HHS Public Access

Author manuscript

J Mol Biol. Author manuscript; available in PMC 2016 May 18.

Published in final edited form as:

J Mol Biol. 2014 November 11; 426(22): 3689–3702. doi:10.1016/j.jmb.2014.09.015.

A highly conserved region essential for NMD in the Upf2 N-terminal domain

Zaineb Fourati[a,b,#], **Bijoyita Roy**[c,#], **Claudia Millan**[d,e], **Pierre-Damien Coureux**[a], **Stéphanie Kervestin**[f], **Herman van Tilbeurgh**[b], **Feng He**[c], **Isabel Usón**[d,e], **Allan Jacobson**[c,*], and **Marc Graille**[a,b,*]

[a]Laboratoire de Biochimie, CNRS, UMR 7654, Ecole Polytechnique, F-91128 Palaiseau Cedex, France

[b]Institut de Biochimie et Biophysique Moléculaire et Cellulaire (IBBMC), CNRS UMR8619, Université Paris Sud F-91405 Orsay Cedex, France

[c]Department of Microbiology and Physiological Systems, University of Massachusetts Medical School, Albert Sherman Center, 368 Plantation Street, Worcester, MA 01655-0122, USA

[d]Instituto de Biología Molecular de Barcelona (IBMB-CSIC), Barcelona Science Park, Baldiri Reixach 15, 08028 Barcelona, Spain

[e]Institució Catalana de Recerca i Estudis Avançats (ICREA), Spain

[f]Institut de Biologie Physico-Chimique, CNRS UPR9073, Université Paris Diderot Sorbonne Paris Cité, F-75005 Paris, France

Abstract

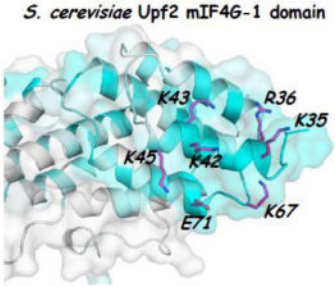
Upf1, Upf2, and Upf3 are the principal regulators of Nonsense-mediated mRNA Decay (NMD), a cytoplasmic surveillance pathway that accelerates the degradation of mRNAs undergoing premature translation termination. These three proteins interact with each other, the ribosome, the translation termination machinery, and multiple mRNA decay factors, but the precise mechanism allowing the selective detection and degradation of nonsense-containing transcripts remains elusive. Here we have determined the crystal structure of the N-terminal mIF4G domain from *Saccharomyces cerevisiae* Upf2 and identified a highly conserved region in this domain that is essential for NMD and independent of Upf2's binding sites for Upf1 and Upf3. Mutations within this conserved region not only inactivate NMD, but also disrupt Upf2 binding to specific proteins, including Dbp6, a DEAD-box helicase. Although current models indicate that Upf2 functions principally as an activator of Upf1 and a bridge between Upf1 and Upf3, our data suggest that it

[*]To whom correspondence should be addressed. Dr Marc Graille. Laboratoire de Biochimie, CNRS, UMR 7654, Ecole Polytechnique, F-91128 Palaiseau Cedex, France. Tel: +33-(0)1-69-33-48-90; Fax: +33-(0)1-69-33-49-09; marc.graille@polytechnique.edu.; Pr Allan Jacobson. Department of Microbiology and Physiological Systems, University of Massachusetts Medical School, Albert Sherman Center, 368 Plantation Street, Worcester, MA 01655-0122, USA. Tel: 1-508-856-2442; Fax: 1-508 856 5920; allan.jacobson@umassmed.edu.

[#]Both authors contributed equally.

may also serve as a platform for the association of additional factors that play roles in premature translation termination and NMD.

