Nucleic Acids Research, 2024, **52**, 6036–6048
https://doi.org/10.1093/nar/gkae323
Advance access publication date: 6 May 2024
Structural Biology

UPF1 helicase orchestrates mutually exclusive interactions with the SMG6 endonuclease and UPF2

Lukas M. Langer*, Katharina Kurscheidt, Jérôme Basquin, Fabien Bonneau, Iuliia Iermak, Claire Basquin and Elena Conti *

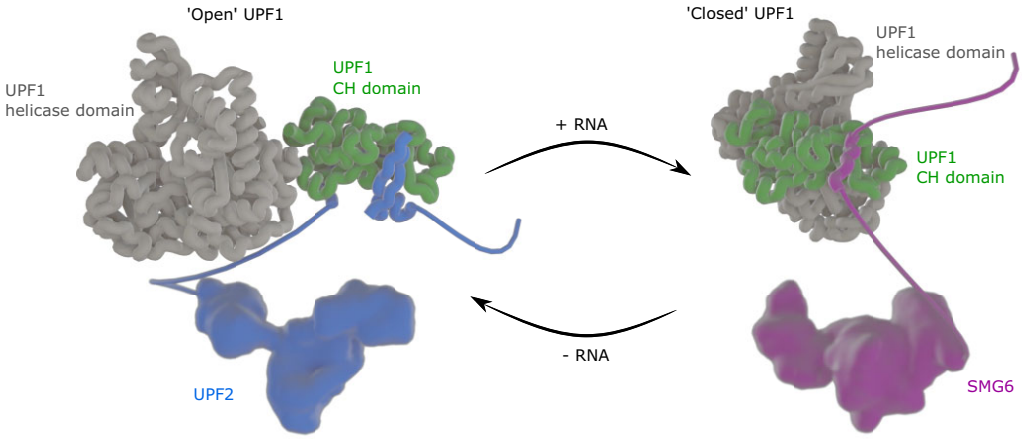
Department of Structural Cell Biology, Max Planck Institute of Biochemistry, Martinsried/Munich D-82152, Germany

*To whom correspondence should be addressed. Tel: +49 89 8578 3606; Fax: +49 89 8578 3605; Email: conti@biochem.mpg.de
Correspondence may also be addressed to Lukas M. Langer. Email: lulanger@biochem.mpg.de

Abstract

Nonsense-mediated mRNA decay (NMD) is a conserved co-translational mRNA surveillance and turnover pathway across eukaryotes. NMD has a central role in degrading defective mRNAs and also regulates the stability of a significant portion of the transcriptome. The pathway is organized around UPF1, an RNA helicase that can interact with several NMD-specific factors. In human cells, degradation of the targeted mRNAs begins with a cleavage event that requires the recruitment of the SMG6 endonuclease to UPF1. Previous studies have identified functional links between SMG6 and UPF1, but the underlying molecular mechanisms have remained elusive. Here, we used mass spectrometry, structural biology and biochemical approaches to identify and characterize a conserved short linear motif in SMG6 that interacts with the cysteine/histidine-rich (CH) domain of UPF1. Unexpectedly, we found that the UPF1–SMG6 interaction is precluded when the UPF1 CH domain is engaged with another NMD factor, UPF2. Based on cryo-EM data, we propose that the formation of distinct SMG6-containing and UPF2-containing NMD complexes may be dictated by different conformational states connected to the RNA-binding status of UPF1. Our findings rationalize a key event in metazoan NMD and advance our understanding of mechanisms regulating activity and guiding substrate recognition by the SMG6 endonuclease.

Graphical abstract



Introduction

Co-translational mRNA quality control pathways ensure the timely elimination of problematic transcripts and are a widely used cellular tool to control eukaryotic gene expression (1,2). Nonsense-mediated mRNA decay (NMD) is one of the major mRNA surveillance mechanisms, originally identified in yeast and human cells for recognizing and degrading mRNAs containing premature termination codons (3,4). Besides performing quality control functions, NMD also acts on physiological mRNAs, effectively regulating their protein expression levels (5). In addition, the pathway leads to the degradation of the aborted nascent polypeptide chains translated from these transcripts (6–8). Activation of the NMD pathway requires translation, *cis*-acting elements present within the defective mRNA, and *trans*-acting factors (9).

In human cells, the major *cis*-acting element is the exon-exon junction complex (EJC), an ensemble of proteins that are assembled on the mRNA during splicing. The EJC forms ∼20–24 nucleotides upstream of spliced junctions, and will become part of the messenger ribonucleoprotein (mRNP) that will be exported out of the nucleus (10). In transiting from the nucleus to the cytoplasm, an outer protein layer associates around the

Received: February 10, 2024. **Revised:** April 6, 2024. **Editorial Decision:** April 8, 2024. **Accepted:** April 25, 2024

EJC core. In particular, the core recruits NMD factors by binding directly to UPF3, which in turn binds UPF2 (11). The cytoplasmic EJCs are normally disassembled as the mRNPs undergo translation (12,13). However, if translation comes to a halt at a premature termination codon (PTC), preventing the disassembly of EJCs downstream of the PTC, NMD factors sense the context of an aberrant mRNP and respond to it.

The pivotal factor in NMD is arguably the RNA helicase UPF1. The helicase core of UPF1 contains the catalytic domains of the RNA-dependent ATPase and is connected to regulatory domains, including the so-called cysteine-histidine-rich domain (CH domain) (14,15). The UPF1 CH domain ($UPF1_{CH}$) binds UPF2: in the case of the human proteins, the interaction has been visualized structurally and shown biochemically to decrease RNA-binding capabilities (11,14,16). The CH-helicase region of human UPF1 is flanked by N- and C-terminal unstructured regions that have also been reported to limit the RNA-binding properties of UPF1, although the mechanisms are once again unclear (17). The N- and C-terminal regions of UPF1 are recognized and phosphorylated by the SMG1-SMG8-SMG9 kinase complex (18–20). When phosphorylated, the C-terminal region interacts with the 14–3–3-like SMG5-SMG7 heterodimer (21,22), which cross-talks with the CCR4–NOT deadenylase complex to shorten the polyadenosine tail (21,23). Although the phosphorylated N-terminus has been implicated in recruiting SMG6 to the mRNP, other parts of the UPF1 N-terminal region have been shown to interact with SMG6 in a phospho-independent manner (24,25). SMG6 contributes a crucial enzymatic activity in the NMD pathway. Through its PIN domain, SMG6 cleaves the transcript in proximity of the PTC, a hallmark of metazoan NMD (26,27). The resulting mRNA fragments are then ultimately degraded via XRN1 from the 5′ to 3′ direction and by the cytoplasmic exosome in the 3′ to 5′ direction (28).

While we have a general overview of the biochemical reactions executed by UPF1 and SMG6 during NMD, we do not know the precise stepwise progression of molecular events occurring between initial PTC recognition to the final decay of the transcript. For example, SMG6 contains short linear motifs (SLIMs) that engage the same surface of the EJC where UPF3 binds, suggesting that the progression of the NMD process unfolds with the coordination of these mutually exclusive interactions (29,30). Yet, when is SMG6 recruited to UPF1 and how do these two factors organize within the larger NMD process? To address the orchestration of enzymatic components in human NMD, we combined biochemistry *in vitro* and in cells with structural biology and mass spectrometry to dissect interactions between the NMD components UPF1 and SMG6.

Materials and methods

Protein expression and purification

Bacterial cells

6xHis-MBP-3C-tagged SMG6 or UPF2 fragments were expressed in *Escherichia coli* pRARE cells at 18°C overnight. Cells were harvested at 6000 rpm for 10 min at 4°C and lysed by sonication in lysis buffer 1 (100 mM Tris–Cl, 500 mM NaCl, 2 mM DTT pH 7.5) supplemented with DNases, lysozyme and cOmplete protease inhibitor cocktail (Roche Diagnostics). Cell debris was pelleted at 25 000 rpm for 30 min at 4°C and the cleared lysate was combined with pre-equilibrated amylose resin (New England Biolabs) for gravity-flow affinity purification. The resin was washed with several column volumes (CV) of lysis buffer 1 and wash buffer 1 (20 mM HEPES-OH, 120 mM NaCl, 1 mM DTT pH 7.4). Bound proteins were eluted using wash buffer supplemented with 25 mM maltose, and proteins were flash frozen using liquid nitrogen following concentration.

TwinStrep-3C-UPF1 CH domain constructs (residues 110–282) were expressed as described above and lysed in lysis buffer 2 (100 mM Tris–Cl, 500 mM NaCl, 2 μM $ZnCl_2$, 2 mM DTT pH 7.5) supplemented with DNases, lysozyme and cOmplete protease inhibitor cocktail (Roche Diagnostics). The cleared lysate was passed over a 5 ml StrepTrap HP column (Cytiva) and the column was washed using lysis buffer 2 and wash buffer 1. The bound protein was eluted using wash buffer 1 supplemented with 2.5 mM desthiobiotin, concentrated and polished using a Superdex 75 HiLoad column (Sigma-Aldrich) equilibrated with wash buffer 2.

UPF1 CH domain used for crystallization (6xHis-3C-UPF1 115–287) was expressed and lysed as described for MBP-tagged protein constructs in lysis buffer 3 (100 mM Tris–Cl, 500 mM NaCl, 2 μM $ZnCl_2$, 5 mM β-mercaptoethanol and 10 mM imidazole pH 7.5) supplemented with DNases, lysozyme and cOmplete protease inhibitor cocktail (Roche Diagnostics). The cleared lysate was passed over a 5 ml HisTrap HP and the column was washed with several CVs of lysis buffer 3 and wash buffer 3 (20 mM HEPES-OH, 120 mM NaCl, 2 μM $ZnCl_2$, 5 mM β-mercaptoethanol and 10 mM imidazole pH 7.4), before elution of bound proteins by increasing the imidazole concentration to 300 mM. The eluate was dialysed overnight into 20 mM HEPES-OH, 120 mM NaCl, 5 mM β-mercaptoethanol pH 7.4 in the presence of His-tagged 3C protease and cleaved tags as well as uncleaved proteins were removed by passing over a HisTrap HP column. The flow-through was concentrated and run over a Superdex 75 HiLoad column (Sigma-Aldrich) using 20 mM HEPES-OH, 120 mM NaCl, 2 mM dithiothreitol pH 7.4 before being concentrated again and flash frozen for storage until further usage.

