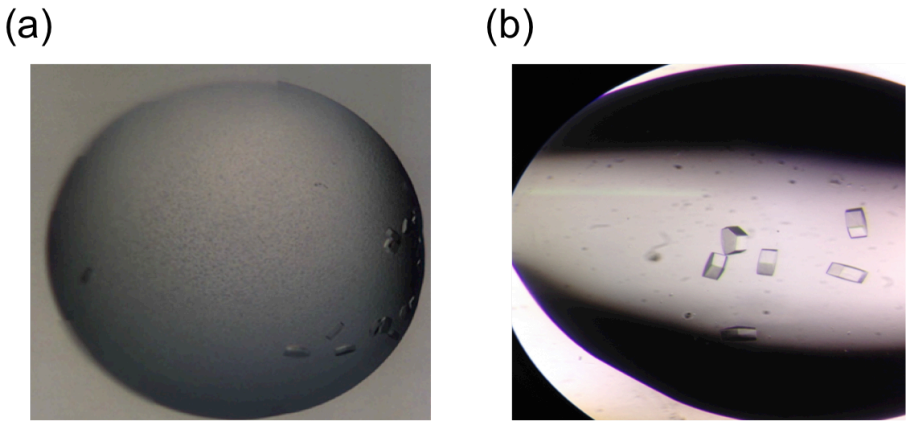


Figure 5.12: Crystals of complex CeSMG8 (1-423)/CeSMG9 (59-375). (a). Crystals grew at room temperature in the condition of 15% PEG 10000, 0.1 M Tris-Cl, PH 8.5 in a sitting drop containing 150 nl protein (8.7 mg/ml) and 150 nl crystallization buffer within 3 days. (b). Crystals grew at 10 °C in the condition of 18% PEG 4 000, 0.1 M Tris-Cl, PH 8.0 in a sitting drop containing 150 nl protein (8.7 mg/ml) and 150 nl crystallization buffer within 3 days.

5.12 Crystallization of complex CeSMG8 (1-423)/CeSMG9 (59-375) with yttrium chloride and structure determination

5.12.1 Crystallization screening and optimization

In order to solve the twinning problem of hexagonal crystals of complex CeSMG8 (1-423)/CeSMG9 (59-375), yttrium chloride was added to the protein as an additive for crystallization screening with the hope that protein can crystallize in a new crystal form without twinning problem, which is based on the previous report[91] that yttrium ions can bind specifically to surface-exposed glutamate and aspartate side chains to create contact of molecules to yield high quality crystals that belong to a new space group. The crystals grew in the initial screening condition shown in figure 5.13 (a) were very small but were not hexagonal, which was a good starting point for optimization. The crystals can be reproduced shown in figure 5.13 (b) by changing the concentration of precipitant and removing ammonium sulfate in the crystallization condition. The crystals were further optimized with hanging drop method and grown at 10 °C. The optimized crystals shown in figure 5.13 (c) have very different morphology as compared to the twinning hexagonal crystals. The final concentration of YCl_3 in protein sample has been constantly 0.11 mM (stock concentration is 1 mM).

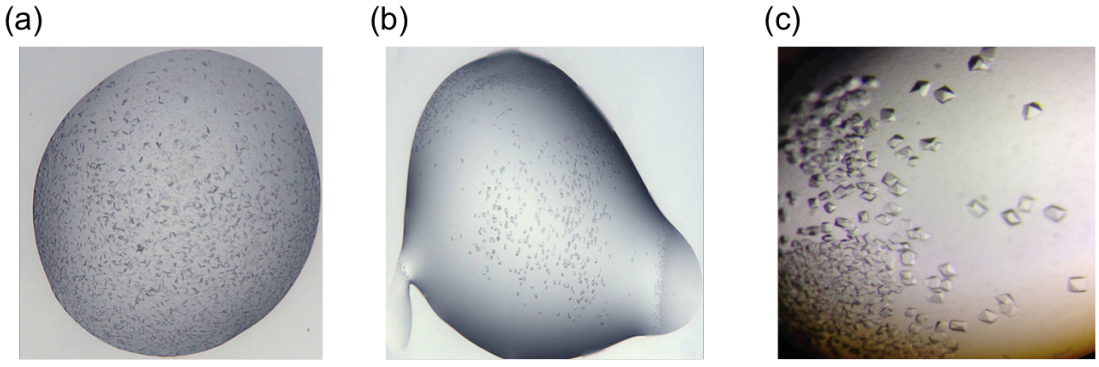


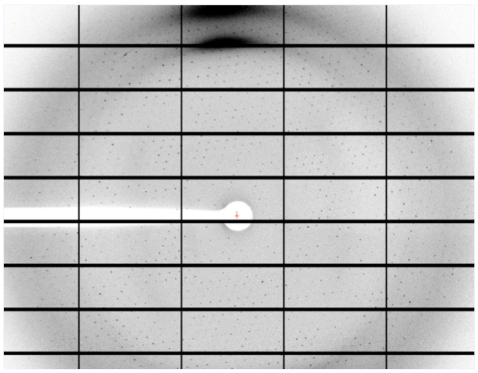
Figure 5.13: Crystals of complex CeSMG8 (1-423)/CeSMG9 (59-375) crystallized in presence of YCl_3. (a). Crystals from initial screening grew at 4 °C in the condition of 25% PEG 3350, 0.1 M Tris-Cl, pH 8.5, 0.2 M ammonium sulfate in a sitting drop containing 150 nl protein (5.9 mg/ml) and 150 nl crystallization buffer. (b). Crystals were reproduced at 4 °C in the condition of 14% PEG 3350, 0.1 M Tris-Cl, pH 8.5 in a sitting drop containing 150 nl protein (5.9 mg/ml) and 150 nl crystallization buffer. (c).

Crystals were optimized at 10 °C and grew in the condition of 12% PEG 3350, 0.1 M Tris-Cl, pH 8.5 in a hanging drop containing equal volume (1 µl) protein (5.1 mg/ml) and crystallization buffer.

5.12.2 Diffraction data of native crystals of complex CeSMG8 (1-423)/CeSMG9 (59-375)

The crystals were further optimized to increase the size by being grown in bigger hanging drop (1.5 µl protein) and changing concentration of protein and precipitant. The diffraction pattern and data processing result are shown below in figure 5.14. The crystal can diffract to around 3.6 Å and no significant twinning can be detected by phenix. Xtriage. The crystal belongs to space group P3221.

(a)



(b)

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR expected	COMPARED	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano
10.82	9098	2215	2254	98.3%	4.6%	4.4%	9088	26.51	5.3%	99.7*	20*	0.872	871
7.75	16111	3826	3836	99.7%	6.6%	5.9%	16094	20.91	7.5%	99.7*	4	0.829	1660
6.35	21707	4936	4939	99.9%	19.8%	18.7%	21695	10.36	22.5%	98.2*	-8	0.759	2229
5.52	25375	5792	5795	99.9%	35.0%	35.2%	25345	6.65	39.8%	95.0*	-5	0.736	2634
4.94	28145	6599	6604	99.9%	37.3%	37.0%	28097	6.58	42.6%	93.4*	-1	0.768	3020
4.51	32217	7258	7261	100.0%	40.0%	40.8%	32176	6.21	45.4%	94.1*	-1	0.760	3360
4.18	32644	7902	7913	99.9%	57.5%	59.7%	32581	4.18	66.0%	88.3*	2	0.746	3658
3.91	35644	8507	8515	99.9%	93.6%	100.5%	35560	2.72	107.2%	78.5*	7	0.731	3947
3.69	27735	7830	8942	87.6%	135.7%	149.4%	26959	1.51	157.9%	60.3*	4	0.674	3148
total	228676	54865	56059	97.9%	27.1%	27.7%	227595	7.01	31.0%	98.6*	2	0.749	24527

Figure 5.14: Diffraction data of native crystals of complex CeSMG8 (1-423)/CeSMG9 (59-375). (a). X-ray diffraction pattern. (b). Data processed by XDS.

5.12.3 Preparation of selenium-methionine derivatized complex CeSMG8 (1-423)/CeSMG9 (59-375)

Selenium-methionine derivatized complex CeSMG8 (1-423)/CeSMG9 (59-375) was expressed and purified for SAD experiment for phasing. Theoretically, 1 selenium-

methionine residue can provide enough phasing power for 100 amino acids. 3.1% residues in CeSMG8 (1-423) and 2.8% residues in CeSMG9 (59-375) are methionine. The procedures for expression and purification of selenium-derivatized protein were described in Methods 4.3.8. The proteins were eluted in a homogeneous peak in gel filtration chromatography shown in figure 5.15 (a), however two additional bands were observed on the SDS-gel shown in figure 5.15 (b). These two additional bands could be either impurities or the degraded fragments of CeSMG8. The incorporation rate of selenium in CeSMG8 and CeSMG9 were analyzed by Mass Spectrometry. The results of molecular weight measurement shown in figure 5.15 (c) and (d) indicate the inhomogeneous incorporation of selenium-methionine in CeSMG8 (1-423) and CeSMG9 (59-375).

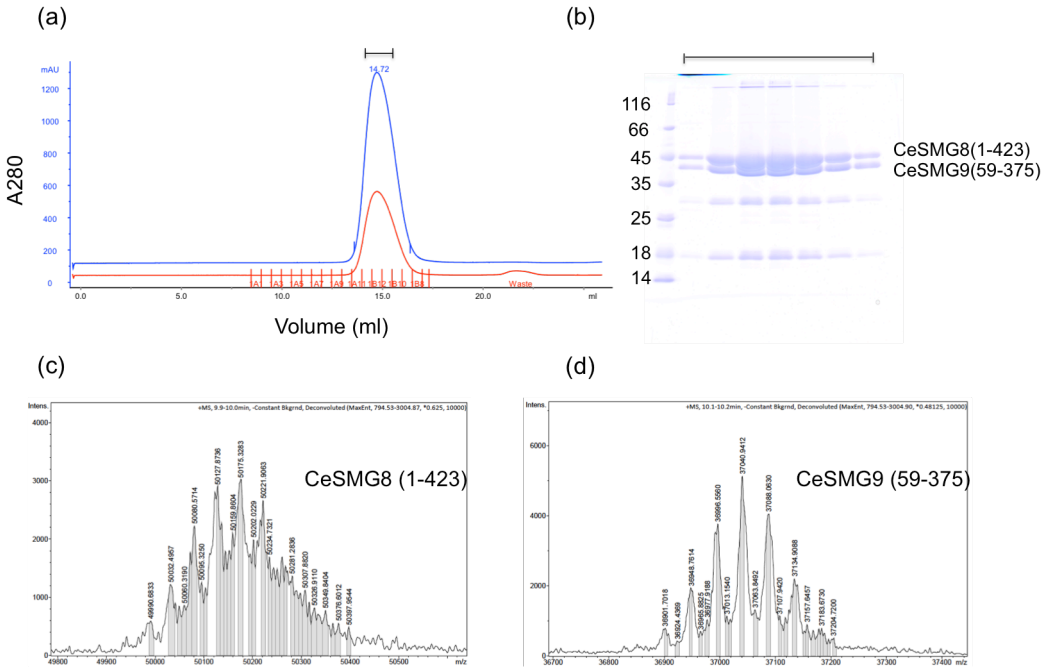
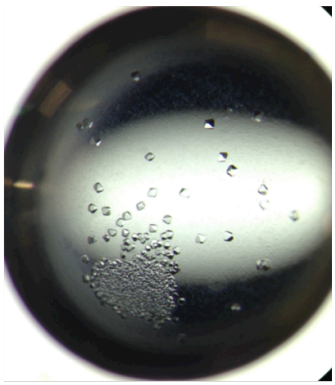


Figure 5.15: Preparation of selenium-methionine derivatized complex CeSMG8 (1-423)/CeSMG9 (59-375). (a). Gel filtration chromatography on a superdex 200 column. (b). 12.5% SDS-PAGE of fractions of peak of gel filtration chromatography shown in (a). (c). Molecular weight of selenium-methionine derivatized CeSMG8 (1-423) measured by Mass Spectrometry. (d). Molecular weight of selenium-methionine derivatized CeSMG9 (59-375) measured by Mass Spectrometry.

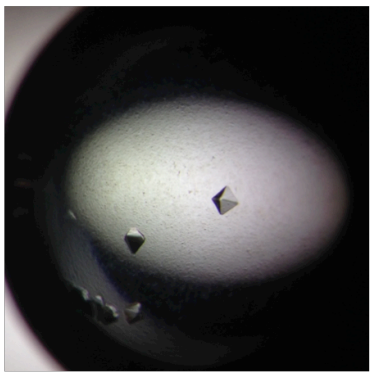
5.12.4 Crystallization of selenium-methionine derivatized complex CeSMG8 (1-423)/CeSMG9 (59-375) and data collection

The selenium-methionine derivatized complex CeSMG8 (1-423)/CeSMG9 (59-375) was crystallized in the same crystallization condition as native protein. However, like the native protein, the crystallization was too fast and produced over 100 very small crystals in one drop shown in figure 5.16 (a), which diffracted weakly and can not produce strong enough anomalous signals for phasing. During data collection, it was found that bigger crystals diffracted better than smaller ones. Therefore, efforts were made to increase the crystal size to get better diffracting crystals. Various methods were tried but only one method worked, which was adding 1 ml paraffin oil on top of reservoir buffer (0.7 ml) in the well based on the knowledge that paraffin oil is very impermeable that can slow down the vapor diffusion process and produce less and bigger crystals in the drop shown in figure 5.16 (b). The dataset collected from optimized crystal was processed with XDS shown in figure 5.16 (c) and the anomalous signal was strong enough for phasing.

