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

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

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

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

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

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

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

RESULTS AND DISCUSSION
Using bioinformatics analyses and proteolysis experiments, we identified regions C. elegans (C. c.) full-length SMG8 (873 residues) and SMG9 (385 residues) as sufficient to form a stable heterodimeric core complex (SMG8c, residues 1–423 and SMG9c, residues 59–375, Fig. 1A) and to yield...
diffracting crystals. After overcoming crystal lattice defects (detailed in Materials and Methods), we solved the structure and refined it at 2.5 Å resolution with R\textsubscript{free} of 26.0% (Table 1).

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

The G domains of SMG8c and SMG9c face each other and interact with part of their convex surfaces (Fig. 1B). In particular, SMG8c helix a2A interacts with SMG9c helices a4 and a3 (patch 1) (in particular Val83\textsubscript{SMG8}, Ile86\textsubscript{SMG8} with Leu258\textsubscript{SMG9}, Leu261\textsubscript{SMG9}) (Fig. 1D). In addition, the stalk domain of SMG8c folds back on the convex surface of SMG9c (patch 2). Here, SMG8c stalk helices a2B and a7 interact with SMG9c helices a7 and a3 (e.g., Ile335\textsubscript{SMG8} and Phe338\textsubscript{SMG8} with Val122\textsubscript{SMG9} and Tyr358\textsubscript{SMG9}) (Fig. 1E). Many of the hydrophobic interface residues observed in the \textit{C. elegans} SMG8c–SMG9c structure are conserved in the human orthologs (Supplemental Figs. 2, 3), suggesting a similar overall structure. To test this prediction, we engineered mutations in human full-length SMG9 (hSMG9) by substituting Met390 (corresponding to \textit{C. elegans} Leu258\textsubscript{SMG9}) and Tyr515 (corresponding to \textit{C. elegans} Tyr358\textsubscript{SMG9}) (Fig. 1D). We transiently coexpressed full-length HA-tagged hSMG8 and Flag-HA-tagged hSMG9 (wild-type, M390R and M390R, Y515R mutants) in HEK293T cells and carried out coimmunoprecipitation assays with Anti-Flag affinity beads, probing with an anti-HA antibody. We found that the interaction of Flag-HA-SMG9 is 63.6 kDa. The mutated residues in human SMG8 and Flag-HA-tagged hSMG9 (wild-type, M390R and M390R, Y515R double mutant) was indeed strongly impaired by the hSMG9 M390R mutant and almost abolished with the hSMG9 M390R, Y515R double mutant (Fig. 1F).

The relative position of the G-like domains in the SMG8c–SMG9c heterodimer is remarkably similar to that observed in active dimeric GTPases of the dynamin family (Supplemental Fig. S1; Daunke and Praefcke 2016), with the two G domains converging at the loops that are known to harbor the nucleotide-binding motifs (G motifs) in canonical GTPases. However, SMG8 lacks the characteristic residues of G motifs. Another difference is that the single-stalk domain in SMG8c–SMG9c has a different position as compared to the conformations observed in dynamin-like proteins (Supplemental Fig. S1; Byrnes et al. 2013). Finally, the SMG8c–SMG9c heterodimer is formed irrespective of nucleotides, while proteins such as Atlastin or GBP1 dimerize in the presence of GTP analogs (Ghosh et al. 2006; Bian et al. 2011; Byrnes and Sondermann 2011).
We tested whether SMG8c–SMG9c can bind guanosine nucleotides. In fluorescence binding assays with mannt-nucleotide derivatives, mant-GDP bound SMG8c–SMG9c and SMG9c with a dissociation constant ($K_d$) of 10 µM and 15 µM, respectively (Fig. 2A). Mant-GTP/βS bound SMG8c–SMG9c with a $K_d$ of 6.5 µM, suggesting a slightly tighter binding in the presence of the guanine γ-phosphate (Fig. 2A). In general, the low-micromolar binding affinities we measured for SMG8c–SMG9c are similar to those reported for GBP1 (Praefcke et al. 1999). We proceeded to obtain our coordinates to progress in the interpretation of cryo-EM structures of human SMG1–SMG8–SMG9 that have been recently resolved at ~20 Å resolution (Fig. 3; Arias-Palomo et al. 2011; Melero et al. 2014; Deniaud et al. 2015). We fitted a homology model of SMG1 with the kinase domain of Atlastin and GBP1 in the G4 motif is not present in SMG9 (Fig. 2B,C). At the corresponding position of Asp218Atlastin, SMG9 features a conserved lysine residue (Lys241SMG9) that stacks with its aliphatic portion on top of the guanine base. With the caveat that motif G5 is largely disordered, none of the interactions in the current structure engage guanine-specific moieties.

We used our coordinates to progress in the interpretation of cryo-EM structures of human SMG1–SMG8–SMG9 that have been recently resolved at ~20 Å resolution (Fig. 3; Arias-Palomo et al. 2011; Melero et al. 2014; Deniaud et al. 2015). We fitted a homology model of SMG1 with the kinase domain of Atlastin and GBP1 in the G4 motif is not present in SMG9 (Fig. 2B,C). At the corresponding position of Asp218Atlastin, SMG9 features a conserved lysine residue (Lys241SMG9) that stacks with its aliphatic portion on top of the guanine base. With the caveat that motif G5 is largely disordered, none of the interactions in the current structure engage guanine-specific moieties.