For the production of full-length unphosphorylated human UPF1 protein, a UPF1 construct fused to a TEV-cleavable C-terminal 6xHis tag (hsUPF1) was expressed in *E. coli* BL21 Star (DE3) pRARE cells at 18°C overnight. Bacterial cells were harvested at 6000 rpm for 10 min at 4°C and subsequently lysed by sonication in wash buffer 4 (20 mM Tris–HCl pH 7.5, 1 M NaCl, 2 mM $MgCl_2$, 20 mM imidazole, 10 % (v/v) glycerol, 2 mM DTT) supplemented with 1 mM AEBSF (PanReac AppliChem), 0.1 % (v/v) NP-40, Benzonase (Merck) and EDTA-free cOmplete protease inhibitor cocktail (Roche Diagnostics). The lysate was cleared by centrifugation for 30 min at 25 000 rpm and filtration and loaded onto a HisTrap FF column (Cytiva) pre-equilibrated with wash buffer 4. The column was washed with 10 CV wash buffer 4 and wash buffer 5 (20 mM Tris–HCl pH 7.5, 500 mM NaCl, 2 mM $MgCl_2$, 20 mM imidazole, 10 % (v/v) glycerol, 2 mM DTT), respectively. Bound proteins were eluted with wash buffer 5 supplemented with 500 mM imidazole. The eluate was combined with His-tagged TEV protease and dialyzed overnight against 20 mM Tris–HCl pH 7.5, 30 mM NaCl, 70 mM KCl, 2 mM $MgCl_2$, 10 % (v/v) glycerol and 2 mM DTT. To separate cleaved UPF1 without affinity tag from cleaved tag, the His-tagged protease and uncleaved UPF1, the dialysed sample was again passed over a HisTrap FF column (Cytiva) and the

flow-through containing cleaved UPF1 was subjected to ion exchange chromatography over a HiTrap Heparin HP column (Cytiva). The column was loaded with the flow-through and subsequently washed with 10 CV Hep A buffer (20 mM Tris–HCl pH 7.5, 70 mM KCl, 2 mM $MgCl_2$, 10 % (v/v) glycerol and 2 mM DTT). Bound proteins were eluted with a linear salt gradient from 70 mM to 1 M KCl. The peak fractions containing full-length UPF1 were pooled and concentrated. For a final polishing step, the concentrated protein sample was run over a Superdex 200 increase 10/300 GL column (Cytiva) with gel filtration buffer (20 mM HEPES pH 7.5, 75 mM KCl, 1 mM $MgCl_2$, 5 μM $ZnCl_2$, 10 % (v/v) glycerol and 2 mM DTT). The peak fractions corresponding to full-length UPF1 were pooled and again concentrated before they were flash-frozen using liquid nitrogen and stored at –80°C until further usage.

Mammalian cells

Proteins were expressed with a N-terminal TwinStrep-3C-tag in either transiently transfected HEK293T cells or stable pools of the same cell line as described before (24). Pelleted cells were washed using ice-cold 1× D-PBS and lysed with a Dounce homogenizer in 2× D-PBS, 5 mM $MgCl_2$ and 1 mM DTT supplemented with DNase I, Benzonase and EDTA-free cOmplete protease inhibitor cocktail (Roche Diagnostics). Cell debris was removed by pelleting (25 000 rpm, 30 min, 4°C) and the cleared lysate was loaded on an affinity column.

For SMG6 constructs, the cleared lysate was loaded onto a StrepTrap XT column (Cytiva) pre-equilibrated with lysis buffer. The column was washed with at least 10 CV of wash buffer (2× PBS, 2.5 mM $MgCl_2$ and 1 mM DTT). Bound proteins were eluted with wash buffer supplemented with 50 mM biotin. The eluate containing full-length SMG6 was concentrated and run over a Superdex 200 increase 10/300 GL column (Cytiva) with wash buffer. The peak fractions containing full-length SMG6 were pooled, concentrated, flash-frozen using liquid nitrogen and stored at –80°C until further usage.

For UPF1 used in cryo-EM studies, the affinity column was washed with 10 CVs of 1× D-PBS, 5 mM $MgCl_2$, 1 μM $ZnCl_2$, 10 % (v/v) glycerol and 1 mM DTT followed by 10 CVs of 20 mM HEPES-OH, 50 mM KCl 1 mM $MgCl_2$, 1 μM $ZnCl_2$, 5 % (v/v) glycerol and 1 mM DTT (Hep A buffer) before being eluted onto a HiTrap Heparin HP column (GE Healthcare) using the Hep A buffer supplemented with 2.5 mM desthiobiotin. After washing with Hep A buffer, bound protein was eluted by a gradient with increasing salt concentration from 50 to 500 mM KCl (Aekta prime FPLC system, GE Healthcare). The peak fractions were analyzed by SDS-PAGE and the fractions carrying full-length UPF1 were pooled, concentrated, flash frozen and stored at -80 °C until further usage.

In vitro pull-down experiments and competition assays

For pull-down experiments with proteins purified from *E. coli*, 1.5 μM of Twinstrep-tagged bait protein was mixed with a 10-fold molar excess of prey protein in 40 μl of a 50 mM Potassium phosphate buffer pH 8 supplemented with 50 mM NaCl, 5 mM $MgCl_2$, 0.1 % (v/v) NP-40 substitute, 5 % (v/v) glycerol and 1 mM DTT. After incubation for at least 30 min at 4°C, input samples were set aside and 15 μl slurry of pre-equilibrated MagStrep 'type3' XT beads (IBA Lifesciences) were added, followed by incubation for 1 h at 4°C. The beads were washed 3 times with 500 μl and 2 times with 300 μl pull-down buffer, and bound protein were eluted by boiling for 2 min in 20 μl sample buffer containing SDS.

For competition assays, complexes were pre-formed by mixing 1.5 μM of Twinstrep-tagged protein with 5 μM of MBP-tagged protein for 15 min at 4°C in 40 μl Phosphate buffer pH 7.3 supplemented with 75 mM NaCl, 5 mM $MgCl_2$, 0.1 % (v/v) NP-40 substitute, 5 % (v/v) glycerol and 1 mM DTT. To challenge the pre-formed complexes, the second MBP-tagged protein was added at concentrations of 0.5, 1.5, 5 and 10 μM for 15 min at 4°C. The complete reaction mixes were then combined with beads for 40 min at 4°C as described for pull-downs above. After taking input samples, the reactions were processed as described above.

Eluates from both pull-down and competition assays were analyzed by running 15 % SDS-PAGE followed by staining with Der Blaue Jonas (GRP).

For *in vitro* pull-down experiments with full-length proteins, 1.0 μM of TwinStrep-tagged SMG6 bait protein (WT or L409E mutant) purified from HEK293T cells was mixed with a 3-fold molar excess of UPF1 purified from *E. coli* as prey protein in 30 μl binding buffer (2× D-PBS supplemented with 5 mM $MgCl_2$, 0.1 % (v/v) NP-40, 5 % (v/v) glycerol and 1 mM DTT). Protein mixtures were incubated for at least 30 min at 4°C under continuous shaking to allow for complex formation, before input samples were taken. Following the addition of 10 μl slurry of pre-equilibrated MagStrep 'type3' XT beads (IBA Lifesciences), the reaction mixes were incubated for 15 min at 4°C. The beads were washed three times with 200 μl binding buffer and once with 200 μl high-salt wash buffer (2× D-PBS supplemented with additional 500 mM NaCl, 5 mM $MgCl_2$, 0.2 % (v/v) NP-40, 5 % (v/v) glycerol and 1 mM DTT). For elution, the washed beads were re-suspended in 15 μl elution buffer (10× Buffer BXT (IBA Lifesciences) diluted to 1× concentration with binding buffer) and incubated for 30 min at 4°C.

For *in vitro* pull-down experiments using purified full-length proteins and UPF1 antibody N, magnetic Dynabeads coupled to ProteinG (Invitrogen) were pre-bound to the UPF1 antibody. Full-length wildtype UPF1 (purified from *E. coli*) and TwS-SMG6 were mixed with a 2-fold molar excess of SMG6 in binding buffer (see above for full-length proteins) and incubated for 30 min at 4°C. Next, the protein samples were combined with the antibody-bound beads and the mixture was kept for 30 min at 4°C on a rotating wheel. After removal of the flow-through, beads were washed multiple times with binding buffer followed by high-salt wash buffer (see above). Finally, the beads were boiled for 2 min in 20 μl SDS-containing sample buffer.

All input samples and eluates were analyzed by SDS-PAGE followed by staining with Der Blaue Jonas (GRP).

Isothermal calorimetry (ITC)

ITC was carried out by using a ITC200 Microcal calorimeter (Microcal; GE Healthcare). All samples were dialyzed in a buffer containing (20 mM HEPES-OH, 100 mM NaCl, 0.5 mM tris(2-carboxyethyl)phosphine pH 7.4). The TwS-3C-UPF1 110–282 sample was concentrated to around 20 μM and MBP-SMG6 and MBP-UPF2 samples were concentrated to around 200 μM. Titrations were carried out at 20°C with 25–30 injections of 1.5 μl of the MBP-SMG6/UPF2 solution into the UPF1 solution. As control for all ITC measurements, the injectant was titrated into buffer. The measurements were

made in duplicate. All data were processed and curves fitted using the Origin software (Microcal).

Antibody-mediated purification of UPF1 mRNPs

K562 cells were diluted to a density of 0.5 × 10^6 cells per ml and grown for approximately 12 h in RPMI medium (Gibco, Thermo Fisher Scientific) supplemented with 10 % (v/v) FBS at 37°C. After trypsin treatment (Gibco, Thermo Fisher Scientific), cells were harvested by pelleting and washing with ice-cold 1× D-PBS. To lyse the cells, they were resuspended in 50 mM Potassium phosphate pH 7.3 supplemented with 10 mM $MgCl_2$ and 0.1 % (v/v) NP-40 and incubated for 20 min on ice. The lysate was cleared by spinning at 10 000 g (10 min, 4°C) and its density determined by measuring absorption at 260 nm. Normalized amounts of lysate were combined with magnetic Dynabeads coupled to ProteinG (Invitrogen) pre-bound to the desired UPF1 antibody (antibody N: Abcam, ab109363 and antibody C: Abcam, ab133564). Lysate and beads were incubated for 1 hour at 4°C to allow for capture of UPF1-containing particles. Next the beads were washed 4 times with 25 CVs 50 mM Potassium phosphate pH 7.3 supplemented with 10 mM $MgCl_2$ and 0.1 % (v/v) NP-40, 1 time with 20 CVs buffer without detergents and directly submitted to mass spectrometric analysis. A replicate was eluted using SDS-containing sample buffer and analyzed by SDS-PAGE followed by Coomassie-staining.