(a)



(b)



(c)

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR expected	COMPARED	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano
8.30	60484	4886	4916	99.4%	2.6%	2.7%	60484	85.02	2.8%	100.0*	97*	6.988	2043
5.92	108443	8621	8621	100.0%	4.3%	4.0%	108443	53.14	4.5%	99.9*	92*	4.829	3910
4.84	147493	11061	11061	100.0%	5.2%	4.9%	147493	45.85	5.4%	99.9*	86*	3.384	5141
4.20	167099	13141	13141	100.0%	5.4%	5.1%	167099	41.87	5.6%	99.9*	73*	2.303	6172
3.76	196443	14878	14879	100.0%	7.7%	7.6%	196443	29.77	8.0%	99.9*	56*	1.655	7038
3.43	215689	16413	16413	100.0%	11.7%	11.9%	215689	19.67	12.2%	99.8*	40*	1.274	7810
3.18	224628	17810	17810	100.0%	21.7%	22.7%	224627	10.57	22.6%	99.4*	23*	0.987	8501
2.98	253109	19220	19222	100.0%	39.3%	41.7%	253109	6.10	40.9%	98.3*	15*	0.859	9217
2.81	260986	20163	20390	98.9%	62.8%	66.3%	260888	3.80	65.4%	95.3*	7	0.777	9620
total	1634374	126193	126453	99.8%	8.0%	8.1%	1634275	24.40	8.4%	100.0*	62*	1.853	59452

Figure 5.16: Crystals of selenium-methionine derivatized complex CeSMG8 (1-423)/CeSMG9 (59-375) and data processing. (a) Crystals grew at 10 °C with hanging drop method in the condition of 11% PEG 3350, 0.1 M Tris-Cl, pH 8.5 in a drop containing equal volume (1.5 µl) of protein (6.7 mg/ml) and crystallization buffer. (b). Optimized crystals grew at 10 °C with hanging drop method in the condition of 11% PEG 3350, 0.1 M Tris-Cl, pH 8.5 in a drop containing equal volume (1.5 µl) of protein (6.7 mg/ml) and crystallization buffer with 1 ml paraffin oil on top of the reservoir buffer in the well. (c). Processed results of dataset collected from an optimized selenium-methionine derivatized crystal.

5.13 Crystal structure of *C. elegans* SMG8-SMG9 core complex

The crystal structure of *C. elegans* SMG8-SMG9 core complex was solved at the resolution of 2.5 Å with procedures described in the Methods 4.9.4. Three copies of CeSMG8-CeSMG9 complex are present in the asymmetric unit forming a ring-like structure in 3 fold non-crystallographic symmetry (Figure 5.17, b). The structure model was refined to the final R-free of 26.3% with good stereochemistry and mostly built except some disordered regions. The structure shows that the previously predicted N-terminal region of CeSMG8 actually consists of two domains, a N-terminal G-like domain (1-323) and a middle helical bundle domain (334-416), connected by a short linker of 10 residues. The C-terminal domain (59-363) of CeSMG9 is also a G-like domain (Figure 5.17, c). Both G-like domains are centered at a mixed β-sheet (of five parallel and one antiparallel β-strands) surrounded by α-helices both on the concave surface (α1, α5) and on the convex surface (α2, α3, α4) (Figure 5.17, d). Both CeSMG8 and CeSMG9 incorporate additional secondary structure elements. Two more helices line the convex surface of each domain (α2A, α7 in CeSMG8 and α4, α7 in CeSMG9). The major structural difference between CeSMG8 and CeSMG9 is the presence in the form of a helical bundle of three C-terminal helices (α7-α9) forming a stalk-like protrusion, reminiscent of the stalk domain found in GTPases of the dynamin-like family such as Atlastin[92] or GBP1[93].

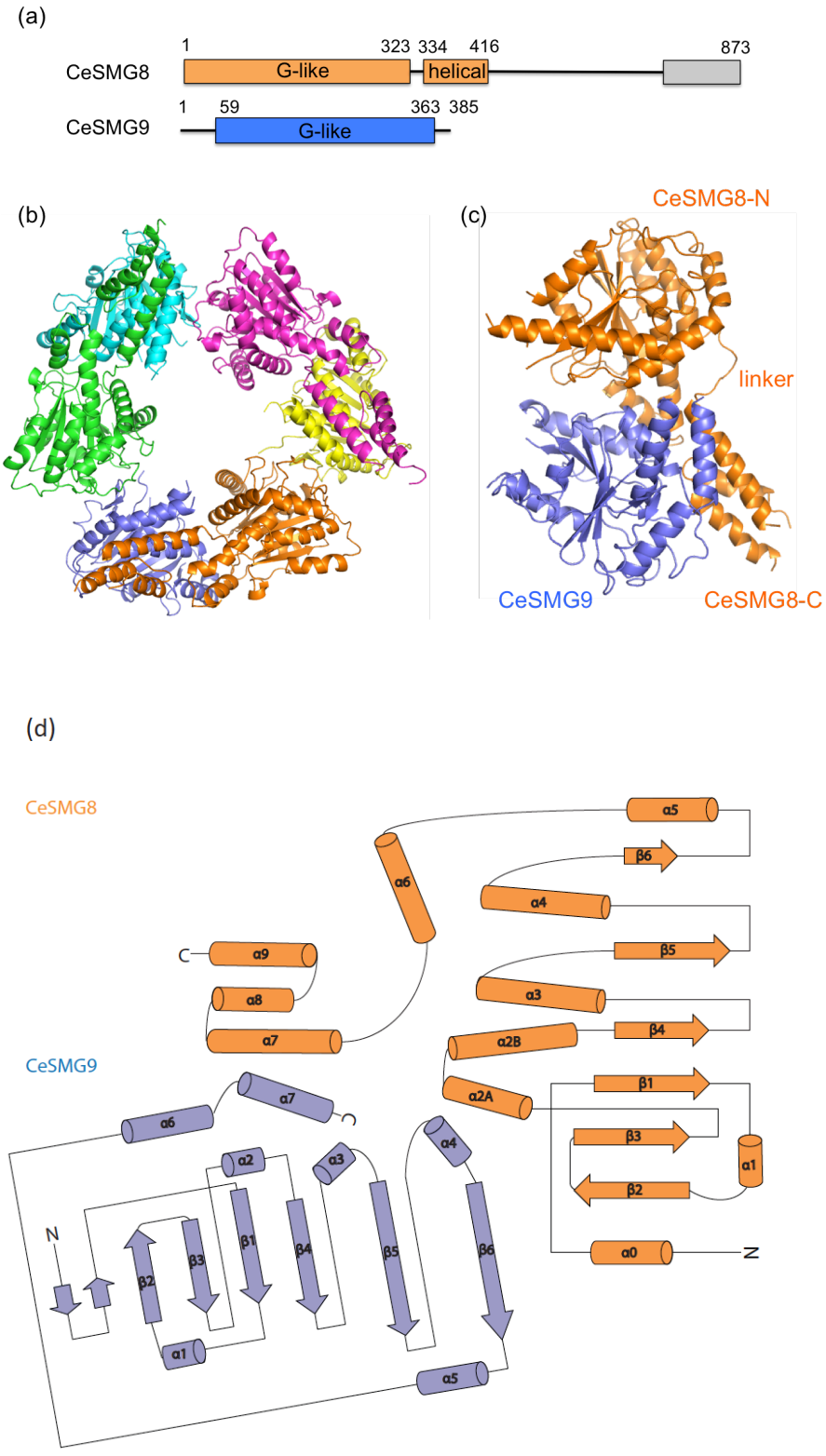


Figure 5.17: The crystal structure of *C. elegans* SMG8-SMG9 core complex. (a). The domain architecture of CeSMG8 and CeSMG9 based on structural information. The CeSMG8 contains a N-terminal G-like domain (in orange), a middle helical bundle domain (in orange) and a distant domain of unknown structure and function (in

grey). The CeSMG9 contains a N-terminal unstructured region and a C-terminal G-like domain (in blue). The domain boundaries are marked with numbers. (b). Three copies of CeSMG8-CeSMG9 complex are in the asymmetric unit forming a ring-like structure. (c). Overall view of the structure model of CeSMG8-CeSMG9 core complex. (d). Topology diagram of CeSMG8 (1-416) and CeSMG9 (59-363). Figure was generated by program Pro-origami[94] with structures of CeSMG8 and CeSMG9 and manually edited.

5.14 CeSMG9 interacts with both domains of CeSMG8

The interaction between CeSMG8 and CeSMG9 is extensive and contributed mainly by helices at the binding interface. The G-like domains face each other and interact with part of their convex surfaces. The buried surface area of CeSMG8 and CeSMG9 calculated with program PISA[95] are 1724.5 Å² (9.5 % of total solvent-accessible area) and 1704.5 Å² (13.7% of total solvent-accessible area), respectively. CeSMG9 G-like domain interacts with CeSMG8 N-terminal G-like domain as well as middle helical bundle domain at two major binding sites. The first binding site is on the G-like domain of CeSMG8 (Arg87, Lys91, Val83, Ile85), which interacts with CeSMG9 (Glu254, Glu257, Tyr202, Leu258) with both charged interaction and hydrophobic interaction (shown below, left). The second binding site is composed by both CeSMG8 G-like domain (Phe99) and helical bundle domain (Ile335, Phe338), which contacts CeSMG9 (Val212, Ile216, Tyr358) with hydrophobic interaction (shown below, right) and anchors the CeSMG8 helical bundle domain on CeSMG9.

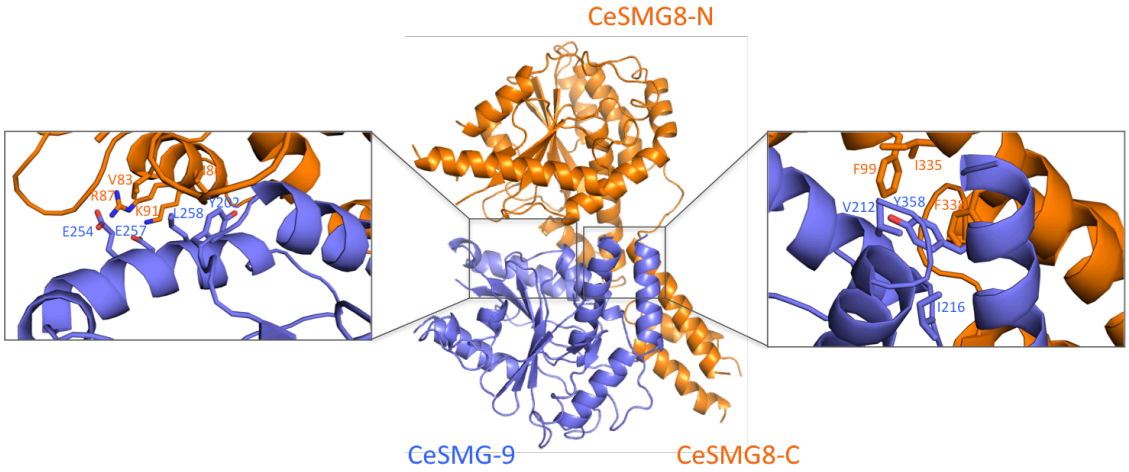


Figure 5.18: Zoom-in views of major interactions between CeSMG8 and CeSMG9. Left: CeSMG9 interacts with the binding site on N-terminal CeSMG8;

middle: overall view of the structure; right: CeSMG9 interacts with a composite binding site on CeSMG8 formed by the N-terminal G-like domain and the middle helical bundle domain.

5.15 SMG8-SMG9 binding interface is conserved in humans

In order to know whether the binding interface of CeSMG8 and CeSMG9 is conserved in human SMG8 and SMG9, Co-immunoprecipitation (IP) assays using mammalian cells were performed by Mahesh Lingaraju in the lab with human SMG8 and SMG9 mutants designed with structure information and sequence conservation. The Met390 and Tyr515 in hSMG9 are conserved with Leu258 and Tyr358 in CeSMG9 respectively based on sequence alignment (supplementary materials, S4). The results of Co-IP show that the single mutation M390R in hSMG9 abolished the hSMG8 binding significantly while the single mutation Y515R did not weaken the binding. The double mutation M390R and Y515R abolished the hSMG8 binding almost completely. The results indicate that the interaction mode of SMG8-SMG9 is conserved in humans and the binding is stronger at the N-terminal binding site than at the C-terminal binding site of SMG8.