Graphical abstract



Upf2 mutant	NMD
WT	☺
K35E/R36E	☠
K42E/K43E/K45E	☠
K67A/E71A	☹

INTRODUCTION

Nonsense-mediated mRNA decay (NMD) is one of three conserved eukaryotic surveillance pathways ensuring mRNA quality control in the cytoplasm[1; 2; 3; 4]. NMD is activated by mechanistic differences between normal and premature translation termination and utilizes three conserved factors (Upf1, Upf2, and Upf3)[5; 6; 7] to couple nonsense codon recognition to the release factors (eRF1 and eRF3), the ribosome, and the mRNA decapping complex[8; 9; 10; 11; 12; 13; 14]. Upf1, the central regulator of NMD[2], encompasses an N-terminal zinc knuckle CH domain and a functional helicase core domain exhibiting RNA binding and ATPase activities[15; 16]. The Upf1 CH domain binds to the C-terminal region of Upf2, the largest component of the Upf1-Upf2-Upf3 surveillance complex[17]. The N-terminal two-thirds of Upf2 include three domains that adopt the same fold as the middle domain of eukaryotic initiation factor eIF4G (mIF4G, Fig. 1A)[18; 19; 20; 21]. The third of these mIF4G domains (mIF4G-3) interacts with the central RNA Recognition Motif (RRM) from Upf3[20]. Accordingly, Upf2 is thought to act as a scaffolding protein that bridges Upf1 and Upf3[15; 17; 22]. Other factors are necessary for NMD in higher eukaryotes[5] and appear to regulate the activity or availability of the Upfs. Thus, in humans, the SMG-1 and SMG-5 to 9 proteins, as well as the PP2A phosphatase, act as NMD enhancers by regulating UPF1's phosphorylation status[23; 24; 25] and the exon-junction complex (EJC) interacts with the human UPF3b isoform[26; 27].

Although multiple mechanistic models have been proposed for NMD in lower and higher eukaryotes[2; 28; 29; 30; 31; 32; 33], little is known about the precise role of the three Upf proteins in the individual steps of NMD: recognition of prematurely terminating mRNAs, targeting of an mRNA and its nascent polypeptide for accelerated degradation, and disassembly of the prematurely terminating mRNP complex. The most elaborate details uncovered to date pertain to Upf1:Upf2 interaction and the possible role of this interaction in activating Upf1's ATPase and helicase activities that are essential for NMD. Prior to interacting with Upf2, Upf1 adopts a closed conformation that appears to optimize its RNA binding capabilities while simultaneously minimizing its ATPase and helicase activities[15; 34]. Upf2 interaction with the Upf1 CH domain[35] triggers a major conformational change of the CH domain, opening the structure of Upf1 while simultaneously reducing its RNA-binding activity and

stimulating its ATPase and helicase activities[34]. These effects are comparable to the consequences of deleting the Upf1 CH domain[15; 34]. While these results provide some insights into the function of Upf2's Upf1-interaction domain, they leave unresolved the role of a large fragment of Upf2, namely the N-terminal half encompassing the first and second mIF4G domains. Previous study has highlighted the critical role of the whole N-terminal domain and particularly the region spanning residues 30–50 of yeast Upf2 mIF4G1 domain for NMD[36]. Very recently, it was also shown that the deletion of the human UPF2 mIF4G-1 or mIF4G-2 domain, or both, completely inhibits NMD without affecting UPF2 protein stability or cellular localization or UPF1 and UPF3 recruitment[19]. However, tethering of these truncated UPF2 proteins to PTC-containing mRNA partially restores NMD, suggesting that the mIF4G-1 and mIF4G-2 UPF2 domains are important for recruitment of UPF2 to PTC-containing mRNAs[19]. Here, we have pursued structural and functional analyses of the role of the Upf2 mIF4G-1 domain in yeast NMD. Its crystal structure led to the characterization of a highly conserved region within this domain that is essential for NMD, and to the identification of proteins whose association with Upf2 is dependent on conserved residues within this domain. Our results suggest that Upf2 may have previously unanticipated mechanistic roles in NMD beyond serving as an activator of Upf1 activity or a bridge between Upf1 and Upf3.