Immunoblotting

Protein samples were separated using SDS-PAGE followed by stain-free visualization (Bio-Rad). The proteins were then blotted onto a PVDF membrane using a wet transfer setup. The membrane was imaged to visualize transferred proteins using the same stain-free imaging technique as for SDS-PAGE. The membrane was incubated with blocking solution (5 % (w/v) milk powder dissolved in 1× PBS, Tween-20 (PBS-T)) to reduce unspecific binding. Next, the primary antibody was added, diluted as recommended in fresh blocking solution (1:30 000 for antibody N, 1:10 000 for antibody C). After incubation, the membrane was washed with PBS-T before being combined with an HRP-conjugated secondary antibody solution (1:10 000 in blocking solution). Subsequent to further incubation, the membrane was washed extensively with PBS-T and once with PBS, before adding development solution (Cytiva Amersham) and signal detection in increments using a LAS4000 Western blot imager.

AlphaFold-Multimer protein structure predictions

Protein structure prediction was performed using AlphaFold version 2.2.0 and full databases with modelling set to 'multimer' (31,32). Each run resulted in 25 predicted structures (based on 5 models) which were further analysed using UCSF ChimeraX version 1.5 (33).

Protein crystallization and structure solution

$UPF1_{CH}$ (residues 115–287) and a SMG6 peptide comprising residues 404–418 (in-house synthesis) were mixed with a 2x molar excess of SMG6 peptide and a $UPF1_{CH}$ concentration of 15 mg/ml. Crystals were grown at 4°C by hanging-drop vapor diffusion from drops formed by equal volumes of protein and crystallization solution (28 % (w/v) PEG 3500, 0.2 M lithium sulfate and 0.1 M Tris pH 7.8). Prior to flash freezing in liquid nitrogen, the crystals were briefly soaked in mother liquor containing 28 % (v/v) ethylene glycol.

Table 1. Data collection and refinement statistics of human UPF1 CH domain–SMG6 404–418 complex

	UPF1 CH domain/SMG6 404–418
Wavelength (Å)	1.00
Resolution range (Å)	43.44–2.597 (2.63–2.6)
Space group	P 1 2_1 1
Unit cell	
a,b,c (Å)	87.165, 90.677, 92.291
α,β,γ (°)	90, 113.36, 90
Total reflections	148 663 (4624)
Unique reflections	75 047 (2523)
Multiplicity	2.0 (1.8)
Completeness (%)	93.69 (83.21)
Mean I/sigma(I)	6.91 (1.67)
Wilson *B*-factor	54.03
***R*-merge**	0.0691 (0.2972)
***R*-meas**	0.09368 (0.4016)
***R*-pim**	0.06278 (0.2679)
CC1/2	0.99 (0.84)
CC*	0.997 (0.955)
Reflections used in refinement	39 991 (1361)
Reflections used for *R*-free	2001 (68)
***R*-work**	0.2812 (0.3257)
***R*-free**	0.2969 (0.3397)
Number of non-hydrogen atoms	7265
Macromolecules	7248
Ligands	17
Solvent	0
Protein residues	936
RMS(bonds) (Å)	0.007
RMS(angles) (°)	1.08
Ramachandran favored (%)	97.64
Ramachandran allowed (%)	2.13
Ramachandran outliers (%)	0.22
Rotamer outliers (%)	1.49
Clashscore	13.98
Average *B*-factor	61.81
Macromolecules	61.80
Ligands	66.11

Statistics for the highest-resolution shell are shown in parentheses.

X-ray data sets were recorded on the 10SA (PX II) beamline at the Paul Scherrer Institute (Villigen, Switzerland) at a wavelength of 1.0 Å using a Dectris Eiger3 16M detector with the crystals maintained at 100 K by a cryocooler. Diffraction data were integrated using XDS and scaled and merged using AIMLESS (34,35); data collection statistics are summarized in Table 1. The structure solution was obtained by molecular replacement using PDB 2IYK as template and 6 copies of UPF1 could be automatically fitted. The model was completed by manual rebuilding in COOT and refined using in Phenix refine (36,37).

Cryo-electron microscopy

Sample preparation

For preparation of cryo-EM grids, ∼1.75 μM full-length TwS-UPF1 purified from mammalian cells was combined with 0.2 mM AMPPNP and optionally with 1.5 molar excess of U_{15}-RNA over UPF1 in a buffer containing 1× D-PBS, 2 mM $MgCl_2$, 1 μM $ZnCl_2$, 1 mM DTT and 0.04 % (v/v) *n*-octyl-β-D-glucoside and incubated for ∼30 min. 3.5 μl of the samples

were then applied to Quantifoil R2.1 Cu 200 grids and plunge frozen in a liquid ethane/propane mix using a Vitrobot Mark IV (Thermo Fisher Scientific).

Data collection

Both datasets were collected on a 200 kV Glacios Cryo-TEM (ThermoFisher Scientific) equipped with a K2 Summit direct detector (Gatan) operated in counting mode. The data sets were recorded at 1.885 Å/pix with a nominal magnification of 22 000×. The defocus ranged from −1.5 μm to −3 μm and the total dose of ∼60 $e^-/Å^2$ was equally distributed over 40 frames.

Data processing

The frames were motion-corrected in RELION 3.1 (38,39) using MotionCor2 (40), CTFs were estimated using the RELION wrapper for CTFFIND4.1 (41) and particles were picked using Gautomatch (https://www.mrc-lmb.cam.ac.uk/kzhang/Gautomatch/). After particle extraction, the stacks were transferred to CryoSPARC v3.3.2 (42) for further processing. Following 2D classification, an ab-initio model was generated and the particles were further sorted into multiple classes using heterogeneous refinement. The final set of particles was refined using non-uniform refinement (Figure 4 Supp. 1).

Mass spectrometry proteomics

Sample preparation—in-gel samples

Gel pieces were washed two times with 150 μl of destaining buffer (25 mM ammonium bicarbonate, 50 % ethanol) and dehydrated by addition of 150 μl of 100 % ethanol. The ethanol was removed and the gel pieces were dried by vacuum centrifugation. Then, 50 μl of ice-cold digestion buffer (25 mM Tris–HCl, 10 % acetonitrile, 10 ng/ul of trypsin) was added. After incubation for 20 min on ice, 50 μl of ammonium bicarbonate buffer (25 mM) was added and the gel pieces were incubated at 37°C overnight. Peptides in the supernatant were collected and more peptides were extracted from the gel pieces by repeated incubation of the gel pieces at 25°C with 100 μl of extraction buffer (3 % TFA, 30 % acetonitrile) and collection of the supernatants. Finally, the gel pieces were dehydrated by incubation at 25°C in 100 μl of 100 % acetonitrile and the supernatant was unified with the supernatants from previous extraction steps. Acetonitrile was removed by vacuum-centrifugation and 70 μl of 2 M Tris–HCl as well as 10 mM tris(2-carboxyethyl) phosphine (TCEP) and 40 mM 2-cloroacetamide (CAA) was added. After incubation for 30 min at 37°C, peptides were acidified to 1 % TFA and desalted using SDB-XC Stage Tips.

LC–MS/MS data acquisition—in-gel samples

Peptides were dissolved in buffer A (0.1 % formic acid) and analyzed by LC–MS/MS. We have used two different LC–MS set-ups for the data acquisition. The LC–MS/MS system was composed of an Easy-nLC 1200 (Thermo Fisher Scientific) coupled to a QExactive HF mass spectrometer (Thermo Fisher Scientific). Peptides were separated on a 30-cm analytical column (inner diameter: 75 microns; packed in-house with ReproSil-Pur C18-AQ 1.9-micron beads, Dr Maisch GmbH) using a 25 min gradient of 10 % to 30 % buffer B (80 % acetonitrile, 0.1 % FA). The mass spectrometer was operated in a data-dependent mode with survey scans from 300 to 1650 m/z (resolution of 60000 at $m/z = 200$), and up to 10 of the top precursors were selected and fragmented using higher energy collisional dissociation (HCD with a normalized collision energy of value of 28). The MS2 spectra were recorded at a resolution of 30 000 (at $m/z = 200$). AGC target for MS and MS2 scans were set to 3×10^6 and 1×10^5, respectively, within a maximum injection time of 100 and 60 ms for MS and MS2 scans, respectively. Dynamic exclusion was set to 30 ms.

Sample preparation—on-beads samples

After removing the buffer, the beads were incubated with SDC buffer containing 1 % sodium deoxycholate (SDC, Sigma-Aldrich), 40 mM 2-cloroacetamide (CAA, Sigma-Aldrich), 10 mM tris(2-carboxyethyl)phosphine (TCEP; Thermo Fisher Scientific) and 100 mM Tris, pH 8.0 at 37°C. After incubation for 20 min at 37°C, the samples were diluted 1:2 with MS grade water (VWR). Proteins were digested overnight at 37°C by addition of 0.5 μg trypsin (Promega). The supernatant was collected with the help of a magnetic rack and was acidified with trifluoroacetic acid (TFA; Merck) to a final concentration of 1 %. Precipitated SDC was removed by centrifugation and the peptide mixture was desalted via SCX StageTips and after elution, the samples were vacuum dried and dissolved in buffer A (0.1 % formic acid).

LC–MS data acquisition—on-beads samples

Purified and desalted peptides were loaded onto a 30-cm column (inner diameter: 75 microns; packed in-house with ReproSil-Pur C18-AQ 1.9-micron beads, Dr Maisch GmbH) via the autosampler of the Thermo Easy-nLC 1200 (Thermo Fisher Scientific) at 60°C. Using the nanoelectrospray interface, eluting peptides were directly sprayed onto the Exploris 480 mass spectrometer (Thermo Fisher Scientific). Peptides were loaded in buffer A (0.1 % formic acid) and separated through the column at a flow rate of 300 nl/min by increasing percentage of buffer B (80 % acetonitrile, 0.1 % formic acid) in the following steps: increase from 5 % buffer B to 30 % buffer B over 40 min followed by an increase to 65 % buffer B over 5 min then 95 % over the next 5 min. Percentage of buffer B was maintained at 95 % for another 5 min. The mass spectrometer was operated in a data-dependent mode with survey scans from 300 to 1650 m/z (resolution of 60 000 at $m/z = 200$), and up to 15 of the top precursors were selected and fragmented using higher energy collisional dissociation (HCD with a normalized collision energy of value of 28). The MS2 spectra were recorded at a resolution of 15000 (at $m/z = 200$). AGC target for MS and MS2 scans were set to 3×10^6 and 1×10^5, respectively, within a maximum injection time of 25 and 28 ms for MS and MS2 scans, respectively. Dynamic exclusion was set to 30 s.