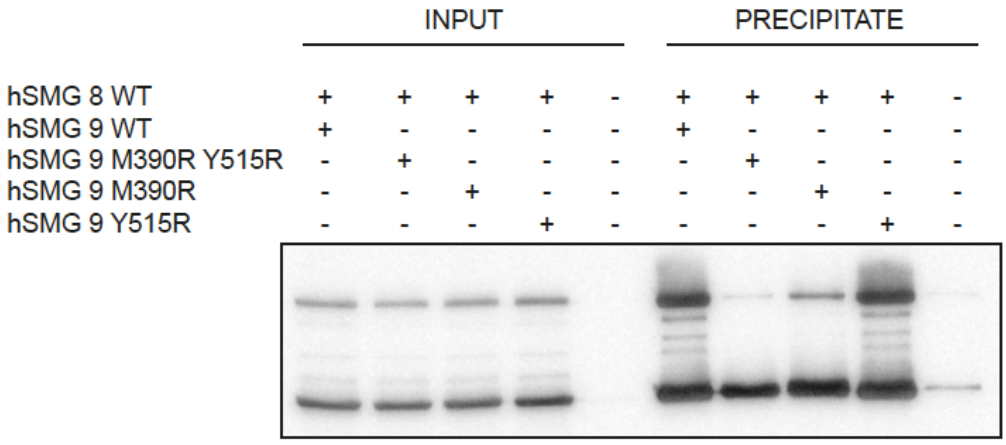


Figure 5.19: Co-IP of human SMG8 and SMG9. Figure is contributed by Mahesh Lingaraju.

5.16 Crystal structure of CeSMG8-CeSMG9-GDP complex

The crystal structure of CeSMG8-CeSMG9 core complex shows that both the N-terminal domain of CeSMG8 and C-terminal domain of CeSMG9 are G-like domains and SMG9 was predicted to be a NTPase based on sequence analysis. To know whether CeSMG8 and CeSMG9 bind GTP or GDP and how they bind, native crystals of complex CeSMG8 (1-423)/CeSMG9 (59-375) shown in figure 5.20 (a) were soaked with either GTP and GTP analogues or GDP. However, datasets can only be collected from GDP-soaked crystals because crystals had severe cracks when soaked with GTP or GTP analogues even for very short time. The GDP-soaked crystals diffracted to 2.64 Å and the structure can be solved by molecular replacement with the Apo structure of complex CeSMG8-CeSMG9 as search model. The GDP-soaked crystal belongs to same space group P3221 and contains also three copies of CeSMG8-CeSMG9 complex in the asymmetric unit. However, the density for a complete GDP molecule presents in only one copy of CeSMG9 shown in figure 5.20 (b) probably due to the crystal packing that blocked the access of GDP to the binding sites on other two copies of CeSMG9. The GDP-bound structure could be refined with final R-free of 28.27% and good stereochemistry. Most of the model could be built except some disordered regions including CeSMG8 residues 193-207, 255-288, 357-381, 422-423 and CeSMG9 residues 125-134, 152-172, 285-307 and 365-375. The GDP binds to the P-loop of CeSMG9. The switch regions in CeSMG9 are not visible in the structure due to their conformational flexibility. No conformational change is observed in GDP-bound CeSMG9 structure as compared to the Apo structure.

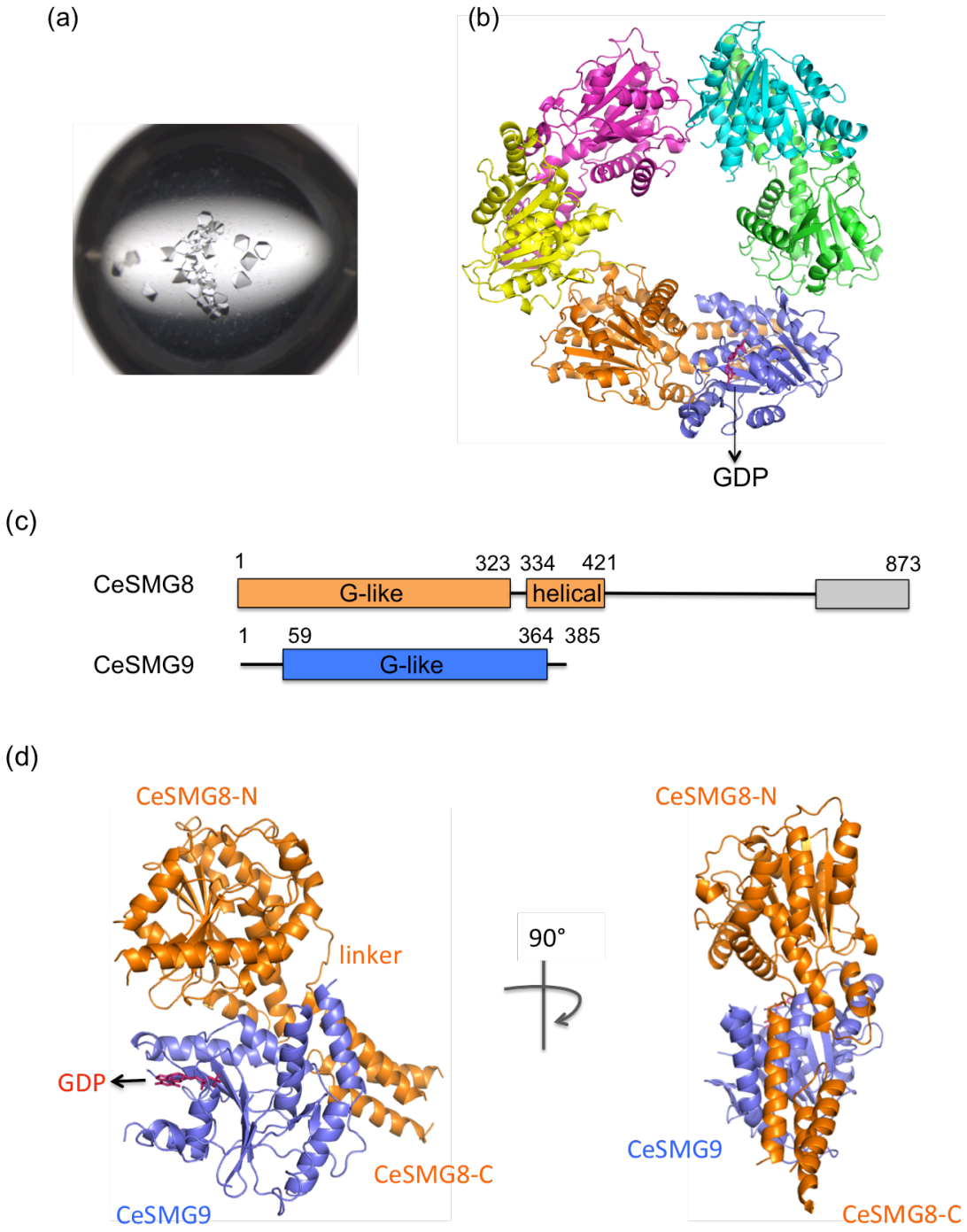


Figure 5.20: The crystal structure of CeSMG8-CeSMG9-GDP complex. (a). Native crystals of complex CeSMG8 (1-423)/CeSMG9 (59-375) used for GDP soaking. The crystals were grown at 10 °C with hanging drop method in the condition of 11% PEG 3350, 0.1 M Tris-Cl, pH 8.5 in a drop containing equal volume (1.5 µl) of protein (6.7 mg/ml) and crystallization buffer with 1 ml paraffin oil on top of the reservoir buffer in the well. (b). Three copies of CeSMG8-CeSMG9 complex are in the asymmetric unit and GDP molecule (shown in red) is bound to only one copy of CeSMG9. (c). Domain

architecture of CeSMG8 and CeSMG9. (d). Overall view of the crystal structure of CeSMG8-CeSMG9-GDP complex shown in two orientations related by the rotation of 90° around the vertical axis. CeSMG8 is shown in orange, CeSMG9 is shown in blue, GDP is shown in red.

5.17 GDP binds to the conserved P-loop in CeSMG9

Comparison with previously determined structures in protein data bank (PDB) using program Dali[96] shows that both G-like domains of CeSMG8 and CeSMG9 resemble the GTPase domain of the dynamin-like family proteins hGBP1 as well as Atlastin. Although both CeSMG8 and CeSMG9 G-like domains share a similar overall fold with G-domain of hGBP1, GDP only binds to the P-loop (G1 motif) of CeSMG9 but not CeSMG8. Comparing the P-loops of CeSMG9 and hGBP1 G-domain, ^{CeSMG9}Lys99 and ^{CeSMG9}Ser100 are structurally conserved with ^{hGBP1}Lys51 and ^{hGBP1}Ser52, which interact with phosphate groups of GMPALF4 via hydrogen bonding. Although parts of switch 1 and 2 are disordered in our GDP-bound structure (residues 124-134 and 152-172, respectively), Asp150^{CeSMG9} in switch 2 is at the position expected for coordinating the magnesium ion while Thr135^{CeSMG9} in switch 1 is well ordered but is 10 Å away from the position expected for coordinating the γ-phosphate. There are two major differences in the phosphate-binding G1-G3 motifs of CeSMG9 as compared to the dynamin-like family proteins (Figure 5.21, d). First, there is a conserved proline residue (Pro153^{CeSMG9}) at the position of switch 2 typically occupied by a glycine. Second, there is a conserved glycine residue (Gly96^{CeSMG9}) in the P loop at the equivalent position of the so–called arginine-finger, an important element for the in-cis stimulation of GTPase activity in dynamin-like proteins[97,98]. Another major difference between CeSMG9 and dynamin-like proteins is at the motifs that bind the base of the nucleotide. The G4 motif of CeSMG9 contains a conserved lysine residue (Lys241^{CeSMG9}) instead of the aspartic acid that in hGBP1 and Atlastin mediates specific interactions with the guanine base. The side chain of Lys241^{CeSMG9} stacks on top of the guanine base with the aliphatic portion and points towards the phosphates of the nucleotide with the positively charged tip.

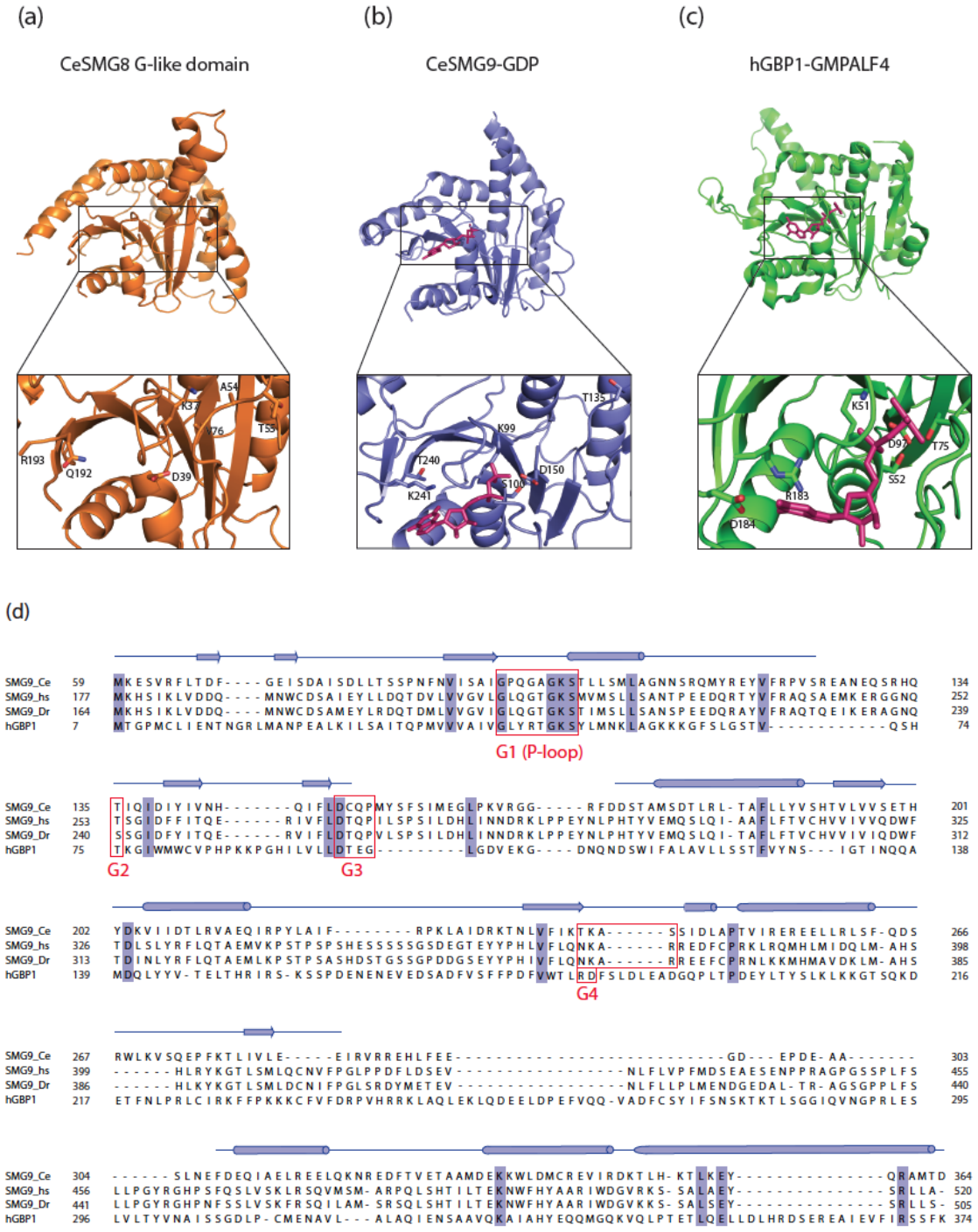


Figure 5.21: Structural comparison of CeSMG8, CeSMG9 and hGBP1 and sequence alignment of SMG9 and hGBP1. (a): Zoom-in view of the region in CeSMG8 that corresponds to the GDP-binding site in CeSMG9 based on structural alignment; (b): Zoom-in view of the GDP-binding site in CeSMG9. Highlighted residues K99 and S100 in P-loop (G1), T135 in G2, D150 in G3, T240 and K241 in G4 are conserved as in classical G motifs in GTPases; (c): Zoom-in view of GMPALF4-binding site in G-domain of hGBP1. (d): Sequence alignment of SMG9 from

Caenorhabditis elegans, *Homo sapiens*, *Danio rerio* and human GBP1. G-motifs are marked with red rectangles. The secondary-structure elements of CeSMG9 are indicated above the sequences.