The start of the G domain of SMG9 is at the center of the density, with nine equivalent positions of the so-called arginine “finger” (Arg77Atlastin) (Fig. 2B,C). Consistent with the absence of such arginine (which stimulates the GTPase activity of dynamin-like proteins in cis), we did not detect convincing GTPase hydrolysis in vitro (data not shown). Another significant difference is at the G4 and G5 loops that bind the base of the nucleotide in dynamin-like proteins. The characteristic guanosine specificity determinant of Atlastin and GBP1 in the G4 motif is not present in SMG9 (Fig. 2B,C). At the corresponding position of Asp218Atlastin, SMG9 features a conserved lysine residue (Lys241SMG9) that stacks with its aliphatic portion on top of the guanine base. With the caveat that motif G5 is largely disordered, none of the interactions in the current structure engage guanine-specific moieties.

<table>
<thead>
<tr>
<th>TABLE 1. Data collection and refinement statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CeSMG8-9-Apo</strong></td>
</tr>
<tr>
<td><strong>Data set</strong></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
</tr>
<tr>
<td>Resolution range (Å)$^a$</td>
</tr>
<tr>
<td>Space group</td>
</tr>
<tr>
<td>a, b, c (Å)</td>
</tr>
<tr>
<td>α, β, γ (%)</td>
</tr>
<tr>
<td>Total reflections$^a$</td>
</tr>
<tr>
<td>Unique reflections$^a$</td>
</tr>
<tr>
<td>Multiplicity$^a$</td>
</tr>
<tr>
<td>Completeness (%)$^a$</td>
</tr>
<tr>
<td>Mean l/sigma(l)$^a$</td>
</tr>
<tr>
<td>R-merge$^a$</td>
</tr>
<tr>
<td>CC1/2$^a$</td>
</tr>
<tr>
<td><strong>Reefinement</strong></td>
</tr>
<tr>
<td>R-work</td>
</tr>
<tr>
<td>R-free</td>
</tr>
<tr>
<td>Average B-factor</td>
</tr>
<tr>
<td>Ligands</td>
</tr>
<tr>
<td><strong>Stereochemistry</strong></td>
</tr>
<tr>
<td>RMS (bonds)</td>
</tr>
<tr>
<td>RMS (angles)</td>
</tr>
<tr>
<td>Ramachandran favored (%)</td>
</tr>
<tr>
<td>Ramachandran allowed (%)</td>
</tr>
<tr>
<td>Ramachandran outliers (%)</td>
</tr>
</tbody>
</table>

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

**MATERIALS AND METHODS**

Protein expression and purification

We analyzed the amino acid sequence of SGM8 and SMG9 proteins from different species in an effort to identify orthologs that would be best suited for crystallization. We selected the *C. elegans* (Ce.) proteins since they are 10%–25% smaller and therefore likely more compact than their human counterparts. *C. elegans* SMG8 (873 residues) and SMG9 (385 residues) were subcloned from a *C. elegans* cDNA library with standard PCR protocols in a single MultiBac expression vector (pFL) (Fitzgerald et al. 2006). SMG8 was cloned into the multiple cloning site 1 (MCS1) of the pFL vector using Xma1 and Nhe1, while SMG9 was cloned into the multiple cloning site 2 (MCS2) using BamHI and Sall. Coexpression was crucial to obtain the heterodimer: Although SMG9 could be expressed soluble when in isolation, SMG8 was insoluble in 25 mM Tris pH 8.0 with 300 mM NaCl, and purified in a soluble form, SMG8 was insoluble when in isolation (data not shown).

Rounds of limited proteolysis and optimization of the expression constructs narrowed down the SMG8c–SMG9c core complex (C.e. SMG8 1–423 and SMG9 59–375). SMG8c–SMG9c were coexpressed in baculo-virus-infected Hi-Five insect cells (Invitrogen) at 26°C for 70 h. Cells were lysed in 25 mM Tris pH 8.0 with 300 mM NaCl, and collected the protein. Protein expression and purification
and 20 mM imidazole was supplemented to the supernatant before loading onto the nickel column. The complex was purified by nickel-based affinity chromatography via a C-terminal hexahistidine tag on C.e. SMG8, and subsequent ion exchange (Heparin HiTrap) and gel-filtration chromatography (Superdex200, equilibrated with 25 mM Tris, 300 mM NaCl, pH 8.0). SelenoMethionine (SeMet) substituted proteins were expressed in insect cells with similar protocols that we reported previously (Halbach et al. 2013). The purification procedure of the SeMet-substituted complex was the same as for the native protein, except that all buffers were degassed and 4 mM β-mercaptoethanol and 2 mM DTT were added before and after elution from the Ni²⁺-NTA resin, respectively. Mass spectrometry analysis showed the presence of ~60% SeMet incorporation in the purified complex.