RESULTS

Crystal structure of the S. cerevisiae Upf2 N-terminal domain

To investigate the biochemical and structural roles of the Upf2 N-terminal region in NMD, we cloned and expressed in *E. coli* two fragments from *Saccharomyces cerevisiae* Upf2 (hereafter designated ScUpf2). One construct corresponds to the first N-terminal mIF4G domain (amino acids 1-360) while the other spans the three predicted Upf2 mIF4G domains (amino acids 1-820, Fig. 1A). Although both ScUpf2 fragments could be purified to homogeneity in large quantities, neither yielded diffracting crystals. The larger fragment proved to be very unstable and two species of lower molecular weight (25 and 35 kDa) were observed on SDS-PAGE after one day of storage at 4°C (data not shown). Analysis of these bands by mass spectrometry following trypsin digestion revealed that the 25 kDa band corresponded to the second and third mIF4G domains while the 35 kDa band corresponded to the first mIF4G Upf2 domain (amino acids 1-310, hereafter named ScUpf2 mIF4G-1). This prompted us to clone, express, and purify the ScUpf2 mIF4G-1 fragment. This domain yielded crystals from which we could collect datasets up to 1.45 Å resolution. Although this structure could not be solved by classical methods (see Materials and Methods for details), the high resolution datasets proved to be crucial for *ab initio* structure determination using the ARCIMBOLDO program ([37]; see Table 1 for dataset and refinement statistics). Indeed, this approach, which combines 16-residue α-helix location with PHASER[38], and density modification and auto-tracing with SHELXE[39] in a supercomputing frame, yielded an electron density map from the correct positioning of only two α-helices (Fig. S1). The high quality of this map allowed building of the model of the entire ScUpf2 mIF4G-1 domain, which was refined against both the 1.45 Å resolution and the 1.65 Å resolution datasets. As the statistics and electron density maps were of significantly better quality for the 1.65 Å resolution dataset, only this model will be discussed.

A highly conserved and functionally essential region in the Upf2 mIF4G-1 domain

ScUpf2 mIF4G-1 is an all α-helical protein composed of 11 α-helices (hA and h1 to h10) belonging to the large HEAT repeat family (Fig. 1B). This domain adopts a typical mIF4G fold composed of 5 pairs of anti-parallel helices stacked against each other to form a solenoid structure. An additional helix (hA) is located at the N-terminus of the domain and folds back onto it on the convex face of the solenoid. The concave and convex faces of the solenoid are formed by the first (odd number) and second (even number) helices of each repeat, respectively. Compared to canonical HEAT repeats that comprise approximately 40 amino acids[40; 41], the size of the ScUpf2 HEAT repeats increases from the N- to the C-terminal end of the mIF4G-1 domain: repeats 1 and 2 are ~40 residues long, while repeats 3, 4, and 5 are 55, 65, and 90 residues long, respectively. The overall structure of this yeast mIF4G1 domain is similar to that of the recently determined corresponding domain from human UPF2 (2.68Å RMSD over 224 Cα, 19% sequence identity; Fig. 1C;[19]). In human UPF2 mIF4G-1, helix hA is much longer than the corresponding helix from ScUpf2 and an additional helix (h8i) is inserted between helices h8 and h9 (Fig. S2). In addition, the region corresponding to residues 25 to 28 from ScUpf2 folds as an α-helix (hB) in human UPF2. Deeper structural comparison reveals that the relatively high rmsd value between both structures results from a narrowed angle between two halves of the mIF4G-1 domain. Indeed, superimposition of the first half of both mIF4G-1 domains revealed a 10° rotation of the C-terminal half (formed by helices h6 to h10 in ScUpf2) relative to the N-terminal half (formed by helices hA and h1 to h5 in ScUpf2) in human UPF2 compared to yeast Upf2, resulting in a more compact overall structure of the human UPF2 mIF4G-1 domain (Fig. 1C). Whether this is due to crystal packing constraints needs to be further investigated by solution approaches.

We next mapped the sequence conservation score of eukaryotic Upf2 proteins at the surface of the ScUpf2 mIF4G-1 domain (Fig. 1D). This analysis revealed two highly conserved regions located at the opposite ends of the domain. One patch is formed by residues covering the C-terminal end of helix h9 to the N-terminal end of helix h10. Residues from this region are strongly conserved in fungi, but differ from residues present at these positions in metazoa (Fig. S2), suggesting that this region may not be critical for NMD in all eukaryotic organisms. The second conserved patch is much larger and is formed by several residues strictly conserved from fungi to metazoa (Fig. 1D and Fig. S2). It is located at the N-terminal end of the domain and is formed by residues from the linker preceding helix h1 and from solvent exposed residues from helices h1 to h3. This patch is centered on helix h1 and also contains Asp31 and Ser32, and Lys35 and Arg36, all of which were previously shown to be important for NMD in yeast[36]. Mapping of the electrostatic potential at the surface of the ScUpf2 mIF4G-1 domain revealed that this conserved N-terminal region is enriched in basic residues (Fig. 1E).