5.18 The non-canonical G4 motif of CeSMG9 allows ATP binding

It is likely that CeSMG9 lost its GTP binding specificity due to the non-canonical G4 motif, therefore it is interesting to know whether CeSMG9 is able to bind ATP. I soaked the native crystal of complex CeSMG8 (1-423)/CeSMG9 (59-375) with ADP in the same way as soaking with GDP. The crystal structure of CeSMG8-CeSMG9-ADP was solved by molecular replacement using the structure of CeSMG8/CeSMG9 complex as searching model. The structure was solved at 2.65 Å and contains three copies of CeSMG8 and CeSMG9 in the asymmetric unit. The model was refined to a final R-work of 24.4% and R-free of 28.3% and good stereochemistry. The density of ADP molecule can be observed at the same position in CeSMG9 where GDP binds. The phosphate groups of ADP interact with the P-loop of CeSMG9 (residues Lys99, Ser100) by forming a number of hydrogen bonds. Like GDP, the adenine base of ADP is not coordinated with specific residues. CeSMG9 lost specificity for nucleotide recognition due to the non-canonical G4 motif and binds both GDP and ADP with its conserved P-loop.

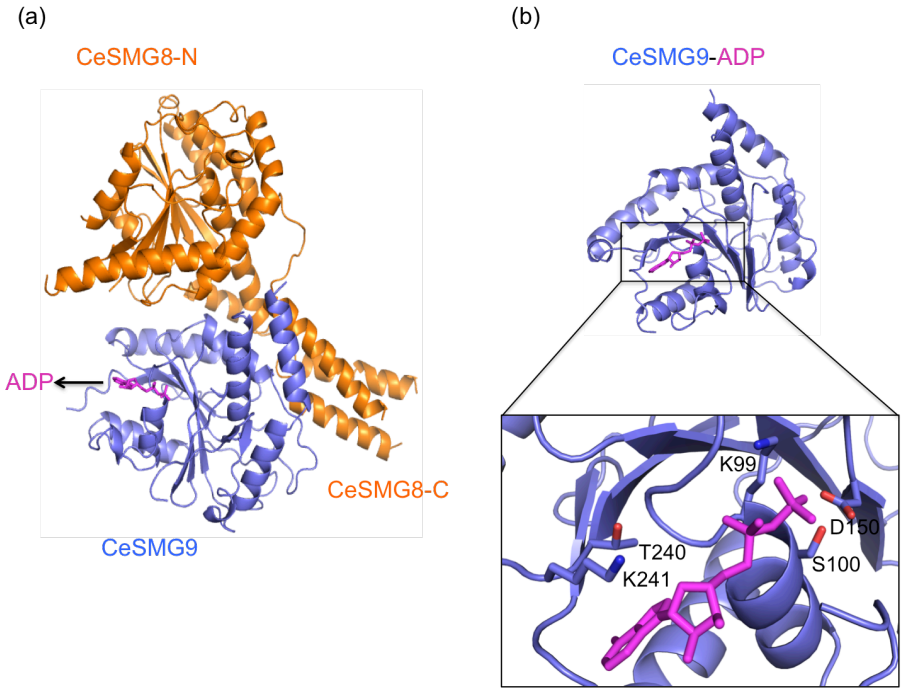


Figure 5.22: Crystal structure of the CeSMG8-CeSMG9-ADP complex. (a): Overall view of structure model. CeSMG8 is shown in orange, CeSMG9 is shown in blue, ADP is shown in magenta. (b): zoom-in view of ADP binding site. Highlighted residues are conserved in G motifs.

5.19 CeSMG9 binds GTP and ATP with micro molar affinity

The binding affinity of CeSMG9 with different nucleotides GTPγS, GDP and ATP were measured with Mant-labeled Fluorescence Anisotropy. The experiment procedure was described in Methods 4.10. CeSMG9 alone binds GDP with affinity of around 10 μM (figure 5.23, a) and the complex CeSMG8/CeSMG9 binds GDP with affinity of around 15 μM (figure 5.23, b). The complex CeSMG8/CeSMG9 binds GTPγS and ATP with slightly higher affinity (6.5 μM and 1.5 μM, respectively) than binds GDP probably due to the interaction between the additional phosphate group and CeSMG9.

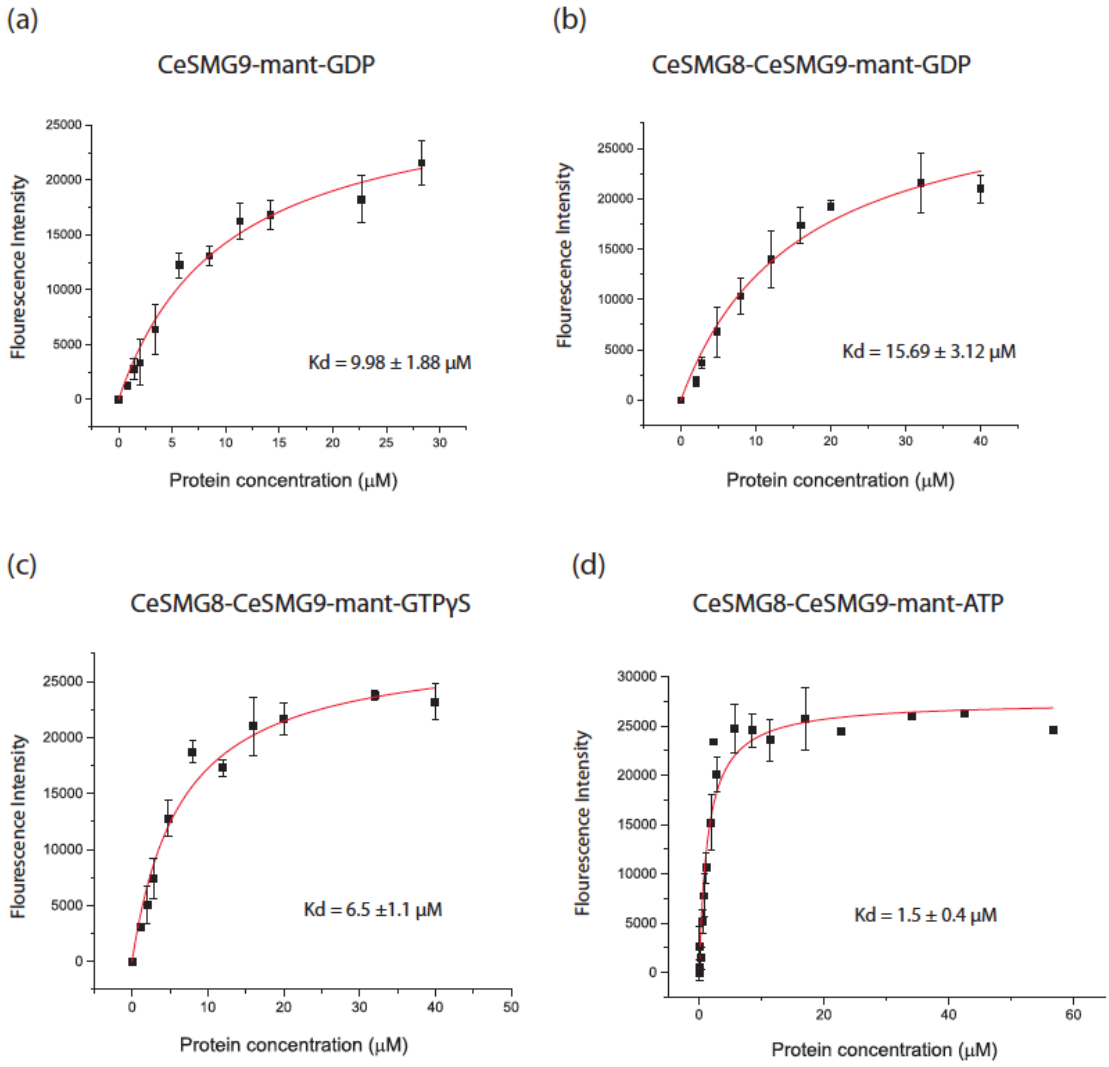


Figure 5.23: Measurement of nucleotides binding affinity of CeSMG9 and CeSMG8-CeSMG9 complex. (a). Measurement of GDP binding affinity of CeSMG9 (59-375), the G-like domain of CeSMG9. (b). Measurement of GDP binding affinity of the complex CeSMG8 (1-423)/CeSMG9 (59-375). (c). Measurement of GTPγS binding affinity of the complex CeSMG8 (1-423)/CeSMG9 (59-375). (d). Measurement of ATP binding affinity of the complex CeSMG8 (1-423)/CeSMG9 (59-375).

5.20 Fitting of CeSMG8-9 into the EM density of Human SMG1-8-9

The previously published negative-stain EM density of human SMG1-8-9 complex [47](EMD-2663) was fitted with the crystal structure of CeSMG8-9 complex together

with the homology model of human SMG1 from another published EM study[99] with UCSF Chimera[100]. Both the N-terminal HEAT repeats and the C-terminal kinase domain model of human SMG1 were fitted as two separate rigid bodies into the density at the contour level of 3.4 as suggested by the authors. The N-terminal HEAT repeats of SMG1 was fitted into the curved and tubular density (the "arm" region) by which the N-terminus of HEAT repeats is at the bottom and the C-terminus of HEAT repeats is close to the kinase domain (correlation coefficient: 0.74). The homology model of kinase domain was fitted into the "head" region of SMG1 (correlation coefficient: 0.77). After fitting of homology models of human SMG1 into the density map, the extra density for human SMG8-SMG9 can be fitted with our crystal structure of CeSMG8-9 core complex (correlation coefficient: 0.69). The fitting of CeSMG8-9 structure in the map shows that SMG9 is at the central position in SMG1-8-9 complex, with the G-motif loops pointing towards the HEAT repeat region of SMG1. The fitting is consistent with previously reported results that SMG9 directly interacts and forms a stable complex with SMG1[48]. Although the N-terminal unstructured region of CeSMG9 is absent in the crystal structure, the N-terminus of SMG9 might be an integral part of the density at the "arm" region. The fitting also indicates that the helix bundle domain of SMG8 points towards the C-terminal "head" region of SMG1, which allows the possible interaction between C-terminal "head" region of SMG1 and the C-terminal low-complexity region of SMG8 and might lead to the inhibition of SMG1 kinase activity. The fitting is supported by the crosslinking data reported previously that the residue Lys869 at the C-terminus of human SMG8 was cross-linked with the residue Lys2993 at the C-insertion of human SMG1[99]. Taken together, the fitting provided a reasonable model of SMG1-8-9 complex supported by the visible fitting quality and the previous knowledge on the interaction mode among SMG1, SMG8 and SMG9.

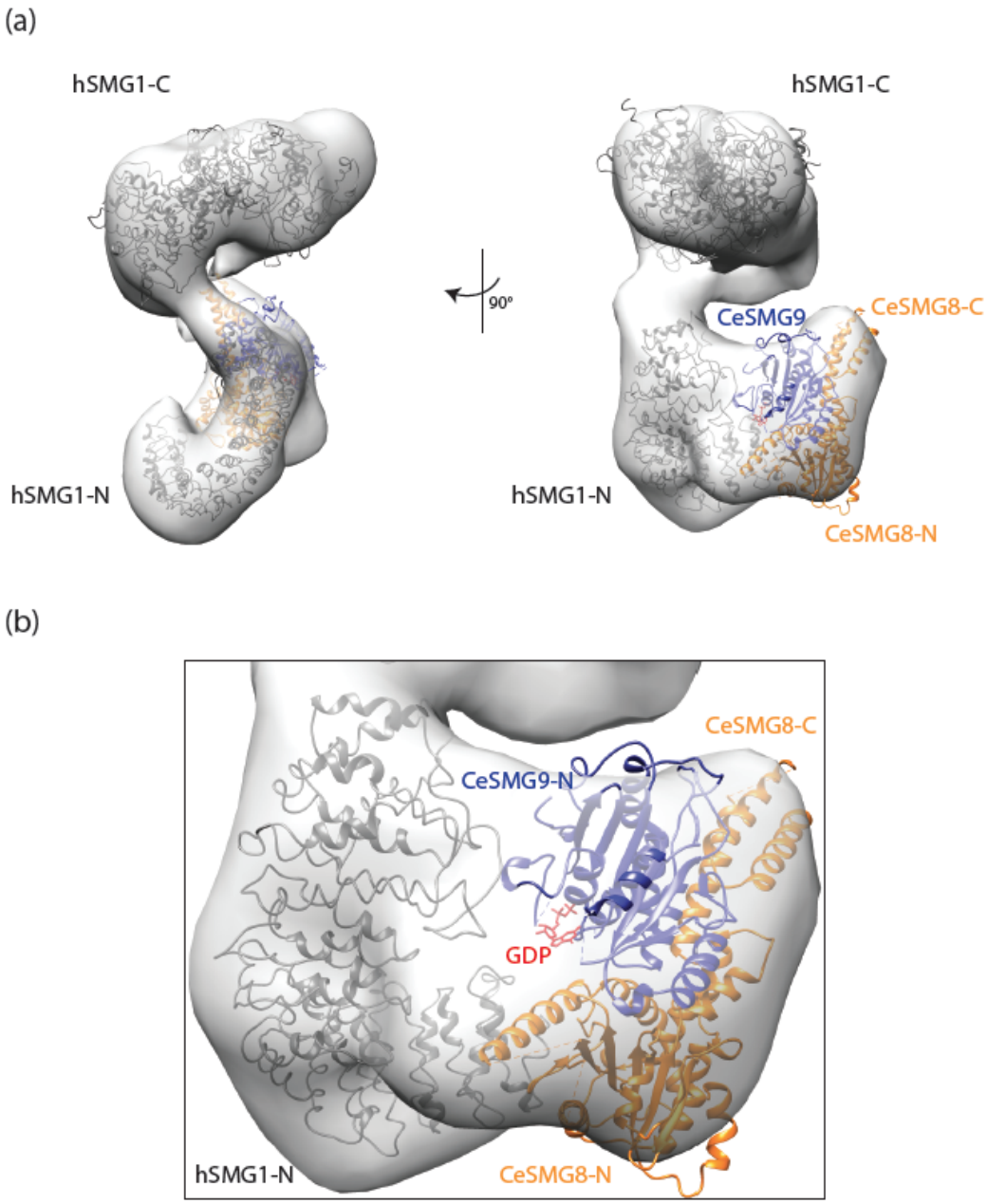


Figure 5.24: Fitting of CeSMG8-9 into the density of human SMG1-8-9. (a). The overview of the fitting in two orientations associated with 90 degrees rotation around a vertical axis. (b). The zoom-in view of models of hSMG1, CeSMG8-9-GDP complex.