Data analysis and visualization

Raw data were processed using the MaxQuant computational platform (version 2.2.0.0) (43) with standard settings applied. Shortly, the peak list was searched against the Uniprot database of human (downloaded in 2023) with an allowed precursor mass deviation of 4.5 ppm and an allowed fragment mass deviation of 20 ppm. MaxQuant by default enables individual peptide mass tolerances, which was used in the search. Cysteine carbamidomethylation, methionine oxidation and N-terminal acetylation were set as variable modifications. Proteins were quantified across samples using the

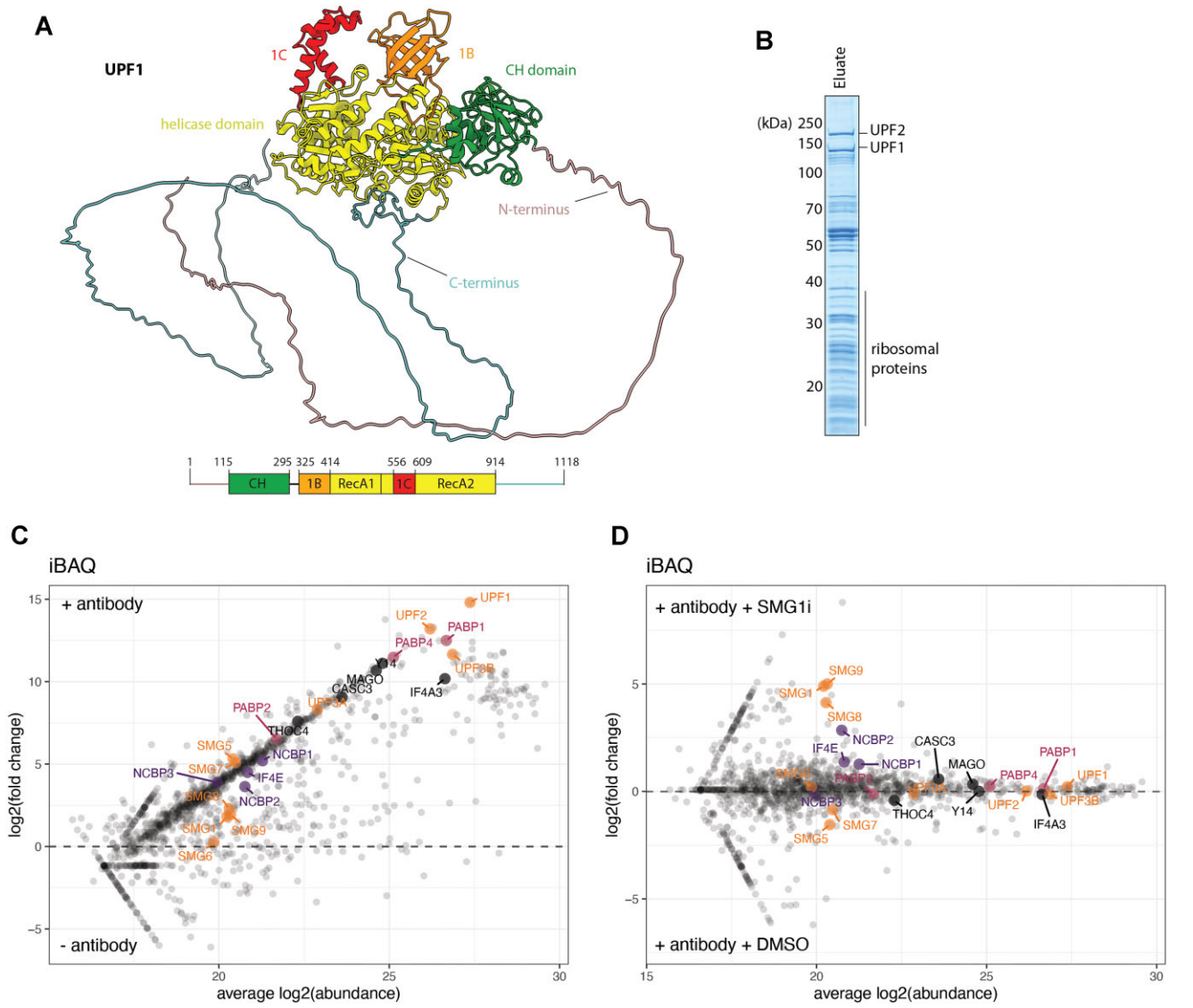


Figure 1. Inventory of UPF1 mRNPs purified from K562 cells. (**A**) AlphaFold2 model of full-length human UPF1 colored by indicated domains. (**B**) Coomassie-stained SDS-PAGE of the protein components pulled-down with UPF1 antibody N. UPF1 and UPF2 proteins identified by peptide mass fingerprinting are labeled (compare Figure 1, Supplementary Figure 2). (**C**) Enrichment analysis comparing pull-downs performed in three replicates using either UPF1 antibody N-bound or free beads. Data points corresponding to selected proteins have been colored and labeled: NMD-specific factors are in orange, EJC components in black, cap-binding proteins in purple and poly(A)-binding proteins are in magenta. (**D**) As (C), but with pull-downs performed using cells grown in the presence of either DMSO or SMG1i dissolved in DMSO.

label-free quantification algorithm in MaxQuant as label-free quantification (LFQ) intensities. Calculation of iBAQ-values (intensity-based absolute quantification of proteins) was enabled. R with the tidyverse collection of packages (44) together with the package proDA (Ahlmann-Eltze 2020 https://doi.org/10.1101/661496) was used to perform statistical analysis and plotting of the mass spectrometry data.

Results

Composition of human UPF1-containing mRNPs under homeostatic conditions

Given the central role of UPF1 in coordinating different NMD factors, we set out to isolate and characterize the composition of endogenous UPF1-containing mRNPs. To purify native UPF1-containing complexes from human cells, we used an antibody-based isolation strategy. We characterized two distinct commercially available monoclonal antibodies raised against UPF1. Immunoblotting against HEK293T cell lysate validated the specificity of both antibodies (Figure 1, Supplementary Figure 1A). Immunoblotting against various regions of recombinantly expressed and purified UPF1 identified the epitopes, with one antibody recognizing the UPF1 N-terminus and the second antibody recognizing the UPF1 C-terminus (referred to as antibody N and C, respectively) (Figure 1A and Figure 1, Supplementary Figure 1B). For purification purposes, each antibody was immobilized to ProteinG-coupled magnetic beads and tested with recombinant full-length UPF1. Both antibodies recognized UPF1 in solution (Figure 1, Supplementary Figure 1C); however, as antibody C was commercially discontinued in the course of this study, we proceeded to characterize UPF1 interactions with antibody N. We then optimized the purification strategy with regards to the cellular

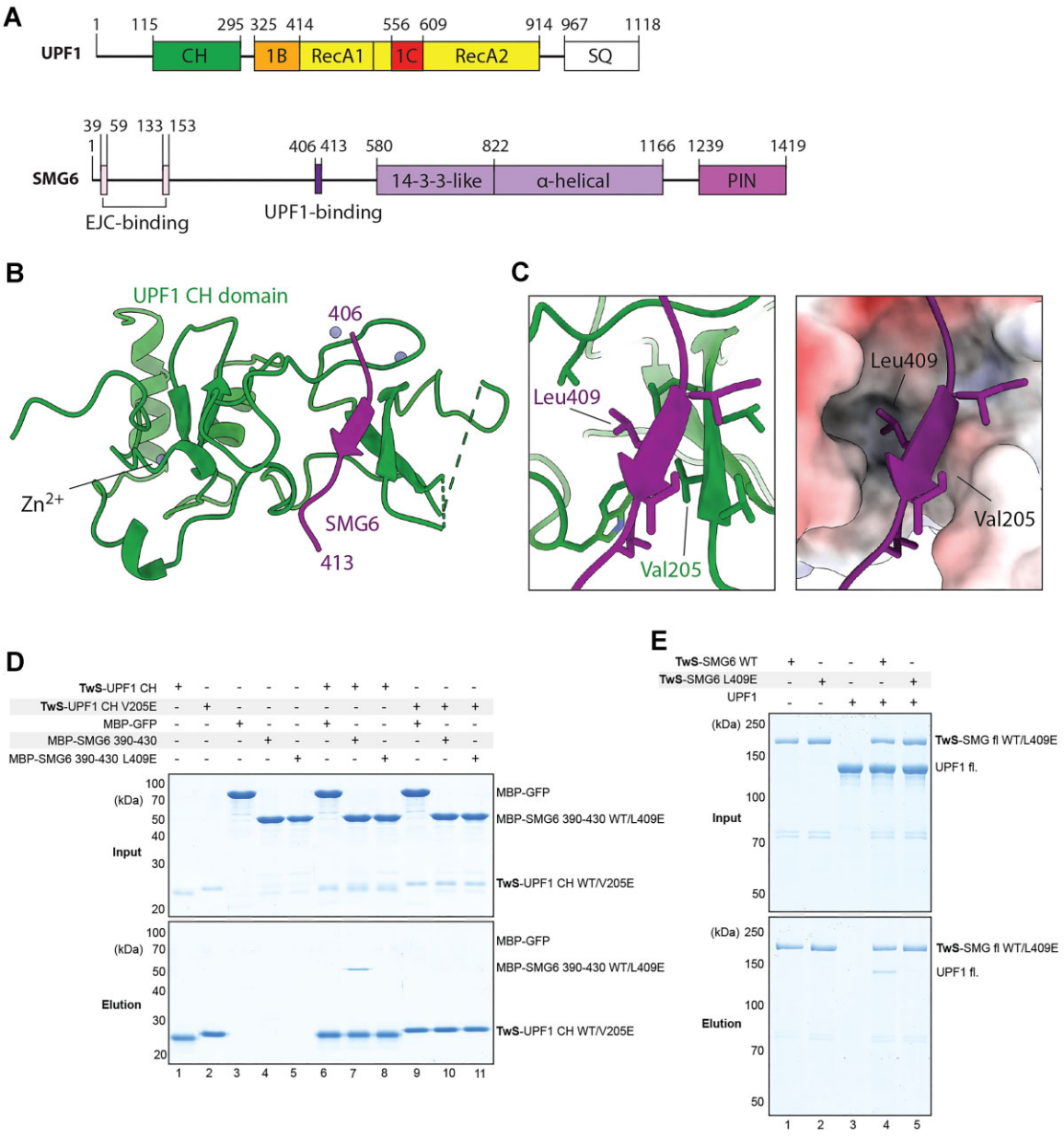


Figure 2. A short linear motif in human SMG6 interacts with the UPF1 CH domain. (**A**) Domain architecture of human UPF1 and SMG6 proteins. (**B**) Cartoon representation of the crystal structure of human UPF1 CH domain (green) bound by a SMG6 peptide (magenta). (**C**) Close-ups highlighting residues involved in the UPF1-SMG6 interaction with the UPF1 CH domain either shown as cartoon or as a surface colored by electrostatic potential. (**D**) Coomassie-stained SDS-PAGE analysis of a pull-down experiment using different truncated purified UPF1 and SMG6 constructs to validate the structural data. (**E**) Coomassie-stained SDS-PAGE analysis of a pull-down experiment using purified human full-length UPF1 and SMG6 proteins.