We addressed the role of the latter conserved region in yeast NMD by mutating strongly conserved residues to generate three *upf2* alleles (Fig. S2). First, we introduced charge inversion along helix h1 by replacing Lys35 and Arg36 as well as Lys42, Lys43 and Lys45 with glutamic acid, creating the K35E/R36E and K42E/K43E/K45E alleles (Fig. 2A). We also mutated Lys67 and Glu71 from helix h3 to alanine to generate the K67A/E71A allele.

The effects of these mutations on NMD *in vivo* were studied by monitoring the abundance of two nonsense-containing NMD reporters, the *CYH2* pre-mRNA and the *ade2-1* mRNA, in *upf2Δ* cells that had been transformed with single-copy plasmids expressing each of the three alleles (Fig. 2B). As previously demonstrated[17], deletion of the *UPF2* gene in *S. cerevisiae* led to stabilization and markedly increased abundance of both the *ade2-1* mRNA and *CYH2* pre-mRNA without affecting the NMD-immune *CYH2* mRNA (Fig. 2B). Expression of an untagged or a HA-tagged plasmid-borne version of wild-type full-length *UPF2* in the *upf2Δ* strain complemented those cells' NMD deficiency and led to reductions of the *CYH2* pre-mRNA and *ade2-1* mRNA (Fig. 2B), yielding levels comparable to those observed in wild-type cells[17]. In contrast, expression of the three HA-tagged full-length *upf2* alleles in the *upf2Δ* strain did not complement the *UPF2* deletion and resulted in the accumulation of both the *ade2-1* mRNA and *CYH2* pre-mRNA (Fig. 2B). These defects in NMD are not due to instability of the respective Upf2 protein mutants as all three are expressed at approximately the same level as wild-type Upf2 (Fig. S3). The three mutants exhibited different levels of NMD inhibition, with the strongest effects (approximately six-fold increase in the levels of the two reporter transcripts compared to wild-type protein) obtained with the K42E/K43E/K45E and K35E/R36E mutants (Fig. 2B). A smaller but still substantial effect (approximately three- to four-fold increase in the levels of the two reporter transcripts) was observed with the K67A/E71A mutant (Fig. 2B). mRNA stabilizing effects of other alleles of K35 and R36 have been observed elsewhere[36].

Since Upf2 has been shown to be associated with polysomes and 80S ribosomes[13; 42; 43], we compared the distributions of wild-type Upf2 (full-length and [1–360] domain) and those of the three Upf2 mutants in polysome profiles (Fig. S4A). We did not observe any differences in Upf2 localization between the wild-type and mutant proteins, further indicating that the mutant proteins are properly folded and that NMD inhibition is not due to defects in the recruitment of these Upf2 mutants to the ribosome. Furthermore, we observed that the Upf2 [1–360] domain is sufficient for association to polysomes (Fig. S4B).

Collectively, these experiments have identified highly conserved residues from Upf2 helices h1 and h3 that are essential for NMD activity. These residues are located within a highly conserved patch, which also encompasses four residues (Asp31, Ser32, Lys35, and Arg36) previously shown to be important for NMD[36].

Interactions with DEAD-box helicases by the ScUpf2 mIF4G-1 conserved region

Members of the HEAT-repeat protein family often act as scaffolds that recruit different protein partners. This is particularly true for proteins encompassing mIF4G domains, such as the translation initiation factor eIF4G[44], the nuclear cap-binding protein CBP80[45; 46], the mRNA export factor Gle1[47], the splicing factor CWC22[48], the Not1 subunit of the Ccr4-Not deadenylase[49], the SLBP-binding protein SLIP1[50], and Upf2[20]. For instance, the human UPF2 mIF4G-3 domain, which is structurally related to the ScUpf2 mIF4G-1 domain (rmsd of 3.3 Å over 181 Cα atoms and 11% sequence identity), interacts with the RRM domain from Upf3[20]. Since crystal structures of these proteins and their bound partners were available, we compared the structures of some of these complexes with our structure of the ScUpf2 mIF4G-1 domain (see Fig. S5 and Fig. 3A). This analysis revealed that the UPF3