6 Discussion and conclusions

Nonsense mediated mRNA decay (NMD) is an important mRNA quality control pathway conserved in eukaryotes. NMD targets aberrant mRNAs carrying premature stop codons (PTCs) for rapid degradation, and in doing so it prevents the accumulation of C-terminally truncated protein products that might otherwise be toxic to cells. NMD involves the concerted action of many trans-acting factors and it is a highly regulated process. The decisive event to trigger NMD is the phosphorylation of UPF1 by a PIKK kinase, SMG1. It was reported that in human cells, SMG8 and SMG9 form a heterodimer that interacts with SMG1 and inhibits its kinase activity. However, several important questions remained to be investigated, such as, how does SMG8 interact with SMG9 to form a heterodimer? What are the biochemical properties of SMG8 and SMG9? How do SMG8 and SMG9 work together to regulate SMG1 kinase activity? In order to address these questions, I carried out structural and biochemical analysis of SMG8 and SMG9. The results in this thesis work provide for the first time detailed structural information of the SMG8-SMG9 core complex from *C. elegans* and the structural basis for the nucleotides binding property of SMG9.

Structural characterization of human SMG8-SMG9 complex was impeded by the fact that both human SMG8 and SMG9 contain many large disordered regions based on secondary structure prediction (supplementary materials, S2), which would cause problems for producing crystals for X-ray crystallography. In comparison, the *C. elegans* SMG8 and SMG9 contain less flexible regions thus were used as subjects for structure determination.

The domain organization of CeSMG8 and CeSMG9 were not previously characterized, secondary structure predictions show that CeSMG8 contains a structured N-terminal region (1-423), followed by a large disordered region and a small C-terminal structured region. CeSMG9 contains two distinct regions, a N-terminal disordered region (1-58) and a C-terminal folded domain (59-385). The corresponding C-terminal domain in human SMG9 was predicted as a putative NTPase with analysis of sequence[48].

The expression tests of CeSMG8 and CeSMG9 were tried in *E. coli* first. The full-length as well as N-terminally truncated CeSMG9 could be expressed and purified in isolation from *E. coli*. The purified CeSMG9 (39-385) and CeSMG9 (59-385) were subjected to crystallization screening. However, no crystal hit was obtained from these

two constructs. The reason could be that CeSMG9 alone is not stable in solution based on the observation of significant precipitation of purified CeSMG9 protein after overnight incubation at 4 °C. CeSMG8 (1-423) was not soluble when it was expressed in isolation in *E. coli* but could be solubilized when it was co-expressed with full-length CeSMG9 and could be co-purified in the form of CeSMG8-CeSMG9 complex, indicating that the N-terminal structured region (1-423) of CeSMG8 is sufficient to bind CeSMG9. Although the CeSMG8-CeSMG9 complex could be purified from *E. coli*, the amount of purified protein obtained from large-scale expression was too little to perform crystallization experiments and in addition, protein degradation occurred after purification and most likely in the N-terminal intrinsically disordered region of CeSMG9.

To solve these two problems, CeSMG8 (1-423) and CeSMG9 (39-385) were co-expressed in insect cells with a baculovirus system. Milligram quantities of purified protein can be obtained and was used for crystallization screening. Hexagonal crystals were obtained within two days and were optimized to single big hexagonal crystals with seeding in sitting drops. Although the optimized crystals could diffract to around 3 Å resolution, crystals suffered severe twinning defects that made structure determination not feasible, particularly because there was no suitable model for phasing by molecular replacement. A first strategy I used to try and solve the twinning problem was to make a shorter construct of CeSMG9 in the CeSMG8-CeSMG9 complex with the hope of changing crystal packing. CeSMG9 (59-385), a further N-terminally truncated CeSMG9 construct that contains only the C-terminal domain was co-expressed with CeSMG8 (1-423) in insect cells. After purification, two distinct bands appeared close to the position of CeSMG9 (59-385) on SDS-PAGE gel, which could be the result of degradation of CeSMG9. Results of mass spectrometry analysis showed that the degradation occurred at the C-terminus of CeSMG9 and identified a shorter fragment CeSMG9 (59-375). However, no new crystal form was obtained with this shorter construct and the same hexagonal crystal can be obtained but still had twinning problem, which suggested that the N-terminal and C-terminal truncation of CeSMG9 did not affect the crystal packing.

It was reported that yttrium chloride could be used in crystallization to affect the crystal packing, based on previous experiences that the yttrium ion can coordinate with surrounding aspartic acid or glutamic acid residues of protein[91]. The usage of yttrium chloride as an additive was a successful strategy to solve the twinning problem. The

complex CeSMG8 (1-423)/CeSMG9 (59-375) was crystallized in presence of yttrium chloride and crystals with a different morphology (pyramid-like) were obtained that were not twinned. Selenium-methionine derivatized crystals were obtained with same crystallization method and in addition the crystal was optimized with paraffin oil to increase the crystal size and boost anomalous signal, which led to the successful structure determination.

The crystal structure of CeSMG8-CeSMG9 core complex was solved at the resolution of 2.5 Å. The structure of CeSMG8 revealed its domain organization that the previously predicted N-terminal structured region (1-423) consists of two distinct domains, a N-terminal G-like domain (1-323) and a middle helical bundle domain (334-416) connected by a linker of 10 residues. The structure of CeSMG9 C-terminal region (59-363) also resembles a G-domain. Both G-like domains of CeSMG8 and CeSMG9 are structurally similar to the GTPase domain of dynamin-like family proteins hGBP1 and Atlastin. This led to the question that whether CeSMG8 and CeSMG9 are able to bind GTP. I soaked the native crystals with either GTP and its analogue or GDP but crystals could only survive when soaked with GDP. Therefore, the crystal structure of GDP-bound CeSMG8-CeSMG9 core complex was solved by molecular replacement. The structure revealed that GDP only binds CeSMG9 but not CeSMG8. Residues in the P-loop responsible for phosphate binding are indeed conserved in CeSMG9 but not in CeSMG8. The GDP and GTP binding affinity of CeSMG9 in vitro is around 10-15 μM and 6.5 μM respectively, as measured by fluorescence anisotropy. In general, the low-micromolar binding affinity we measured for CeSMG8-CeSMG9 is similar to that reported for hGBP1[101]. Although CeSMG9 binds GDP with its conserved P-loop, the base recognition is not specific: because the non-canonical G4 motif of SMG9 lacks the conserved Aspartic acid residue that recognizes the guanine base by specific hydrogen bonding. The unspecific nucleotide binding property of CeSMG9 was confirmed by determining the crystal structure of ADP-bound CeSMG8-CeSMG9 complex, which was solved by using crystals soaked with ADP. The structure revealed that the binding mode of ADP is the same as GDP. The ADP molecule binds to the conserved P-loop of CeSMG9 and the adenine base is not coordinated by specific residues. The ATP-binding affinity of CeSMG9 is around 1.5 μM measured by fluorescence anisotropy. Like CeSMG9, human SMG9 also contains a conserved P-loop and a non-canonical G4 motif therefore it is very likely that human SMG9 also binds nucleotides without selectivity. In addition, both

CeSMG9 and hSMG9 did not show GTPase activity in GTPase assays. This is possibly because the glycine residue that is critical for GTP hydrolysis in classical GTPases (G3 motif DxxG) is replaced with a residue proline in SMG9 (G3 motif DxxP). In this context, we note that SMG9 has been reported to interact with a GTPase-activating protein[102], which might activate SMG9 activity by providing a catalytic residue.

The relative position of the G-like domains in the CeSMG8-CeSMG9 heterodimer is remarkably similar to that observed in active dimeric GTPases of the dynamin family[97,98]. In particular, CeSMG8 and CeSMG9 converge at the loops that are known to harbor the nucleotide-binding motifs (G motifs) in canonical GTPases. While many of the conserved G-motif residues are present in CeSMG9, in CeSMG8 they are absent. Furthermore, the single stalk domain in the CeSMG8-CeSMG9 heterodimer has a different relative position as compared to the conformations observed in dynamin-like proteins[92,93]. Finally, while proteins such as Atlastin or hGBP1 dimerize in the presence of GTP analogues[92,97], the SMG8-SMG9 heterodimer is formed in the absence of nucleotides.

CeSMG8 interacts with CeSMG9 extensively along the binding interface. Both G-like domain and helical bundle domain of CeSMG8 interact with the G-like domain of CeSMG9 with charged interaction and hydrophobic interaction. Co-immunoprecipitation assays with human SMG8 and SMG9 mutants revealed that the interaction mode between CeSMG8 and CeSMG9 is conserved between hSMG8 and hSMG9. The different effects of two hSMG9 mutants in abolishing hSMG8-binding indicated that the binding is stronger at the N-terminal binding site than at the C-terminal binding site of SMG8. The stronger binding at the N-terminal binding site of SMG8 can be explained by the structural information of CeSMG8-9 that both hydrophobic and charged interactions and also more residues are involved at the N-terminal binding site of CeSMG8, while fewer residues are involved at the C-terminal binding site of CeSMG8. I speculate that it would be easier and more efficient for the conformation-flexible C-terminal helical bundle domain of SMG8 to dock on SMG9 if the N-terminal domain of SMG8 is already tightly bound on SMG9.

The fitting of CeSMG8-9 structure into the EM density of human SMG1-8-9 provided a reasonable model for SMG1-8-9 complex. The pseudo-atomic model of SMG1-8-9 complex revealed that SMG9 is in the central position and critical for the integrity of the whole complex. The G-like domain of SMG9 interacts with the N-terminal HEAT repeats of SMG1. Although the N-terminal unstructured region of CeSMG9 is absent

in the crystal structure, the N-terminus of SMG9 might be an integral part of the density at the "arm" region. Our fitting model also provided the structural insights into how SMG8 probably interacts with SMG1. The SMG1-8-9 model revealed that the N-terminus of SMG1 is close to the G-like domain of SMG8 and their possible interaction is supported by the previous published crosslinking data[99] which showed the residue Lys170 at the N-terminus of HEAT repeats of SMG1 was cross-linked with the residue Lys290 in the SMG8 G-like domain. The conformation of CeSMG8 helical bundle domain is fixed by docking on the CeSMG9, which would orient the SMG8 C-terminal region to the C-terminal "head" region of SMG1 as revealed in the fitting model of SMG1-8-9 and is consistent with the previous published crosslinking data[99] which showed that the residue Lys869 at the C-terminus of human SMG8 was cross-linked with the residue Lys2993 at the C- insertion region of human SMG1. Therefore, SMG8 is likely to interact with both the N-terminus and C-terminus of SMG1 with its N-terminal G-like domain and C-terminus, respectively. The model also has predictive value, as it raises the hypothesis that the switch regions of the SMG9 G-like domain might become ordered upon SMG1 binding and therefore that the nucleotide-binding state of SMG9 might impact on the conformation of the entire complex.

To conclude, in this thesis, I have solved the crystal structures of *C. elegans* SMG8-9 core complex in its Apo, GDP and ADP-bound forms at the resolution of 2.49, 2.64, 2.65 Å respectively. The structures characterized interactions between CeSMG8 and CeSMG9, CeSMG9 and nucleotides. The conservation of SMG8-9 interaction was confirmed by the Co-IP experiment with human SMG8-9. In addition, the fitting model of SMG1-8-9 provided structural insights into how SMG8 and SMG9 interact with SMG1 to assemble a stable complex. This fitting model together with previously published biochemical data have helped us significantly to understand the molecular mechanism underlying the regulation of kinase SMG1 by its two regulators, SMG8 and SMG9. This model paves the way to future studies to understand whether and how the nucleotide-binding properties of SMG8-SMG9 affect the function of the SMG1 kinase.