Crystallization and structure determination

C.e. SMG8c–CeSMG9c crystallized by vapor diffusion in several PEG conditions at pH 8.0 and 10°C. These initial crystals diffrazed to ~3.0 Å resolution and could be processed in a hexagonal space-group, but analysis of the cumulative intensity distribution showed the presence of merohedral twinning with a twin fraction close to 0.5. Additive screening allowed us to identify yttrium chloride as an effective chemical compound to overcome the twinning problem. The best un-twinned crystals were grown by hanging-drop vapor diffusion in drops formed by equal volumes (1.5 µL) of protein (6.8 mg/mL in gel filtration buffer supplemented with 0.11 mM YCl₃) and crystallization buffer (10% PEG3350, 0.1M Tris pH 8.5). SeMet crystals were obtained using the same conditions, but adding tris(2-carboxyethyl)phosphine (TCEP, to limit SeMet oxidation) and covering the reservoir buffer with paraffin oil (to slow drop evaporation and increase crystal size). All crystals were cryo-protected with the crystallization buffer supplemented with 25% ethylene glycol prior to cryo-cooling and data collection.

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

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

Nucleotide-binding experiments

The affinities for GDP were determined by fluorescence measurements on an Infinite M1000 Pro (Tecan). Experiments were carried out at 21°C in a buffer containing 25 mM Tris pH 8.5, 150 mM NaCl, and 5 mM MgCl₂. Increasing protein concentrations were incubated with 1.67 µM of methylanthraniloyl (mant) labeled GDP for 30 min at room temperature. The experiments were carried out with the fragments crystallized, since the full-length proteins were prone to degradation of the low complexity sequences. Fluorescence of mant-GDP was excited at 355 nm and emission spectra were then monitored from 400 to 500 nm, with emission maxima detected at 448 nm. The intrinsic protein fluorescence as well as the mant-nucleotide background was subtracted from the curves. Curve fittings were done with Origin with a one-to-one binding model and are consistent with the presence of one molecule of nucleotide per heterodimer. Curves were done in triplicate. Similar approaches were used to determine the binding affinities for GTPyS.

Coimmunoprecipitation assays

Both the SMG8 and SMG9 were cloned in a vector containing the EF-1 α promoter and with an N-terminal Flag tag and N-terminal...
HA tag using EcoRI and NotI restriction sites. HEK293T cells were cultured in Dulbecco’s modified Eagle medium containing 10% fetal bovine serum (Gibco), 100 U/mL penicillin, and 0.1 mg/mL streptomycin (Gibco) at 32°C/5% CO2. Plasmids were transfected with polyethyleneimine (Polysciences Inc., 1 mg/mL) for protein interaction studies. HEK293T cells were collected from confluent six-well plates after 72 h of transient transfection. Cells were lysed in 0.5 mL of lysis buffer containing 50 mM Tris, pH 7.4, 150 mM NaCl, supplemented with protease inhibitor cocktail (Roche) and DNase I. The lysate was centrifuged at 16,000g for 30 min at 4°C. Twelve microliters of Anti-Flag M2 sepharose beads (Sigma) were added to supernatant for 1 h at 4°C. Beads were washed four times with 1 mL of buffer containing 50 mM Tris, pH 7.4, 300 mM NaCl, and proteins were eluted with 25 µL of lysis buffer supplemented with 100 µg/mL flag peptide (Sigma-Aldrich, F3290). Eluted proteins were run on 12% polyacrylamide gels and transferred onto polyvinylidene difluoride membrane (0.45 µm pore size) (Millipore Immobilon-P) for Western blotting. Anti-HA rabbit polyclonal antibody (Covance, MMS-101 R) antibody and horseradish peroxidase-coupled goat anti-mouse (Millipore, AQ502A) secondary antibody were used in combination with ECL prime Western blotting detection reagent (GE healthcare) for detection of Flag-HA and HA-tagged proteins via Western blotting.

DATA DEPOSITION
The coordinates have been deposited in the Protein Data Bank with accession codes 5NK5 (SMG8-SMG9) and 5NKK (SMG8-SMG9-GDP).

SUPPLEMENTAL MATERIAL
Supplemental material is available for this article.

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

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

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

REFERENCES


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

Liang Li, Mahesh Lingaraju, Claire Basquin, et al.

RNA 2017 23: 1028-1034 originally published online April 7, 2017
Access the most recent version at doi:10.1261/rna.061200.117

Supplemental Material
http://rnajournal.cshlp.org/content/suppl/2017/04/07/rna.061200.117.DC1

References
This article cites 27 articles, 8 of which can be accessed free at:
http://rnajournal.cshlp.org/content/23/7/1028.full.html#ref-list-1

Open Access
Freely available online through the RNA Open Access option.

Creative Commons License
This article, published in RNA, is available under a Creative Commons License (Attribution 4.0 International), as described at http://creativecommons.org/licenses/by/4.0/.

Email Alerting Service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.

To subscribe to RNA go to:
http://rnajournal.cshlp.org/subscriptions

© 2017 Li et al.; Published by Cold Spring Harbor Laboratory Press for the RNA Society