context, in particular, by establishing a fast pull-down protocol that allowed purifying UPF1-containing complexes from cleared K562 cell lysate without background binding (Figure 1, Supplementary Figure 1D) (45). In the antibody-based purification, SDS-PAGE analysis and peptide mass fingerprinting identified UPF1, UPF2 and ribosomal proteins (Figure 1, Supplementary Figure 2). These data suggested that the protocol we established allows us to purify intact endogenous UPF1 mRNPs, with the RNA-binding status of UPF1 itself within these particles remaining unspecified.

To determine the composition and relative abundance of proteins present in the UPF1 mRNP purification, we obtained quantitative proteomics data by mass spectrometry analysis of triplicate samples pulled-down using antibody N. Pull-downs without antibody served as a control. Quantification of the label-free proteomic data was performed using intensity-based absolute quantification (iBAQ) (46). The triplicates displayed strong consistency, with a marked enrichment of UPF1 and numerous ribosomal proteins as compared to the control (Figure 1, Supplementary Figure 3). We identified different NMD factors, such as UPF1, UPF2, UPF3B as well as the EJC core proteins eIF4A3 and MAGO, amongst the most strongly enriched proteins (Figure 1 Supplementary Figure 3B). In addition, we detected other mRNA-binding proteins specifically enriched in the antibody-mediated pull-downs as compared to the control, for example, YBOX proteins, hnRNPs, PABPs and ribosomal proteins (Figure 1, Supplementary Figure 3B). When comparing relative enrichment of pro-

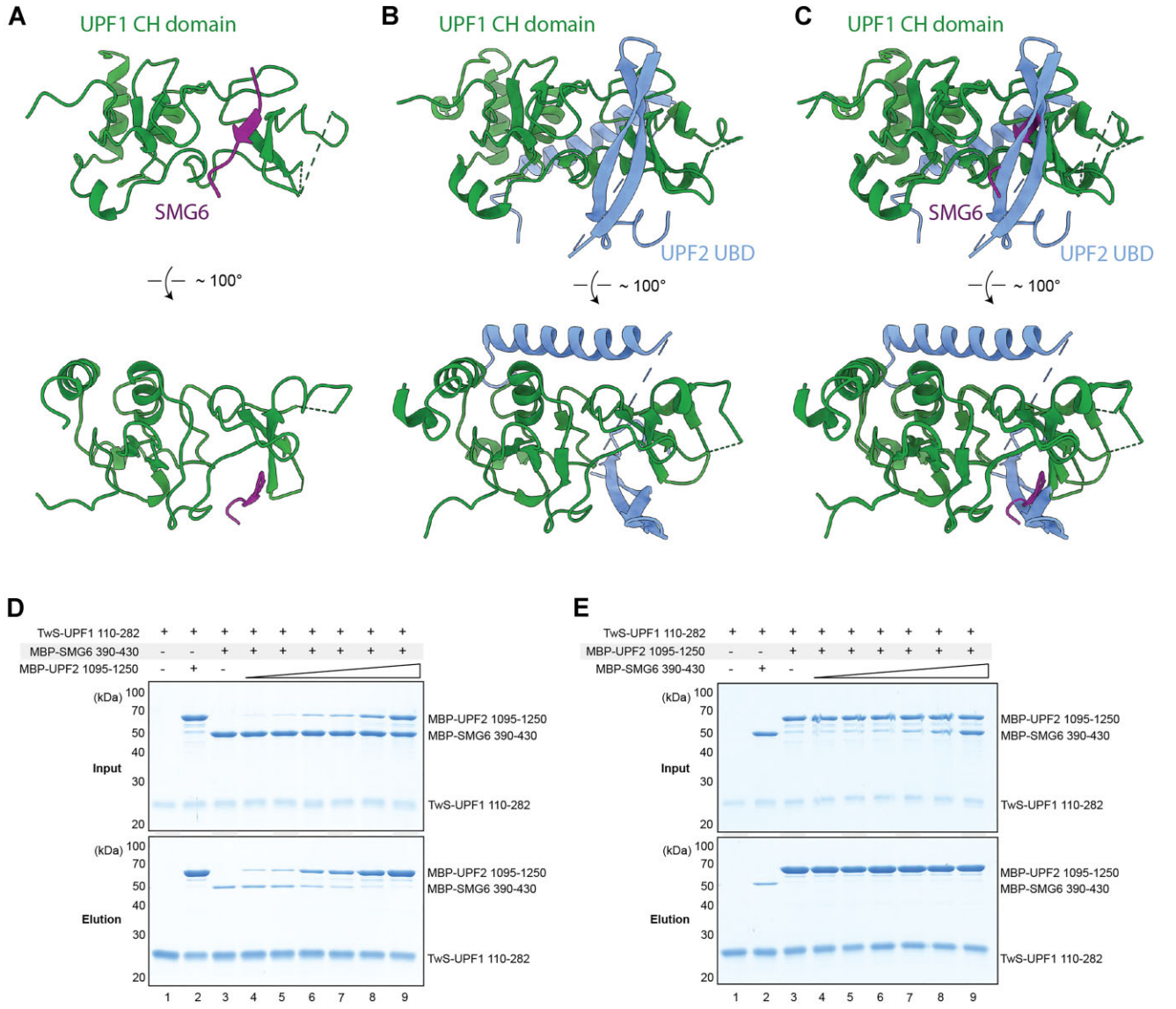


Figure 3. SMG6 and UPF2 compete for binding to the UPF1 CH domain. (**A**, **B**) Model of SMG6 (this study) or UPF2$_{CHBD}$ (PDB: 2WJV) bound to human UPF1 CH domain shown in two different orientations. (**C**) Superposition of the models shown in (A) and (B). (**D**) Coomassie-stained SDS-PAGE of a competition assay using a pre-formed TwinStrep-tagged (TwS) UPF1 CH domain—SMG6 complex with increasing amounts of UPF2$_{CHBD}$ added. (**E**) As (D), but starting with a pre-formed TwinStrep-tagged (TwS) UPF1 CH domain—UPF2$_{CHBD}$ complex with increasing amounts of SMG6 added.

teins by antibody N pull-down to the control, all major NMD factors appeared to be enriched in the purified mRNPs, with the notable exception of the SMG6 endonuclease (Figure 1C). To assess whether the exclusion of SMG6 may be due to the inability of the UPF1 antibody N to detect this interaction, we carried out an analogous antibody-based purification with recombinant UPF1 and SMG6 (Figure 1, Supplementary Figure 3C). This control experiment in the context of purified proteins showed that UPF1 antibody N is in principle capable of co-precipitating UPF1 and SMG6 (Figure 1, Supplementary Figure 3C), in turn suggesting that the exclusion of SMG6 from UPF1-containing mRNPs is genuine (Figure 1C).

Next, we assessed the capability of the endogenous UPF1 purification strategy to detect changes in mRNP composition by comparing the data obtained from cells treated with either SMG1i (a selective inhibitor of SMG1 kinase activity that is solubilized in the presence of DMSO) or with DMSO only as control (47,48). When comparing triplicates of both samples, we identified an enrichment of the endogenous SMG1-8-9 kinase complex in the SMGi-treated sample (Figure 1D). This observation is in agreement with our previous findings that recombinant SMG1-8-9 kinase complex can efficiently bind UPF1 in the presence of the SMG1 inhibitor, even if SMG1-8-9 cannot phosphorylate UPF1 (47). The analysis also showed the depletion of SMG5-SMG7 in the SMGi-treated sample, in line with previous data that SMG5–SMG7 heterodimer interacts with UPF1 in a phospho-dependent manner (32). Finally, endogenous SMG6 levels were unchanged in the SMG1i-treated sample, consistent with earlier studies indicating that the UPF1 phosphorylation status does not impact SMG6 interaction with UPF1 (22,24,25). Puzzled by the finding that the NMD factor SMG6 appears to be selectively excluded from UPF1-containing mRNPs in normal conditions, we set out to better understand the molecular mechanisms underpinning the UPF1–SMG6 interaction.