Supplementary materials

S1. Table 1

Data Collection and Refinement Statistics			
Dataset	CeSMG8-9-Apo	CeSMG8-9-GDP	CeSMG8-9-ADP
Wavelength (Å)	0.979	0.9785	1.2547
Resolution range (Å)[a]	52.36 - 2.493 (2.583 - 2.493)	47.48 - 2.64 (2.734 - 2.64)	62.71 - 2.649 (2.744 - 2.649)
Space group	P 32 2 1	P 32 2 1	P 32 2 1
a, b, c (Å)	111.085 111.085 374.474	110.605 110.605 360.266	111.129 111.129 376.232
α, β, γ (°)	90 90 120	90 90 120	90 90 120
Total reflections[a]	3768766 (360260)	764628 (71339)	752149 (22133)
Unique reflections[a]	94467 (9210)	76172 (7433)	76844 (6129)
Multiplicity[a]	39.9 (39.1)	10.0 (9.6)	9.8 (3.6)
Completeness (%)[a]	1.00 (0.98)	1.00 (0.99)	0.97 (0.78)
Mean I/sigma(I)[a]	27.25 (1.61)	12.39 (1.85)	16.40 (0.76)
R-merge[a]	0.1374 (2.587)	0.1198 (1.09)	0.1019 (1.358)
CC1/2[a]	1 (0.783)	0.998 (0.793)	0.999 (0.349)
Refinement			
R-work	0.2161	0.2329	0.2442
R-free	0.2631	0.2827	0.2830
Average B-factor	86.56	76.57	88.62
Ligands		GDP, Mg, EDO	ADP
Water	48	78	39
Stereochemistry			
RMS (bonds)	0.007	0.005	0.010
RMS (angles)	1.01	0.84	1.06
Ramachandran favored (%)	95	95	96
Ramachandran allowed (%)	3.9	4.5	3.6
Ramachandran outliers (%)	0.9	0.75	0.39
[a]Values in parentheses correspond to the highest-resolution shell.			

S2. Prediction of disordered regions in SMG8 and SMG9

Sequences of SMG8 and SMG9 from both *C. elegans* and humans were submitted to DISOPRED3[103] to predict the intrinsic disordered regions. Amino acids in the input sequence are considered disordered when the blue line is above the grey dashed line, that is the confidence score higher than 0.5. The orange line shows the confidence of disordered protein binding residue predictions. The results shown below indicate that SMG8 and SMG9 from *C. elegans* contain less disordered regions than their human orthologues. The results also show that both CeSMG8 and hSMG8 contain a large disordered region in the middle of their sequences, which separates the N-terminal and C-terminal structured regions. The predictions also show that both CeSMG9 and hSMG9 contain a N-terminal large disordered region followed by a C-terminal structured region. The results suggest that SMG8 as well as SMG9 might have similar domain organizations in *C. elegans* and humans.

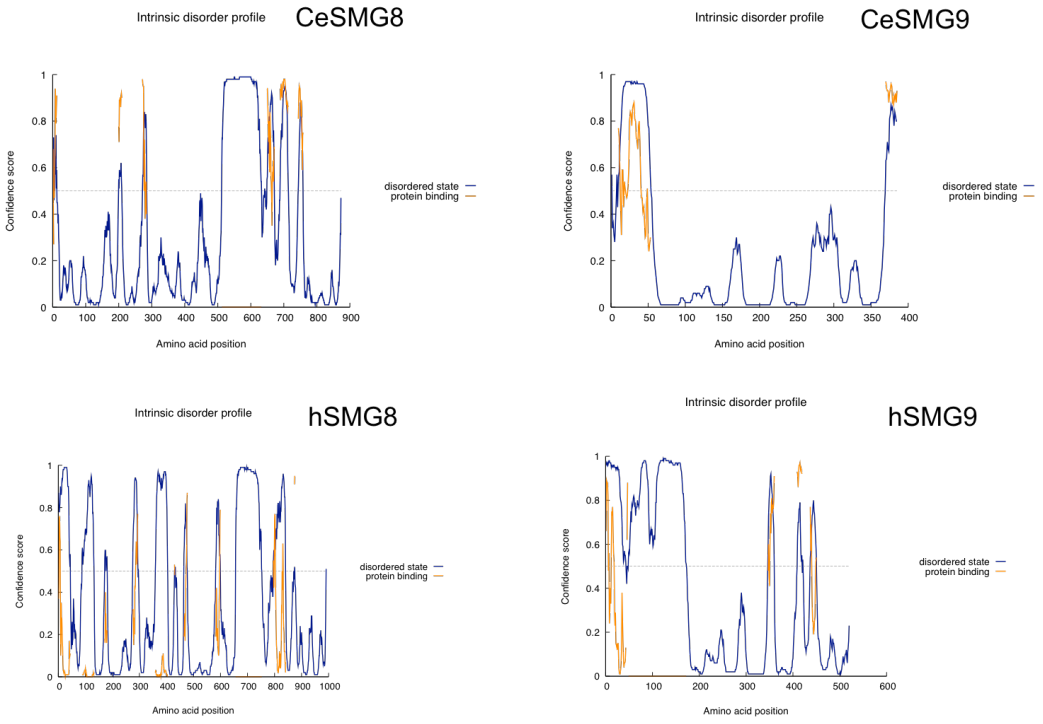
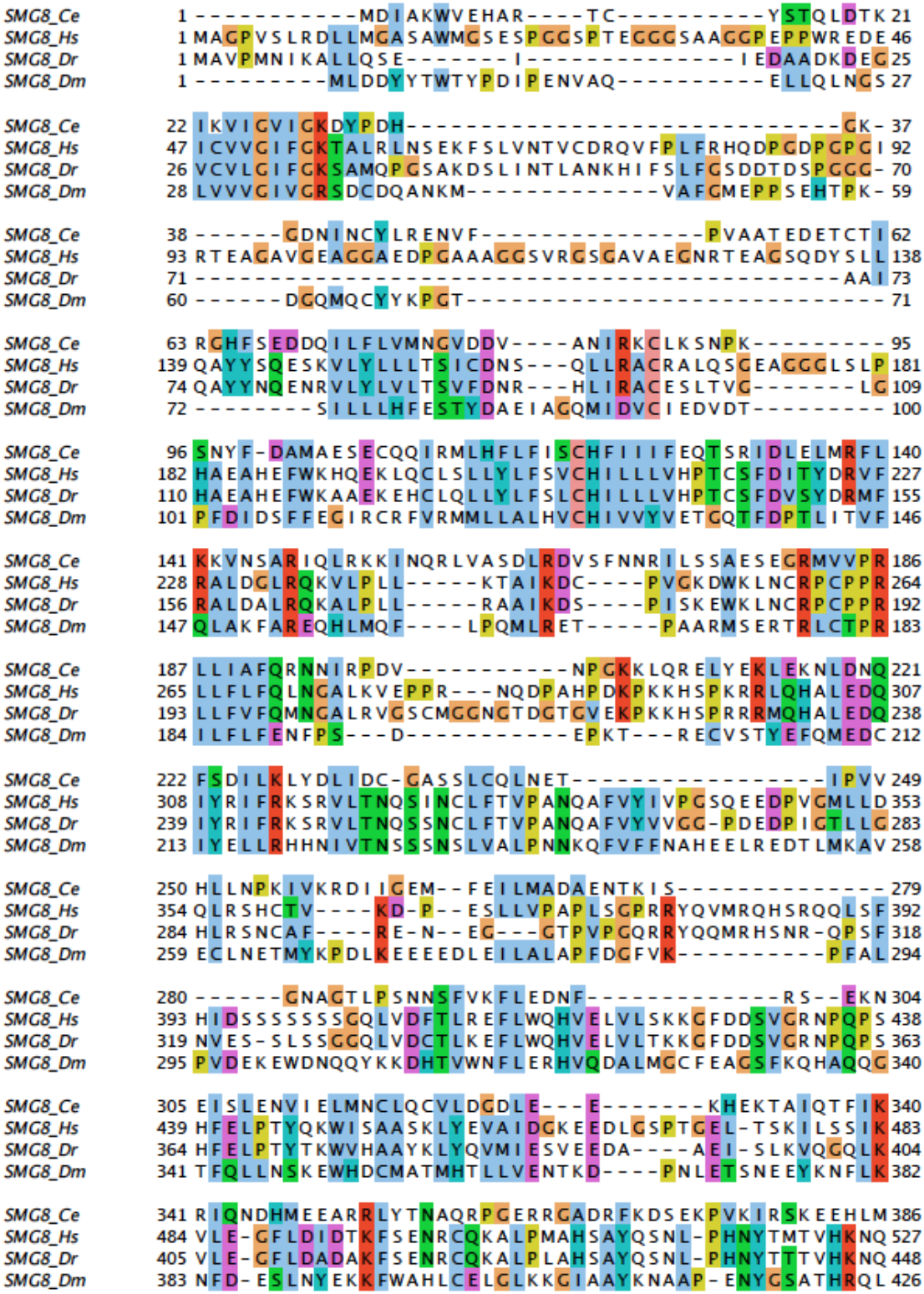


Figure S1: Prediction of disordered regions in SMG8 and SMG9 from *C. elegans* and humans.

S3. Sequence alignments of SMG8



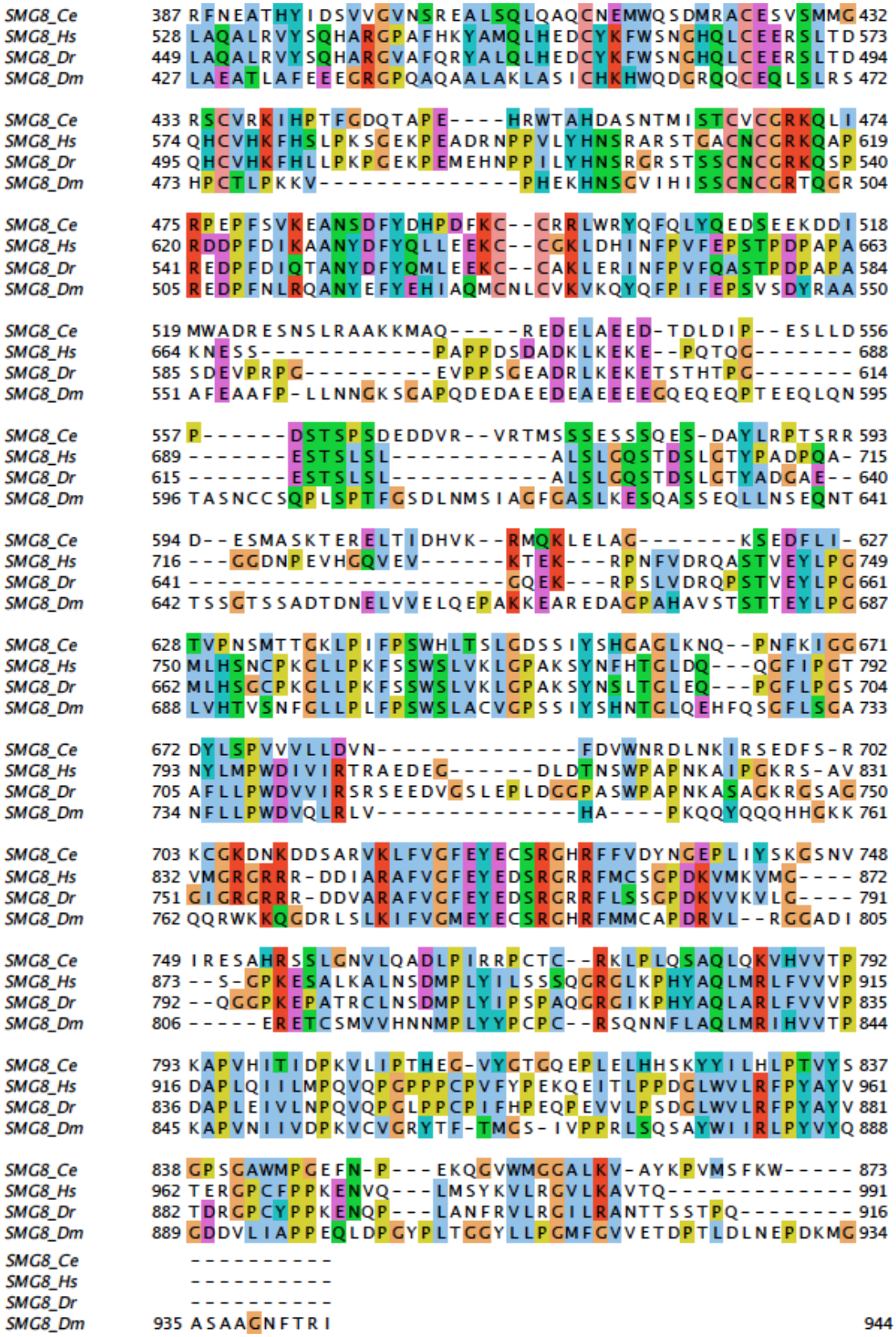


Figure S2: Sequence alignment of SMG8. *Ce: Caenorhabditis elegans; Hs: Homo sapiens; Dr: Danio rerio; Dm: Drosophila melanogaster.*