RRM domain, Caf1, and the SBM motif from Dbp5 bind at the center of the convex face on top of the mIF4G domain (UPF2 mIF4G-3, Not1, and SLIP1 mIF4G domains, respectively; Fig. S5A–C). CBP20 also interacts in the center of the mIF4G domain from CBP80, but on the opposite side of this domain relative to the UPF3 binding site on the UPF2 mIF4G-3 domain (compare panels A and D from Fig. S5). The regions from the ScUpf2 mIF4G-1 domain that structurally match with these interacting regions from the above mentioned complexes are not conserved within Upf2 orthologues, indicating that the Upf2 mIF4G-1 domain is unlikely to recruit partners through these regions. Interestingly, the eIF4A, eIF4AIII and Dbp5 helicases interact through their RecA2 domains with the N-terminal part of the mIF4G domain from eIF4G, CWC22 and Gle1, respectively. Superimposition of the structure of eIF4A–eIF4G (Fig. 3A), eIF4AIII CWC22[48] and Dbp5-Gle1 complexes onto the ScUpf2 mIF4G-1 domain reveals that the eIF4G, CWC22 and Gle1 residues contacting respectively eIF4A, eIF4AIII and Dbp5 RecA2 domains are located within the helices that structurally match ScUpf2 helices h1 to h3, *i.e.*, they correspond to the functionally important conserved region of the ScUpf2 mIF4G-1 domain.

Based on these observations, we immunoprecipitated full-length HA-tagged wild-type or mutant Upf2 from *S. cerevisiae* cell lysates and identified the associated proteins by mass spectrometry (Tables 2 and S2). As expected, immunoprecipitation of wild-type Upf2 yielded significant enrichment of Upf1 and Upf3 (44-fold and 14-fold, respectively, Table S2). Additionally, proteins involved in ribosome biogenesis and assembly represented a major class of proteins associated with full-length Upf2 (Table S2). Our immunoprecipitations did not, however, recover Hrp1, a protein that Wang et al. had identified as a Upf2-interactor[36]. Surprisingly, a comparison of the sets of proteins co-immunoprecipitated with wild-type or mutant Upf2 revealed that Upf2's association with Upf3 was impaired in the K67A/E71A mutant (Table 2A). The interaction between Upf2 and Upf3 is known to be mediated via the C-terminus of Upf2. However, it is possible that introduction of the K67/E71A mutation in the N-terminus can alter the stability of this interaction by a mechanism that is not clear. Two proteins were also found to be specifically associated with wild-type Upf2, but not with the three mutant proteins (Table 2B): Dbp6, an essential putative ATP-dependent DEAD-box RNA helicase required for 60S ribosomal subunit assembly as well as Rpc19, a subunit of both RNA polymerases I and III. Given the propensity of mIFG4 domains to interact with DEAD-box helicases, we focused on Dbp6 and used an *in vitro* binding assay to confirm the Upf2 interaction with Dbp6 (Fig. 3B). Purified GST-Dbp6 (from *E. coli*) and yeast lysates expressing HA-Upf2 were incubated and subjected to immunoprecipitation using anti-HA beads. Purified GST (from Sigma) and a yeast strain expressing the HA tag alone were used as controls for background binding. Analysis of the immunoprecipitates showed co-immunoprecipitation of GST-Dbp6 with wild-type HA-Upf2, consistent with the mass spectrometry data (Fig. 3B). Even though the Upf2 N-terminal mutants did not manifest co-immunoprecipitation of Dbp6 by mass spectrometry (data not shown), GST-Dbp6 was found to co-immunoprecipitate with the N-terminal mutant proteins *in vitro*, albeit at lower abundance (Fig. 3B). Furthermore, the association of wild-type Upf2 or its variants with Dbp6, was still observed after RNase treatment, implying that these interactions are unlikely to be RNA-mediated (data not shown). We further tested whether Dbp6 has a readily detectable role in NMD. However,

when compared to otherwise wild-type cells, those harboring a disruption of the *DBP6* gene did not change the levels of the *CYH2* pre-mRNA, suggesting that any function Dbp6 might have in NMD is not rate-limiting for mRNA decay (data not shown).