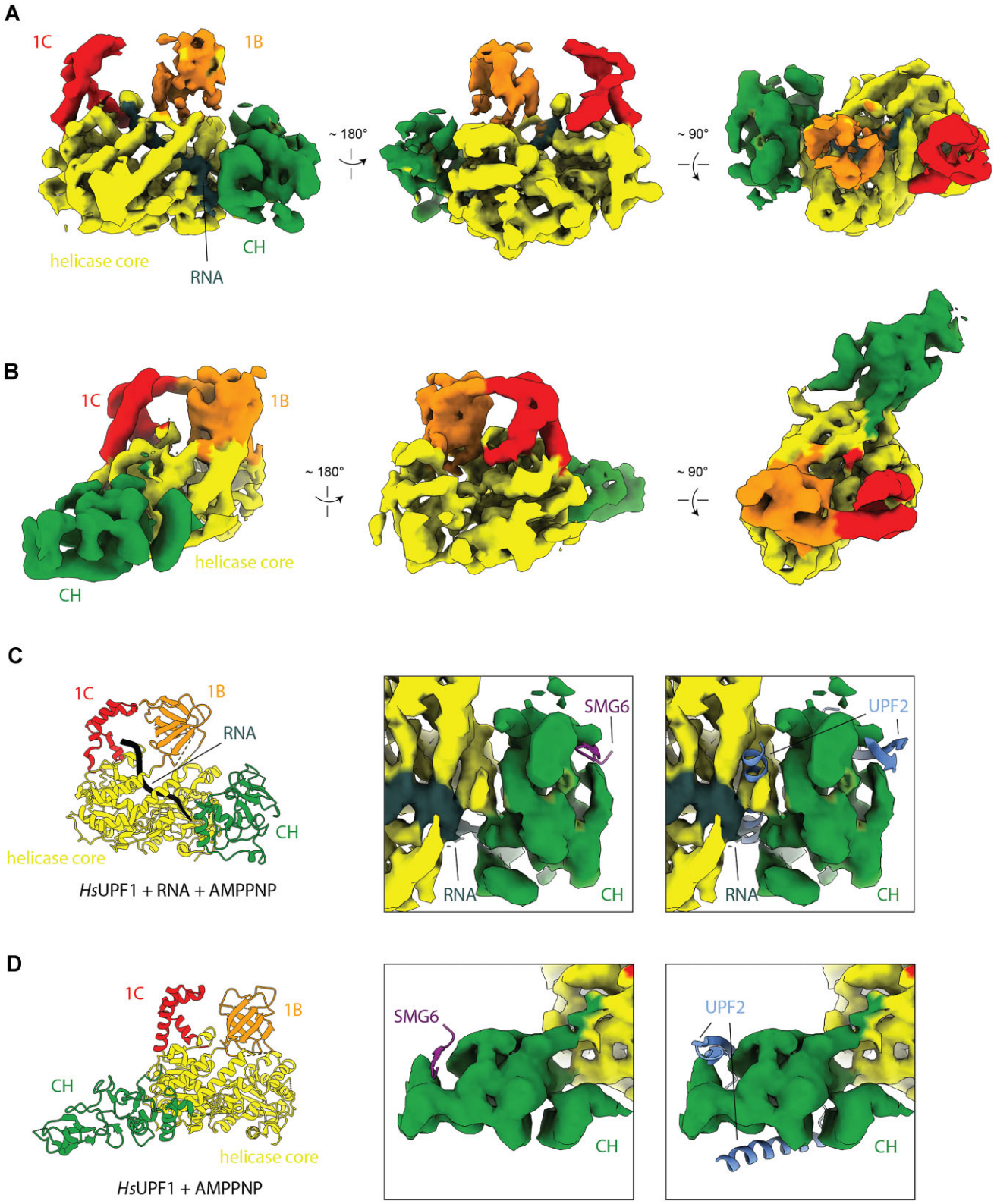


Figure 4. Cryo-EM reconstructions of full-length human UPF1 in its RNA-bound or RNA-free state. (**A**) Cryo-EM density of full-length human UPF1 reconstituted with AMPPNP and U_{15}-RNA shown in three different orientations. Domains and RNA colored as indicated. (**B**) Cryo-EM density of full-length human UPF1 reconstituted with AMPPNP shown in three different orientations. Domains colored as indicated and helicase domain in the same orientation as in (A). (**C**) Model of RNA-bound UPF1 and close-ups of the corresponding CH domain density fitted with models for SMG6 (this study) or $UPF2_{CHBD}$ (PDB: 2WJV). (**D**) Model of RNA-free UPF1 and close-ups of the corresponding CH domain density fitted with models for SMG6 (this study) or $UPF2_{CHBD}$ (PDB: 2WJV).

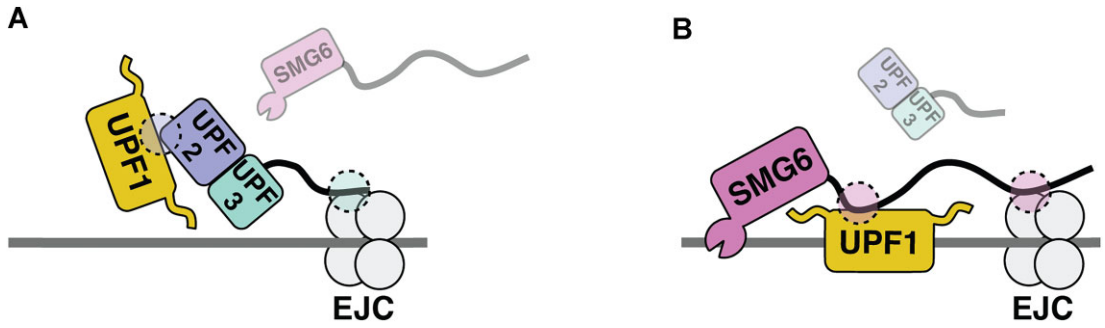


Figure 5. Model for two ways of bridging UPF1 to the EJC via mutually exclusive interactions. Cartoon models highlighting how UPF1 could be linked to the EJC in two distinct complexes—*via* UPF2-UPF3B (panel **A**) or *via* SMG6 (panel **B**). The two interaction sites that are mutually exclusive between either complex are highlighted by transparent circles.

A short linear motif in SMG6 interacts with the UPF1 CH domain

In previous work, we mapped the phospho-independent interaction between human UPF1 and SMG6 to a segment spanning residues 207–580 in the N-terminal unstructured region of SMG6 (Figure 2A) (24). When using AlphaFold-Multimer (31,32) to predict potential binding interfaces between this segment of SMG6 and full-length UPF1, we obtained consistent predictions between the UPF1 CH domain ($UPF1_{CH}$) and SMG6 residues 406–413 (Figure 2 Supplementary Figure 1 and 2), which are highly conserved across metazoans (Figure 2, Supplementary Figure 1D). We proceeded to test and thereby validate the computational model experimentally.

First, we crystallized and determined the 2.6 Å resolution structure of the complex between $UPF1_{CH}$ (residues 115–287) and the SMG6 SLIM encompassing residues 404–418, confirming the model obtained by Alpha-Fold Multimer predictions (Figure 2A–C and Supplementary Figure 2 and Table 1). The $UPF1_{CH}$ domain adopts a compact fold with three structural zinc ions, characterized by a central four-stranded anti-parallel β-sheet connected to a three-stranded antiparallel β-sheet via a linker, as described before (15,49). These two β-sheets of $UPF1_{CH}$ are arranged in a perpendicular fashion forming a hydrophobic cavity at their interface. The SMG6 SLIM docks as a β-strand in this cavity, interacting in particular via Gly407 and Leu409.

Next, we engineered structure-based mutations to disrupt the $UPF1_{CH}$–SMG6 interface and tested their effect in pull-down assays with purified proteins (Figure 2D, E and Figure 2, Supplementary Figure 1C). TwinStrep-tagged $UPF1_{CH}$ (TwS-$UPF1_{CH}$) could efficiently pull down a maltose-binding protein (MBP)-tagged SMG6 construct encompassing residues 390–430 (hereby referred to as SMG6 CH-binding domain or $SMG6_{CHBD}$), and this interaction was abolished when mutating the conserved SMG6 residue Leu409. The detrimental effect of this mutation also correlated with AlphaFold predictions comparing either wild-type (WT) or Leu409Glu SMG6 mutant (Figure 2, Supplementary Figure 2A and B). Conversely, mutation of the conserved Val205 in TwS-$UPF1_{CH}$ impaired the interaction with MBP-$SMG6_{CHBD}$ (Figure 2D). When using TwinStrep-tagged UPF1 full-length, we again observed an interaction with MBP-SMG6 390–430, which was disrupted by the Leu409Glu as well as a Gly407Glu mutation. Consistently, an unrelated TwinStrep-tagged protein did not interact with MBP-SMG6 390–430 (Figure 2, Supplementary Figure 1C). We next assessed the impact of the mutations in the context of both full-length proteins. Untagged full-length UPF1 was obtained from overexpression and purification from bacterial cells, to eliminate potential contributions of UPF1 phosphorylation. Full-length TwinStrep-tagged SMG6 (TwS-$SMG6_{FL}$), either WT or Leu409Glu mutant, was overexpressed and purified from HEK293T cells. In a pull-down experiment, WT TwS-$SMG6_{FL}$ interacted with untagged UPF1, whereas the Leu409Glu mutant strongly reduced binding (Figure 2E). In conclusion, these structural and biochemical results indicate that we have identified a UPF1-binding site within SMG6 that is crucial for the interaction of the full-length UPF1 and SMG6 proteins *in vitro*.

SMG6 and UPF2 compete for binding to the UPF1 CH domain

The $UPF1_{CH}$ is known to also recognize UPF2. In particular, UPF2 residues 1095–1250 ($UPF2_{CHBD}$) wrap around $UPF1_{CH}$, binding with a β-hairpin motif at the hydrophobic cavity formed between the two perpendicular $UPF1_{CH}$ β-sheets and with an α-helix at the diametrically opposite surface (15). Upon superposing the $UPF2_{CHBD}$-bound and $SMG6_{CHBD}$-bound structures, it became evident that the $SMG6_{CHBD}$ and the $UPF2_{CHBD}$ β-hairpin occupy the same surface of $UPF1_{CH}$ (Figure 3A and B) and the concomitant binding of both would therefore be incompatible (Figure 3C). To test the hypothesis of mutually exclusive interactions, we performed competition assays using purified proteins (Figure 3D and E). We pre-formed complexes between TwS-$UPF1_{CH}$ with either MBP-$SMG6_{CHBD}$ or MBP-$UPF2_{CHBD}$ and challenged them with increasing amounts of the putative competing protein (MBP-$UPF2_{CHBD}$ or MBP-$SMG6_{CHBD}$, respectively). We observed that MBP-$UPF2_{CHBD}$ could readily replace MBP-$SMG6_{CHBD}$ from TwS-$UPF1_{CH}$ (Figure 3D). When using the same concentration range, MBP-$SMG6_{CHBD}$ could not replace MBP-$UPF2_{CHBD}$ (Figure 3E). These results are consistent with the structural analysis showing that $UPF2_{CHBD}$ has an additional $UPF1_{CH}$ binding element (the α-helix) and therefore a more extensive binding interface as compared to $SMG6_{CHBD}$. Consistently, binding affinity measurements using isothermal calorimetry revealed that TwS-$UPF1_{CH}$ engages MBP-$UPF2_{CHBD}$ with a higher affinity (K_D = 1.46 ± 0.25 μM) as compared to MBP-$SMG6_{CHBD}$ (K_D = 4.28 ± 0.25 μM) (Figure 3, Supplementary Figure 1). These results not only demonstrate that $SMG6_{CHBD}$ and $UPF2_{CHBD}$ compete for binding to $UPF1_{CH}$, but also that $UPF2_{CHBD}$ prevails over

$SMG6_{CHBD}$ in the conditions tested, as it recognizes $UPF1_{CH}$ via a larger interaction interface and with higher affinity.