S4. Sequence alignments of SMG9

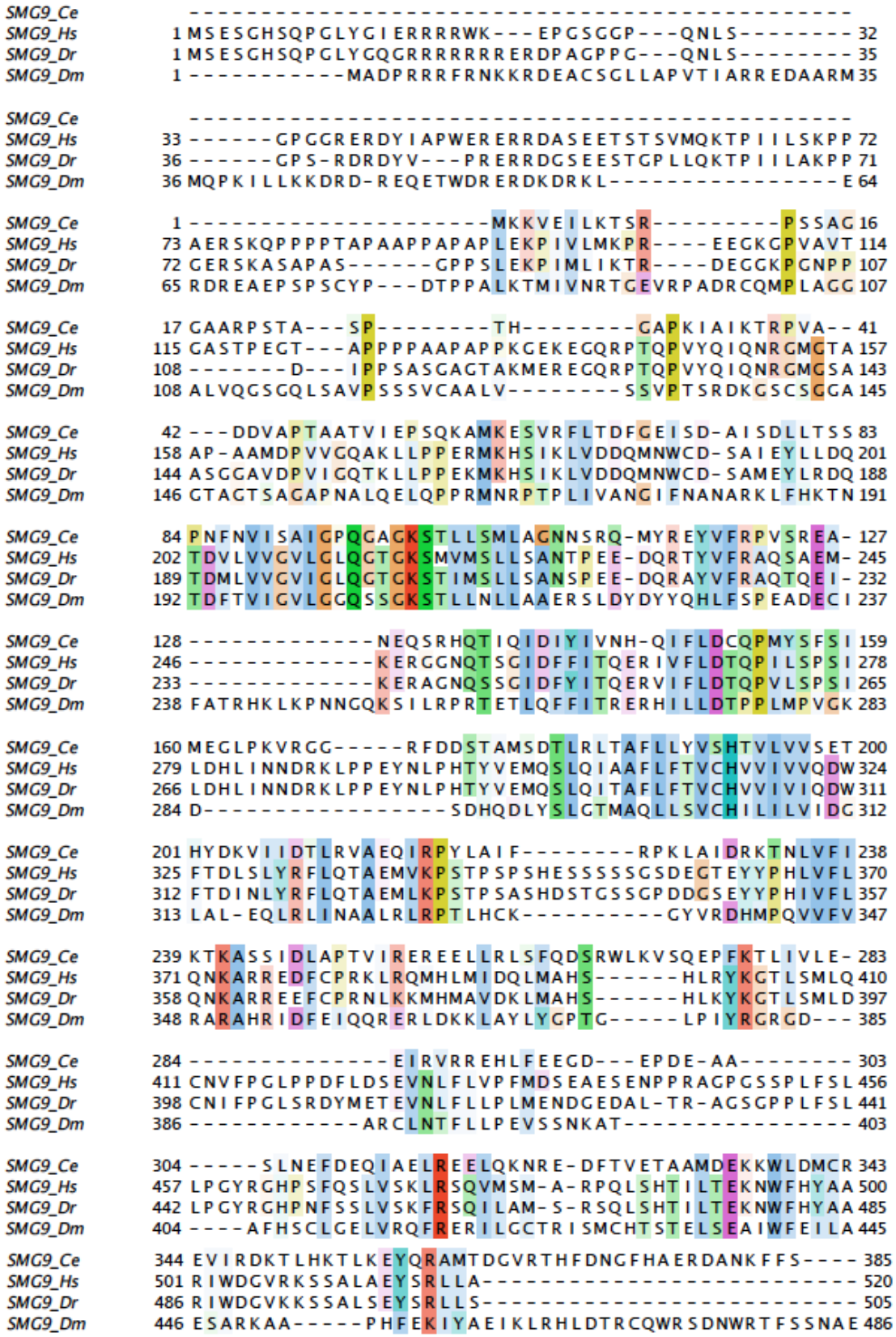


Figure S3: Sequence alignment of SMG9. *Ce: Caenorhabditis elegans; Hs: Homo sapiens; Dr: Danio rerio; Dm: Drosophila melanogaster.*

References

1 Karousis, E. D., Nasif, S. & Muhlemann, O. Nonsense-mediated mRNA decay: novel mechanistic insights and biological impact. *Wiley Interdiscip Rev RNA* **7**, 661-682, doi:10.1002/wrna.1357 (2016).

2 Doma, M. K. & Parker, R. Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature* **440**, 561-564, doi:10.1038/nature04530 (2006).

3 Shoemaker, C. J., Eyler, D. E. & Green, R. Dom34:Hbs1 promotes subunit dissociation and peptidyl-tRNA drop-off to initiate no-go decay. *Science* **330**, 369-372, doi:10.1126/science.1192430 (2010).

4 Graille, M. & Seraphin, B. Surveillance pathways rescuing eukaryotic ribosomes lost in translation. *Nat Rev Mol Cell Biol* **13**, 727-735, doi:10.1038/nrm3457 (2012).

5 Frischmeyer, P. A. *et al.* An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science* **295**, 2258-2261, doi:10.1126/science.1067338 (2002).

6 Klauer, A. A. & van Hoof, A. Degradation of mRNAs that lack a stop codon: a decade of nonstop progress. *Wiley Interdiscip Rev RNA* **3**, 649-660, doi:10.1002/wrna.1124 (2012).

7 Losson, R. & Lacroute, F. Interference of nonsense mutations with eukaryotic messenger RNA stability. *Proc Natl Acad Sci U S A* **76**, 5134-5137 (1979).

8 Maquat, L. E., Kinniburgh, A. J., Rachmilewitz, E. A. & Ross, J. Unstable beta-globin mRNA in mRNA-deficient beta o thalassemia. *Cell* **27**, 543-553 (1981).

9 Peltz, S. W., Brown, A. H. & Jacobson, A. mRNA destabilization triggered by premature translational termination depends on at least three cis-acting sequence elements and one trans-acting factor. *Genes Dev* **7**, 1737-1754 (1993).

10 Rehwinkel, J., Raes, J. & Izaurralde, E. Nonsense-mediated mRNA decay: Target genes and functional diversification of effectors. *Trends Biochem Sci* **31**, 639-646, doi:10.1016/j.tibs.2006.09.005 (2006).

11 Guan, Q. *et al.* Impact of nonsense-mediated mRNA decay on the global expression profile of budding yeast. *PLoS Genet* **2**, e203, doi:10.1371/journal.pgen.0020203 (2006).

12 Wittmann, J., Hol, E. M. & Jack, H. M. hUPF2 silencing identifies physiologic substrates of mammalian nonsense-mediated mRNA decay. *Mol Cell Biol* **26**, 1272-1287, doi:10.1128/MCB.26.4.1272-1287.2006 (2006).

13 Chan, W. K. *et al.* An alternative branch of the nonsense-mediated decay pathway. *EMBO J* **26**, 1820-1830, doi:10.1038/sj.emboj.7601628 (2007).

14 Ramani, A. K. *et al.* High resolution transcriptome maps for wild-type and nonsense-mediated decay-defective Caenorhabditis elegans. *Genome Biol* **10**, R101, doi:10.1186/gb-2009-10-9-r101 (2009).

15 Huang, L. & Wilkinson, M. F. Regulation of nonsense-mediated mRNA decay. *Wiley Interdiscip Rev RNA* **3**, 807-828, doi:10.1002/wrna.1137 (2012).

16 He, F. & Jacobson, A. Nonsense-Mediated mRNA Decay: Degradation of Defective Transcripts Is Only Part of the Story. *Annu Rev Genet* **49**, 339-366, doi:10.1146/annurev-genet-112414-054639 (2015).

17 Ge, Y. & Porse, B. T. The functional consequences of intron retention: alternative splicing coupled to NMD as a regulator of gene expression. *Bioessays* **36**, 236-243, doi:10.1002/bies.201300156 (2014).

18 Gaba, A., Jacobson, A. & Sachs, M. S. Ribosome occupancy of the yeast CPA1 upstream open reading frame termination codon modulates nonsense-mediated mRNA decay. *Mol Cell* **20**, 449-460, doi:10.1016/j.molcel.2005.09.019 (2005).

19 Nyiko, T., Sonkoly, B., Merai, Z., Benkovics, A. H. & Silhavy, D. Plant upstream ORFs can trigger nonsense-mediated mRNA decay in a size-dependent manner. *Plant Mol Biol* **71**, 367-378, doi:10.1007/s11103-009-9528-4 (2009).

20 Kebaara, B. W. & Atkin, A. L. Long 3'-UTRs target wild-type mRNAs for nonsense-mediated mRNA decay in Saccharomyces cerevisiae. *Nucleic Acids Res* **37**, 2771-2778, doi:10.1093/nar/gkp146 (2009).

21 Kertesz, S. *et al.* Both introns and long 3'-UTRs operate as cis-acting elements to trigger nonsense-mediated decay in plants. *Nucleic Acids Res* **34**, 6147-6157, doi:10.1093/nar/gkl737 (2006).

22 Leeds, P., Peltz, S. W., Jacobson, A. & Culbertson, M. R. The product of the yeast UPF1 gene is required for rapid turnover of mRNAs containing a premature translational termination codon. *Genes Dev* **5**, 2303-2314 (1991).

23 Leeds, P., Wood, J. M., Lee, B. S. & Culbertson, M. R. Gene products that promote mRNA turnover in Saccharomyces cerevisiae. *Mol Cell Biol* **12**, 2165-2177 (1992).

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

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

26 Page, M. F., Carr, B., Anders, K. R., Grimson, A. & Anderson, P. SMG-2 is a phosphorylated protein required for mRNA surveillance in Caenorhabditis elegans and related to Upf1p of yeast. *Mol Cell Biol* **19**, 5943-5951 (1999).

27 Yamashita, A. *et al.* 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, doi:10.1101/gad.1767209 (2009).

28 Hug, N., Longman, D. & Caceres, J. F. Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Res* **44**, 1483-1495, doi:10.1093/nar/gkw010 (2016).

29 Chakrabarti, S. *et al.* Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. *Mol Cell* **41**, 693-703, doi:10.1016/j.molcel.2011.02.010 (2011).

30 Weng, Y., Czaplinski, K. & Peltz, S. W. Genetic and biochemical characterization of mutations in the ATPase and helicase regions of the Upf1 protein. *Mol Cell Biol* **16**, 5477-5490 (1996).

31 Franks, T. M., Singh, G. & Lykke-Andersen, J. Upf1 ATPase-dependent mRNP disassembly is required for completion of nonsense- mediated mRNA decay. *Cell* **143**, 938-950, doi:10.1016/j.cell.2010.11.043 (2010).

32 Fiorini, F., Boudvillain, M. & Le Hir, H. Tight intramolecular regulation of the human Upf1 helicase by its N- and C-terminal domains. *Nucleic Acids Res* **41**, 2404-2415, doi:10.1093/nar/gks1320 (2013).

33 Chamieh, H., Ballut, L., Bonneau, F. & Le Hir, H. NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. *Nat Struct Mol Biol* **15**, 85-93, doi:10.1038/nsmb1330 (2008).

34 Weng, Y., Czaplinski, K. & Peltz, S. W. ATP is a cofactor of the Upf1 protein that modulates its translation termination and RNA binding activities. *RNA* **4**, 205-214 (1998).

35 Melero, R. *et al.* The cryo-EM structure of the UPF-EJC complex shows UPF1 poised toward the RNA 3' end. *Nat Struct Mol Biol* **19**, 498-505, S491-492, doi:10.1038/nsmb.2287 (2012).

36 Clerici, M. *et al.* Unusual bipartite mode of interaction between the nonsense-mediated decay factors, UPF1 and UPF2. *EMBO J* **28**, 2293-2306, doi:10.1038/emboj.2009.175 (2009).

37 Kadlec, J., Izaurralde, E. & Cusack, S. The structural basis for the interaction between nonsense-mediated mRNA decay factors UPF2 and UPF3. *Nat Struct Mol Biol* **11**, 330-337, doi:10.1038/nsmb741 (2004).

38 Serin, G., Gersappe, A., Black, J. D., Aronoff, R. & Maquat, L. E. Identification and characterization of human orthologues to Saccharomyces cerevisiae Upf2 protein and Upf3 protein (Caenorhabditis elegans SMG-4). *Mol Cell Biol* **21**, 209-223, doi:10.1128/MCB.21.1.209-223.2001 (2001).

39 Lykke-Andersen, J., Shu, M. D. & Steitz, J. A. Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon. *Cell* **103**, 1121-1131 (2000).

40 Conti, E. & Izaurralde, E. Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. *Curr Opin Cell Biol* **17**, 316-325, doi:10.1016/j.ceb.2005.04.005 (2005).

41 Chan, W. K. *et al.* A UPF3-mediated regulatory switch that maintains RNA surveillance. *Nat Struct Mol Biol* **16**, 747-753, doi:10.1038/nsmb.1612 (2009).

42 Kunz, J. B., Neu-Yilik, G., Hentze, M. W., Kulozik, A. E. & Gehring, N. H. Functions of hUpf3a and hUpf3b in nonsense-mediated mRNA decay and translation. *RNA* **12**, 1015-1022, doi:10.1261/rna.12506 (2006).

43 Buchwald, G. *et al.* 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**, 10050-10055, doi:10.1073/pnas.1000993107 (2010).

44 Yamashita, A., Ohnishi, T., Kashima, I., Taya, Y. & Ohno, S. 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, doi:10.1101/gad.913001 (2001).