DISCUSSION

Despite extensive studies within the last two decades, the mechanism by which NMD distinguishes nonsense-containing mRNAs and activates their accelerated decay remains elusive[2]. The Upf1, Upf2, and Upf3 proteins are central to NMD and, they interact with each other[17], the ribosome[13; 42], the eRF1 and eRF3 translation termination factors[9; 11; 51; 52; 53], and the mRNA decay machinery[10; 14]. Upf1 plays a pivotal role in NMD and its ATPase and RNA helicase activities are regulated by phosphorylation and dephosphorylation events in metazoa[24; 25], as well as by intra- and intermolecular interactions[15; 34; 54; 55], some of which are mediated by Upf2. Beyond its role in activating Upf1, Upf2 also serves as a scaffolding protein that bridges Upf1 and Upf3[17]. While it is clear that the C-terminal half of Upf2 is responsible for interaction with Upf1 and Upf3, a limited number of residues located at the N-terminal part of the Upf2 protein have also been shown to be important for NMD activity[36], but their functional significance has been unknown.

Here, we have analyzed the structure and function of the N-terminal mIF4G-1 domain of *S. cerevisiae* Upf2 and identified a highly conserved region on helices h1 and h3. Not only is this domain essential for NMD[36] (Fig. 2B), but its function may well be independent of the activities of Upf1 and Upf3. This conclusion comes from observations that this domain does not encompass the known binding sites for Upf1 and Upf3[17; 20; 35], and exhibits polyribosomal and 40S ribosomal subunit association independently of the remainder of Upf2 (Fig. S4B–C). A comparison of our Upf2 mIF4G-1 structure with other mIF4G domains indicated that this domain may exert part of its activity by interacting with protein partners (Fig. 3A and Fig. S5). Co-immunoprecipitation studies coupled to mass spectrometry analyses revealed that the DEAD box helicase Dbp6 and the Rpc19 protein interact with wild-type Upf2, but were incapable of interacting with any of the Upf2 mutant proteins. *In vitro* binding assays using purified Dbp6 indicated weakened interactions with the three mutant Upf2 proteins, suggesting that the respective mutated residues might contribute to the structure of an interaction site.

The Dbp6 DEAD-box helicase is of particular interest for several reasons. First, mIF4G domains are known to interact with and regulate the activity of DEAD-box helicases[44; 47; 48; 50; 56]. Second, the mIF4G region involved in helicase binding structurally matches the highly conserved patch from Upf2 mIF4G-1 (Fig. 3A), which we have shown to be important for NMD (Fig. 2). Third, beyond the NMD-specific Upf1 helicase, other helicases have been reported to participate in NMD or to interact with NMD factors. Dbp2, a DEAD-box helicase which, like Dbp6 is involved in ribosome biogenesis, manifests a two-hybrid interaction with Upf1[10] and functions in the NMD pathway[57]. Recently, p68/DDX5, the human Dbp2 orthologue, was shown to interact with human UPF3 and to enhance the NMD-dependent degradation of its own mRNA, as well as mRNAs encoding DDX17 (another DEAD-box helicase highly similar to DDX5) and SMG-5, which are all characterized by long 3′-UTR regions[58]. Furthermore, the C-terminal fragment from human

UPF2 (encompassing the mIF4G-3 domain and the region involved in UPF1-CH binding) was shown to interact with eIF4AI in a yeast two-hybrid assay[59]. Collectively, these results are consistent with the notions that mIFG4 domains are scaffold proteins that appear to have intrinsic DEAD-box helicase binding properties and that the RNA surveillance and ribosome biogenesis pathways have complex interrelationships[60]. Based on these observations we hypothesized that the physical interaction between Upf2 and Dbp6 might have a functional implication. However, we did not observe any effect of Dbp6 disruption on NMD. Further work is needed to understand the biological significance of Upf2-Dbp6 interaction.