The conformation of RNA-bound UPF1 is compatible with binding to SMG6 but not UPF2

The observations above rationalized why SMG6 is prevented from associating with UPF1-mRNPs under normal conditions, i.e., due to preferential binding of $UPF2_{CHBD}$, but in turn raised the question of what conditions or signals might trigger the switch for recruiting $SMG6_{CHBD}$ to $UPF1_{CH}$. Previous structural work on the human UPF1–$UPF2_{CHBD}$ complex has shown that the CH domain adopts an open conformation with respect to the helicase core (15). In contrast, the structure of budding yeast Upf1 bound to RNA has revealed a diametrically opposite position of the CH domain with respect to the helicase core, adopting a closed, RNA clamping conformation (14). Biochemical data have recently demonstrated that binding of UPF2 to UPF1 strongly reduces UPF1's affinity for RNA (16). Taking these data into account, we reasoned that structural conformations associated with the RNA-binding status of UPF1 might impact its ability to interact with either UPF2 or SMG6.

We therefore set out to generate a more comprehensive picture of the structural conformations assumed by full-length human UPF1, including the RNA-bound and RNA-free states. We reconstituted purified full-length human UPF1 with a non-hydrolyzable ATP analogue (AMPPNP) in the presence and absence of a short RNA (U_{15}), and subjected this sample to single-particle cryo-EM analysis. In the presence of RNA, the resulting reconstruction reached an overall resolution of ∼6 Å, allowing us to assign all known structured domains of UPF1 (Figure 4A and Supplementary Figure 1), aside from the unstructured N- and C-termini for which there was no discernible density. In the RNA-bound form of human $UPF1_{FL}$, the CH domain adopts a closed conformation (Figure 4, Supplementary Figure 2A and C) similar to the one previously observed in the crystal structure of yeast RNA-bound $Upf1_{CH\text{-}helicase}$ (except for slightly different relative orientations of the individual domains, Figure 4, Supplementary Figure 2A). In the absence of RNA, the resulting reconstruction reached a similar resolution and showed that the UPF1 CH domain adopts an open conformation, similar to the one identified in a crystal structure of human UPF1 CH helicase bound to $UPF2_{CHBD}$ but different from the closed RNA-bound state (Figure 4B and Supplementary Figure 1B and Supplementary Figure 2B, C). Thus, the open conformation is not triggered by UPF2 binding, but rather appears to be an inherent feature of the RNA-free state. It seems likely that UPF1 cycles between the open and closed conformations during its catalytic activity that is stimulated by UPF2 (14). The UPF1 reconstruction in the absence of RNA also showed two additional features.

First, while most of the unstructured N and C-termini appeared to remain disordered, density corresponding to the N-terminal segment that precedes RecA1 occupied part of the RNA-binding channel (where the RNA 3′-end would reside) (Figure 4, Supplementary Figure 2D). Second, the so-called 'regulatory loop' (protruding from domain 1B in the helicase core) exhibited ordered density and extended towards the RNA-binding channel (where the RNA 5′-end would reside) (Figure 4, Supplementary Figure 2E). As a note, the corresponding loop in the less abundant UPF1 isoform is even more extended, and in both UPF1 isoforms the loop has indeed been implicated in modulating the RNA-binding properties *in vitro* and *in vivo* (50,51). More broadly, the autoinhibiting interactions of the UPF1 N-terminus and regulatory loop that we observed in the reconstruction of the RNA-free state are expected to persist upon binding to UPF2, as they appear to be specific to the open conformation of UPF1 (Figure 4, Supplementary Figure 2E) (16), consistent with the low RNA-binding affinity of UPF1 when in complex with UPF2. Furthermore, parts of the $UPF2_{CHBD}$ binding site are accessible only in the RNA-free state and not in the RNA-bound conformation of UPF1 (Figure 4C and D). In contrast, the $SMG6_{CHBD}$ binding site is fully accessible in both the RNA-free and RNA-bound conformations of human $UPF1_{FL}$. Taken together, our cryo-EM data show that the RNA-free and UPF2-bound conformations of human UPF1 are structurally distinct from the RNA-bound conformation, and that in the latter the UPF2-binding site is partially blocked. These observations suggest that the switch of UPF1 from the UPF2-bound state to the SMG6-bound state may be regulated by RNA binding.

Discussion

In this study, we isolated endogenous UPF1-containing mRNPs from human cells and found that they exhibited a significant enrichment of all known NMD factors, with the notable exception of the SMG6 endonuclease. SMG6 carries out the endonucleolytic cleavage event in NMD that provides the entry point to the processive exoribonucleases for degrading the body of the mRNA (26,27). It is thus a conceptually reasonable notion that SMG6 may be actively excluded from mRNPs in homeostatic conditions, as its action is effectively a death knell for the transcript. This finding raised the question of how is SMG6 recruited to UPF1 once an mRNA has been committed to NMD-mediated degradation.

It was previously shown that SMG6 interacts with UPF1 in a phosphorylation-independent manner (24,25). Using biochemical and structural methods, we found that a short linear motif in the SMG6 N-terminus interacts with the UPF1 CH domain. SMG6 is recognized at the same surface of the UPF1 CH domain where UPF2 also binds, but in the latter case with more extensive interactions and higher affinity (15). Indeed, in a biochemically reconstituted system, the presence of UPF2 precludes the binding of SMG6 to the UPF1 CH domain. Interestingly, another region of SMG6 has been shown to engage in mutually exclusive interactions with UPF3, as they both recognize the same surface of the EJC (29). Taken together, these observations lead to a model whereby mutually exclusive and competing protein-protein interactions underlie rearrangements in UPF1-containing mRNPs at different stages of the NMD pathway, from an initial EJC-UPF3-UPF2-UPF1 assembly to a later EJC-SMG6-UPF1 assembly.

The compositional rearrangements of UPF1 from a UPF2-bound complex to a SMG6-bound complex are connected to conformational rearrangements of the helicase from an RNA-free state to an RNA-bound state. Structural data of full-length human UPF1 obtained using single-particle cryo-EM revealed that the RNA-bound 'closed' conformation is structurally distinct from the 'open' conformation characteristic of both the UPF2-bound and RNA-free states. In the open RNA-free conformation, UPF1 is auto-inhibited by the insertion of unstructured regions in the RNA-binding channel. These findings rationalize previous observations that UPF1 does not efficiently bind RNA when bound to UPF2 (16).

Importantly, they demonstrate that the closed conformation of RNA-bound UPF1 is incompatible with UPF2 binding but compatible with SMG6 binding. The finding that the RNA-dependent conformational switch of UPF1 is connected to the accessibility of distinct NMD factors suggests that UPF2 displacement and SMG6 recruitment are orchestrated around RNA binding to UPF1. It is currently unclear which comes first—the RNA-dependent conformational switch or the protein conformational switch—in the transition from an initial EJC–UPF3–UPF2–UPF1 complex to a later EJC–SMG6–UPF1–RNA complex (Figure 5). Likely candidates in initiating either switch are UPF1 phosphorylation and translational termination events. Incorporating these events and the factors involved in them in future studies will be required to have a broader and more comprehensive framework of the steps that ultimately lead to the specific degradation of metazoan NMD substrates.

Data availability

The PDB file describing the crystal structure of human $UPF1_{CH}$ in complex with $SMG6_{CHBD}$ has been deposited to the Protein Data Bank with the accession code 8RXB. The EM densities can be downloaded from the Electron Microscopy Data Bank. RNA-bound UPF1: EMD-19451. RNA-free UPF1: EMD-19450. A file containing the proteomics data is attached to this manuscript and data are available via ProteomeXchange with identifier PXD049474.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We thank Barbara Steigenberger at the MPIB Biochemistry Core facility for mass spectrometry data acquisition. We are grateful to Tilman Schäfer and Daniel Bollschweiler at the MPIB cryo-EM facility for outstanding infrastructure and J. Rajan Prabu and Christian Benda for excellent maintenance of computational infrastructure. We thank Courtney Long for invaluable help in preparing the manuscript. We are grateful to Petra Birle, Tatjana Krywcun and Elisabeth Stegmann for assistance and all members of our group for thoughtful discussions. The authors are grateful to Robert Bridges, Rosalind Franklin University of Medicine and Science, and the Cystic Fibrosis Foundation for providing the SMG1 inhibitor.

Author contributions: Lukas M. Langer: Conceptualization, Investigation, Supervision, Formal analysis, Funding acquisition, Methodology, Validation, Writing—original draft, Writing—review & editing. Katharina Kurscheidt: Methodology and Investigation. Jerome Basquin: Methodology and Investigation. Fabien Bonneau: Investigation and Formal analysis. Claire Basquin: Investigation and Formal analysis. Iuliia Iermak: Conceptualization, Writing-review & editing. Elena Conti: Conceptualization, Supervision, Funding acquisition, Writing—original draft, Writing—review & editing.

Funding

Max Planck Gesellschaft; European Research Council Advanced Investigator Grant GOVERNA [101054447 to E.C.]; Deutsche Forschungsgemeinschaft [DFG SFB1035 to E.C.]; Boehringer Ingelheim Fonds PhD fellowship (to L.L.). Funding for open access charge: Max Planck Gesellschaft.

Conflict of interest statement

None declared.