45 Yamashita, A. Role of SMG-1-mediated Upf1 phosphorylation in mammalian nonsense-mediated mRNA decay. *Genes Cells* **18**, 161-175, doi:10.1111/gtc.12033 (2013).

46 Denning, G., Jamieson, L., Maquat, L. E., Thompson, E. A. & Fields, A. P. Cloning of a novel phosphatidylinositol kinase-related kinase: characterization of the human SMG-1 RNA surveillance protein. *J Biol Chem* **276**, 22709-22714, doi:10.1074/jbc.C100144200 (2001).

47 Arias-Palomo, E. *et al.* The nonsense-mediated mRNA decay SMG-1 kinase is regulated by large-scale conformational changes controlled by SMG-8. *Genes Dev* **25**, 153-164, doi:10.1101/gad.606911 (2011).

48 Fernandez, I. S. *et al.* Characterization of SMG-9, an essential component of the nonsense-mediated mRNA decay SMG1C complex. *Nucleic Acids Res* **39**, 347-358, doi:10.1093/nar/gkq749 (2011).

49 Fukuhara, N. *et al.* SMG7 is a 14-3-3-like adaptor in the nonsense-mediated mRNA decay pathway. *Mol Cell* **17**, 537-547, doi:10.1016/j.molcel.2005.01.010 (2005).

50 Jonas, S., Weichenrieder, O. & Izaurralde, E. 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**, 211-225, doi:10.1101/gad.206672.112 (2013).

51 Ohnishi, T. *et al.* Phosphorylation of hUPF1 induces formation of mRNA surveillance complexes containing hSMG-5 and hSMG-7. *Mol Cell* **12**, 1187-1200 (2003).

52 Loh, B., Jonas, S. & Izaurralde, E. The SMG5-SMG7 heterodimer directly recruits the CCR4-NOT deadenylase complex to mRNAs containing nonsense codons via interaction with POP2. *Genes Dev* **27**, 2125-2138, doi:10.1101/gad.226951.113 (2013).

53 Kashima, I. *et al.* SMG6 interacts with the exon junction complex via two conserved EJC-binding motifs (EBMs) required for nonsense-mediated mRNA decay. *Genes Dev* **24**, 2440-2450, doi:10.1101/gad.604610 (2010).

54 Chakrabarti, S., Bonneau, F., Schussler, S., Eppinger, E. & Conti, E. Phospho-dependent and phospho-independent interactions of the helicase UPF1 with the NMD factors SMG5-SMG7 and SMG6. *Nucleic Acids Res* **42**, 9447-9460, doi:10.1093/nar/gku578 (2014).

55 Eberle, A. B., Lykke-Andersen, S., Muhlemann, O. & Jensen, T. H. SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. *Nat Struct Mol Biol* **16**, 49-55, doi:10.1038/nsmb.1530 (2009).

56 Glavan, F., Behm-Ansmant, I., Izaurralde, E. & Conti, E. Structures of the PIN domains of SMG6 and SMG5 reveal a nuclease within the mRNA surveillance complex. *EMBO J* **25**, 5117-5125, doi:10.1038/sj.emboj.7601377 (2006).

57 Huntzinger, E., Kashima, I., Fauser, M., Sauliere, J. & Izaurralde, E. SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. *RNA* **14**, 2609-2617, doi:10.1261/rna.1386208 (2008).

58 Nicholson, P., Josi, C., Kurosawa, H., Yamashita, A. & Muhlemann, O. A novel phosphorylation-independent interaction between SMG6 and UPF1 is essential for human NMD. *Nucleic Acids Res* **42**, 9217-9235, doi:10.1093/nar/gku645 (2014).

59 Le Hir, H., Izaurralde, E., Maquat, L. E. & Moore, M. J. The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J* **19**, 6860-6869, doi:10.1093/emboj/19.24.6860 (2000).

60 Le Hir, H., Sauliere, J. & Wang, Z. The exon junction complex as a node of post-transcriptional networks. *Nat Rev Mol Cell Biol* **17**, 41-54, doi:10.1038/nrm.2015.7 (2016).

61 Longman, D., Plasterk, R. H., Johnstone, I. L. & Caceres, J. F. Mechanistic insights and identification of two novel factors in the C. elegans NMD pathway. *Genes Dev* **21**, 1075-1085, doi:10.1101/gad.417707 (2007).

62 Gatfield, D., Unterholzner, L., Ciccarelli, F. D., Bork, P. & Izaurralde, E. Nonsense-mediated mRNA decay in Drosophila: at the intersection of the yeast and mammalian pathways. *EMBO J* **22**, 3960-3970, doi:10.1093/emboj/cdg371 (2003).

63 Ballut, L. *et al.* The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. *Nat Struct Mol Biol* **12**, 861-869, doi:10.1038/nsmb990 (2005).

64 Tange, T. O., Shibuya, T., Jurica, M. S. & Moore, M. J. Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. *RNA* **11**, 1869-1883, doi:10.1261/rna.2155905 (2005).

65 Andersen, C. B. *et al.* Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. *Science* **313**, 1968-1972, doi:10.1126/science.1131981 (2006).

66 Bono, F., Ebert, J., Lorentzen, E. & Conti, E. The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell* **126**, 713-725, doi:10.1016/j.cell.2006.08.006 (2006).

67 Longman, D. *et al.* DHX34 and NBAS form part of an autoregulatory NMD circuit that regulates endogenous RNA targets in human cells, zebrafish and Caenorhabditis elegans. *Nucleic Acids Res* **41**, 8319-8331, doi:10.1093/nar/gkt585 (2013).

68 Melero, R. *et al.* The RNA helicase DHX34 functions as a scaffold for SMG1-mediated UPF1 phosphorylation. *Nat Commun* **7**, 10585, doi:10.1038/ncomms10585 (2016).

69 Durand, S. & Lykke-Andersen, J. Nonsense-mediated mRNA decay occurs during eIF4F-dependent translation in human cells. *Nat Struct Mol Biol* **20**, 702-709, doi:10.1038/nsmb.2575 (2013).

70 Rufener, S. C. & Muhlemann, O. eIF4E-bound mRNPs are substrates for nonsense-mediated mRNA decay in mammalian cells. *Nat Struct Mol Biol* **20**, 710-717, doi:10.1038/nsmb.2576 (2013).

71 Kurosaki, T. & Maquat, L. E. Nonsense-mediated mRNA decay in humans at a glance. *J Cell Sci* **129**, 461-467, doi:10.1242/jcs.181008 (2016).

72 Nagy, E. & Maquat, L. E. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci* **23**, 198-199 (1998).

73 Amrani, N. *et al.* A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature* **432**, 112-118, doi:10.1038/nature03060 (2004).

74 Behm-Ansmant, I., Gatfield, D., Rehwinkel, J., Hilgers, V. & Izaurralde, E. A conserved role for cytoplasmic poly(A)-binding protein 1 (PABPC1) in nonsense-mediated mRNA decay. *EMBO J* **26**, 1591-1601, doi:10.1038/sj.emboj.7601588 (2007).

75 Singh, G., Rebbapragada, I. & Lykke-Andersen, J. A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay. *PLoS Biol* **6**, e111, doi:10.1371/journal.pbio.0060111 (2008).

76 Kononenko, A. V. *et al.* GTP-dependent structural rearrangement of the eRF1:eRF3 complex and eRF3 sequence motifs essential for PABP binding. *Nucleic Acids Res* **38**, 548-558, doi:10.1093/nar/gkp908 (2010).

77 Cosson, B. *et al.* Poly(A)-binding protein and eRF3 are associated in vivo in human and Xenopus cells. *Biol Cell* **94**, 205-216 (2002).

78 Eberle, A. B., Stalder, L., Mathys, H., Orozco, R. Z. & Muhlemann, O. Posttranscriptional gene regulation by spatial rearrangement of the 3' untranslated region. *PLoS Biol* **6**, e92, doi:10.1371/journal.pbio.0060092 (2008).

79 Ivanov, P. V., Gehring, N. H., Kunz, J. B., Hentze, M. W. & Kulozik, A. E. Interactions between UPF1, eRFs, PABP and the exon junction complex suggest an integrated model for mammalian NMD pathways. *EMBO J* **27**, 736-747, doi:10.1038/emboj.2008.17 (2008).

80 Kashima, I. *et al.* Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. *Genes Dev* **20**, 355-367, doi:10.1101/gad.1389006 (2006).

81 Silva, A. L., Ribeiro, P., Inacio, A., Liebhaber, S. A. & Romao, L. Proximity of the poly(A)-binding protein to a premature termination codon inhibits mammalian nonsense-mediated mRNA decay. *RNA* **14**, 563-576, doi:10.1261/rna.815108 (2008).

82 Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132, doi:10.1107/S0907444909047337 (2010).

83 Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr A* **64**, 112-122, doi:10.1107/S0108767307043930 (2008).

84 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).

85 McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674, doi:10.1107/S0021889807021206 (2007).

86 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132, doi:10.1107/S0907444904019158 (2004).

87 Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**, 352-367, doi:10.1107/S0907444912001308 (2012).

88 Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539, doi:10.1038/msb.2011.75 (2011).

89 Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191, doi:10.1093/bioinformatics/btp033 (2009).

90 Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **58**, 1948-1954 (2002).

91 Zhang, F., Zocher, G., Sauter, A., Stehle, T. & Schreiber, F. Novel approach to controlled protein crystallization through ligandation of yttrium cations. *Journal of Applied Crystallography* **44**, 755-762, doi:10.1107/s0021889811017997 (2011).

92 Byrnes, L. J. & Sondermann, H. Structural basis for the nucleotide-dependent dimerization of the large G protein atlastin-1/SPG3A. *Proc Natl Acad Sci U S A* **108**, 2216-2221, doi:10.1073/pnas.1012792108 (2011).

93 Prakash, B., Praefcke, G. J., Renault, L., Wittinghofer, A. & Herrmann, C. Structure of human guanylate-binding protein 1 representing a unique class of GTP-binding proteins. *Nature* **403**, 567-571, doi:10.1038/35000617 (2000).

94 Stivala, A., Wybrow, M., Wirth, A., Whisstock, J. C. & Stuckey, P. J. Automatic generation of protein structure cartoons with Pro-origami. *Bioinformatics* **27**, 3315-3316, doi:10.1093/bioinformatics/btr575 (2011).

95 Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-797, doi:10.1016/j.jmb.2007.05.022 (2007).

96 Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**, W545-549, doi:10.1093/nar/gkq366 (2010).

97 Ghosh, A., Praefcke, G. J., Renault, L., Wittinghofer, A. & Herrmann, C. How guanylate-binding proteins achieve assembly-stimulated processive cleavage of GTP to GMP. *Nature* **440**, 101-104, doi:10.1038/nature04510 (2006).

98 Byrnes, L. J. *et al.* Structural basis for conformational switching and GTP loading of the large G protein atlastin. *EMBO J* **32**, 369-384, doi:10.1038/emboj.2012.353 (2013).

99 Deniaud, A. *et al.* 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, doi:10.1093/nar/gkv668 (2015).

100 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).

101 Praefcke, G. J. *et al.* Identification of residues in the human guanylate-binding protein 1 critical for nucleotide binding and cooperative GTP hydrolysis. *J Mol Biol* **344**, 257-269, doi:10.1016/j.jmb.2004.09.026 (2004).

102 Takeda, S. *et al.* Role of a tyrosine phosphorylation of SMG-9 in binding of SMG-9 to IQGAP and the NMD complex. *Biochem Biophys Res Commun* **410**, 29-33, doi:10.1016/j.bbrc.2011.05.099 (2011).

103 Buchan, D. W., Minneci, F., Nugent, T. C., Bryson, K. & Jones, D. T. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res* **41**, W349-357, doi:10.1093/nar/gkt381 (2013).

Acknowledgements

Finally, I would like to thank all the people who have contributed to this work.
I am very grateful to my supervisor Prof. Dr. Elena Conti for giving me the great opportunity of doing my PhD training in this outstanding laboratory. Her strong support and strategic advices have definitely played the fundamental role in helping me to complete my PhD project successfully. I am also very grateful to Dr. Jérôme Basquin for his great help and direct supervision to this project and teaching me a lot of knowledge and hands-on experiences in crystallography. Thanks to Professor Hervé Le Hir, Dr. Esben Lorentzen and Dr. Christian Biertuempfel for participating in my thesis advisory committee and providing many great suggestions. To Fabien Bonneau and Mahesh Lingaraju for their contributions to this project and sharing their knowledge and insights with me. To Elfriede Eppinger for her excellent work in insect cell facility. Thanks to Claire Basquin for great help in biophysical experiments. Karina Valer-Saldana, Sabine Pleyer, Ariane Fisher for setting up so many crystallization plates and Ariane also for her help in expression test. Thanks to Drs. Sebastian Falk, Christian Benda, Rajan Prabu, Eva Kowalinski and Ingmar Schäfer for sharing their knowledge and experience with me in structural biology and generous help during my work in the lab. Also thanks to Judith Ebert, Peter Reichelt, Walter Erhardt, Steffen Schuessler, Marc Baumgaertner, Petra Lee and Ulrike Goldschmitt for all their help in the lab!
Thanks to all the people in Conti department for being so nice to me and making this place so great to be!
I would like to thank the International Max Planck Research School PhD program for providing many important training activities.
Thanks to the RNPnet for the financial support.
In the end, I want to thank my family for their support for me to pursue my PhD degree in Germany.