References

1. Monaghan,L., Longman,D. and Caceres,J.F. (2023) Translation-coupled mRNA quality control mechanisms. *EMBO J.*, **42**, e114378.
2. D'Orazio,K.N. and Green,R. (2021) Ribosome states signal RNA quality control. *Mol. Cell*, **81**, 1372–1383.
3. Celik,A., He,F. and Jacobson,A. (2017) NMD monitors translational fidelity 24/7. *Curr. Genet.*, **63**, 1007–1010.
4. Kurosaki,T., Popp,M.W. and Maquat,L.E. (2019) Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat. Rev. Mol. Cell Biol.*, **20**, 406–420.
5. Karousis,E.D. and Muhlemann,O. (2022) The broader sense of nonsense. *Trends Biochem. Sci.*, **47**, 921–935.
6. Chu,V., Feng,Q., Lim,Y. and Shao,S. (2021) Selective destabilization of polypeptides synthesized from NMD-targeted transcripts. *Mol. Biol. Cell*, **32**, ar38.
7. Udy,D.B. and Bradley,R.K. (2022) Nonsense-mediated mRNA decay uses complementary mechanisms to suppress mRNA and protein accumulation. *Life Sci. Alliance.*, **5**, e202101217.
8. Inglis,A.J., Guna,A., Galvez-Merchan,A., Pal,A., Esantsi,T.K., Keys,H.R., Frenkel,E.M., Oania,R., Weissman,J.S. and Voorhees,R.M. (2023) Coupled protein quality control during nonsense-mediated mRNA decay. *J. Cell Sci.*, **136**, jcs261216.
9. Kishor,A., Fritz,S.E. and Hogg,J.R. (2019) Nonsense-mediated mRNA decay: The challenge of telling right from wrong in a complex transcriptome. *Wiley Interdiscip Rev RNA*, **10**, e1548.
10. Le Hir,H., Sauliere,J. and Wang,Z. (2016) The exon junction complex as a node of post-transcriptional networks. *Nat. Rev. Mol. Cell Biol.*, **17**, 41–54.
11. Chamieh,H., Ballut,L., Bonneau,F. and Le Hir,H. (2008) 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.
12. Dostie,J. and Dreyfuss,G. (2002) Translation is required to remove Y14 from mRNAs in the cytoplasm. *Curr. Biol.*, **12**, 1060–1067.
13. Gehring,N.H., Lamprinaki,S., Kulozik,A.E. and Hentze,M.W. (2009) Disassembly of exon junction complexes by PYM. *Cell*, **137**, 536–548.
14. Chakrabarti,S., Jayachandran,U., Bonneau,F., Fiorini,F., Basquin,C., Domcke,S., Le Hir,H. and Conti,E. (2011) Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. *Mol. Cell*, **41**, 693–703.
15. Clerici,M., Mourao,A., Gutsche,I., Gehring,N.H., Hentze,M.W., Kulozik,A., Kadlec,J., Sattler,M. and Cusack,S. (2009) Unusual bipartite mode of interaction between the nonsense-mediated decay factors, UPF1 and UPF2. *EMBO J.*, **28**, 2293–2306.
16. Xue,G., Maciej,V.D., Machado de Amorim,A., Pak,M., Jayachandran,U. and Chakrabarti,S. (2023) Modulation of RNA-binding properties of the RNA helicase UPF1 by its activator UPF2. *RNA*, **29**, 178–187.
17. Fiorini,F., Boudvillain,M. and Le Hir,H. (2013) Tight intramolecular regulation of the human Upf1 helicase by its N- and C-terminal domains. *Nucleic Acids Res.*, **41**, 2404–2415.
18. Denning,G., Jamieson,L., Maquat,L.E., Thompson,E.A. and Fields,A.P. (2001) Cloning of a novel phosphatidylinositol kinase-related kinase: characterization of the human SMG-1 RNA surveillance protein. *J. Biol. Chem.*, **276**, 22709–22714.
19. Yamashita,A., Ohnishi,T., Kashima,I., Taya,Y. and Ohno,S. (2001) Human SMG-1, a novel phosphatidylinositol 3-kinase-related protein kinase, associates with components of the mRNA

surveillance complex and is involved in the regulation of nonsense-mediated mRNA decay. *Genes Dev.*, **15**, 2215–2228.
20. Langer,L.M., Gat,Y., Bonneau,F. and Conti,E. (2020) Structure of substrate-bound SMG1-8-9 kinase complex reveals molecular basis for phosphorylation specificity. *eLife*, **9**, e57127.
21. Ohnishi,T., Yamashita,A., Kashima,I., Schell,T., Anders,K.R., Grimson,A., Hachiya,T., Hentze,M.W., Anderson,P. and Ohno,S. (2003) Phosphorylation of hUPF1 induces formation of mRNA surveillance complexes containing hSMG-5 and hSMG-7. *Mol. Cell*, **12**, 1187–1200.
22. Fukuhara,N., Ebert,J., Unterholzner,L., Lindner,D., Izaurralde,E. and Conti,E. (2005) SMG7 is a 14-3-3-like adaptor in the nonsense-mediated mRNA decay pathway. *Mol. Cell*, **17**, 537–547.
23. Loh,B., Jonas,S. and Izaurralde,E. (2013) 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.
24. Chakrabarti,S., Bonneau,F., Schussler,S., Eppinger,E. and Conti,E. (2014) Phospho-dependent and phospho-independent interactions of the helicase UPF1 with the NMD factors SMG5-SMG7 and SMG6. *Nucleic Acids Res.*, **42**, 9447–9460.
25. Nicholson,P., Josi,C., Kurosawa,H., Yamashita,A. and Muhlemann,O. (2014) A novel phosphorylation-independent interaction between SMG6 and UPF1 is essential for human NMD. *Nucleic Acids Res.*, **42**, 9217–9235.
26. Huntzinger,E., Kashima,I., Fauser,M., Sauliere,J. and Izaurralde,E. (2008) SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. *RNA*, **14**, 2609–2617.
27. Eberle,A.B., Lykke-Andersen,S., Muhlemann,O. and Jensen,T.H. (2009) SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. *Nat. Struct. Mol. Biol.*, **16**, 49–55.
28. Schoenberg,D.R. and Maquat,L.E. (2012) Regulation of cytoplasmic mRNA decay. *Nat. Rev. Genet.*, **13**, 246–259.
29. Kashima,I., Jonas,S., Jayachandran,U., Buchwald,G., Conti,E., Lupas,A.N. and Izaurralde,E. (2010) SMG6 interacts with the exon junction complex via two conserved EJC-binding motifs (EBMs) required for nonsense-mediated mRNA decay. *Genes Dev.*, **24**, 2440–2450.
30. Buchwald,G., Ebert,J., Basquin,C., Sauliere,J., Jayachandran,U., Bono,F., Le Hir,H. and Conti,E. (2010) Insights into the recruitment of the NMD machinery from the crystal structure of a core EJC-UPF3b complex. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 10050–10055.
31. Evans,R., O'Neill,M., Pritzel,A., Antropova,N., Senior,A., Green,T., Žídek,A., Bates,R., Blackwell,S., Yim,J., *et al.* (2021) Protein complex prediction with AlphaFold-Multimer. bioRxiv doi: https://doi.org/10.1101/2021.10.04.463034, 04 October 2021, preprint: not peer reviewed.
32. Jumper,J., Evans,R., Pritzel,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Zidek,A., Potapenko,A., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.
33. Goddard,T.D., Huang,C.C., Meng,E.C., Pettersen,E.F., Couch,G.S., Morris,J.H. and Ferrin,T.E. (2018) UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.*, **27**, 14–25.
34. Evans,P.R. and Murshudov,G.N. (2013) How good are my data and what is the resolution? *Acta. Crystallogr. D Biol. Crystallogr.*, **69**, 1204–1214.
35. Kabsch,W. (2010) Integration, scaling, space-group assignment and post-refinement. *Acta. Crystallogr. D Biol. Crystallogr.*, **66**, 133–144.
36. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta. Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
37. Liebschner,D., Afonine,P.V., Baker,M.L., Bunkoczi,G., Chen,V.B., Croll,T.I., Hintze,B., Hung,L.W., Jain,S., McCoy,A.J., *et al.* (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.*, **75**, 861–877.
38. Zivanov,J., Nakane,T., Forsberg,B.O., Kimanius,D., Hagen,W.J., Lindahl,E. and Scheres,S.H. (2018) New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife*, **7**, e42166.
39. Scheres,S.H. (2012) RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.*, **180**, 519–530.
40. Zheng,S.Q., Palovcak,E., Armache,J.P., Verba,K.A., Cheng,Y. and Agard,D.A. (2017) MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods*, **14**, 331–332.
41. Rohou,A. and Grigorieff,N. (2015) CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.*, **192**, 216–221.
42. Punjani,A., Rubinstein,J.L., Fleet,D.J. and Brubaker,M.A. (2017) cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods*, **14**, 290–296.
43. Cox,J. and Mann,M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.*, **26**, 1367–1372.
44. Wickham,H., Averick,M., Bryan,J., Chang,W., McGowan,L., François,R., Grolemund,G., Hayes,A., Henry,L. and Hester,J. (2019) Welcome to the tidyverse. *J. Open Source Softw.*, **10**, 1686.
45. Bonneau,F., Basquin,J., Steigenberger,B., Schafer,T., Schafer,I.B. and Conti,E. (2023) Nuclear mRNPs are compact particles packaged with a network of proteins promoting RNA-RNA interactions. *Genes Dev.*, **37**, 505–517.
46. Schwanhausser,B., Busse,D., Li,N., Dittmar,G., Schuchhardt,J., Wolf,J., Chen,W. and Selbach,M. (2011) Global quantification of mammalian gene expression control. *Nature*, **473**, 337–342.
47. Langer,L.M., Bonneau,F., Gat,Y. and Conti,E. (2021) Cryo-EM reconstructions of inhibitor-bound SMG1 kinase reveal an autoinhibitory state dependent on SMG8. *eLife*, **10**, e72353.
48. Gopalsamy,A., Bennett,E.M., Shi,M., Zhang,W.G., Bard,J. and Yu,K. (2012) Identification of pyrimidine derivatives as hSMG-1 inhibitors. *Bioorg. Med. Chem. Lett.*, **22**, 6636–6641.
49. Kadlec,J., Guilligay,D., Ravelli,R.B. and Cusack,S. (2006) Crystal structure of the UPF2-interacting domain of nonsense-mediated mRNA decay factor UPF1. *RNA*, **12**, 1817–1824.
50. Gowravaram,M., Bonneau,F., Kanaan,J., Maciej,V.D., Fiorini,F., Raj,S., Croquette,V., Le Hir,H. and Chakrabarti,S. (2018) A conserved structural element in the RNA helicase UPF1 regulates its catalytic activity in an isoform-specific manner. *Nucleic Acids Res.*, **46**, 2648–2659.
51. Fritz,S.E., Ranganathan,S., Wang,C.D. and Hogg,J.R. (2022) An alternative UPF1 isoform drives conditional remodeling of nonsense-mediated mRNA decay. *EMBO J.*, **41**, e108898.

Received: February 10, 2024. **Revised:** April 6, 2024. **Editorial Decision:** April 8, 2024. **Accepted:** April 25, 2024