Recently, a structure derived by cryo-electron microscopy revealed the first molecular architecture of the human UPF-EJC complex, with UPF2 adopting a ring-like structure compatible with being a central scaffold that interacts with UPF1 and UPF3[61]. In this complex, the UPF2 ring structure is closed by the proximity of the UPF2 mIF4G-1 domain and the subcomplex formed by the UPF1 CH and UPF2 C-terminal domains[35]. Mass spectrometry analysis of different glutaraldehyde cross-linked human UPF-EJC complexes identified UPF2 peptides (residues 112–137, 146–163, 334–348 and 370–379, corresponding to 1–16, 27–29, 210–227 and 249–258 from ScUpf2) that are protected upon UPF1 binding. These results suggest that the UPF2 N-terminal domain might contact the UPF1 CH domain directly, a conclusion supported by earlier co-immunoprecipitation studies demonstrating that deletion of residues 94–133 from human UPF2 weakened its interaction with UPF1[22]. Melero and colleagues were able to fit known coordinates of the EJC, UPF1-UPF2, and UPF3-UPF2 mIF4G-3 complexes, as well as low homology models for the UPF2 mIF4G-1 and mIF4G-2 domains, into their cryo-EM map[61]. More recently, Clerici et al have docked their new mIF4G1 and mIF4G2 structures from human UPF2 into this cryo-EM map[19]. This revealed that these two mIF4G domains are responsible for the positioning of the EJC in the UPF-EJC complex, in agreement with their proposed scaffolding role. This interaction was shown to be dispensable for triggering NMD, which is not surprising since the EJC or its equivalent has not been identified in *S. cerevisiae* and this complex is also not required for splicing-dependent NMD in *S. pombe*[62]. However, in this model, residues from UPF2 mIF4G-1 domain that are protected from glutaraldehyde cross-linking upon UPF1 binding are not oriented towards UPF1 but face the solvent. Hence, we have reconsidered this model, taking into account: 1) the flexibility of the mIF4G-1 domain and in particular of the long h9 and h10 helices (Fig. 1C); and 2) the information derived from mass spectrometry analysis of different glutaraldehyde cross-linked human UPF-EJC complexes as constraints. In our fit, which exhibits a correlation coefficient of 0.85 (Fig. 4A), the UPF2 mIF4G-1 domain is rotated by almost 180° compared to the fit proposed by Clerici et al (Fig. 4B). This illustrates the difficulty to fit with certainty coordinates of protein domains into medium resolution cryo-EM maps (16 Å). In our model, the UPF2 mIF4G-1 domain contacts the UPF1 CH-UPF2 C-terminal complex through helix hA while the long helices h9 and h10 are oriented towards the UPF2 mIF4G-2 domain and the EJC. The interaction of the Upf2 mIF4G-1 domain with Upf1 is supported in yeast as we have observed that the N-terminal fragment of Upf2 encompassing amino acids 1-360 could not associate with polysomes in the absence of Upf1 (Fig. S6), suggesting that the Upf2 N-terminus may come in contact with Upf1 at the interface of the ribosome at a PTC. Importantly, in our model the highly conserved patch in the Upf2 mIF4G-1 domain is solvent exposed and fully accessible

for possible interactions with additional factors (Fig. 4A). This model, and unanticipated observations such as the failure to recover Upf3 in co-immunoprecipitation experiments with the *upf2* K67A/E71A mutant, clearly implies that multiple steps of complex mRNP remodeling accompany NMD.

In conclusion, the crystal structure of the first mIF4G domain from *S. cerevisiae* Upf2 has led to the identification of a highly conserved patch that is important for NMD in yeast. This region appears to be involved in interactions with protein partners, including Dbp6, a DEAD-box helicase. The functional importance of this interaction remains to be established, but connections between other RNA surveillance pathways and ribosome biogenesis have recently been highlighted[60]. Collectively, the importance of all these N-terminal Upf2 residues for NMD function indicates that Upf2 may well have an important mechanistic role in NMD beyond serving as an activator of Upf1 activity and a bridge between Upf1 and Upf3[15; 17; 20; 34; 35; 36; 61].

MATERIAL AND METHODS

Cloning, expression and purification of Upf2 proteins

The sequence encoding C-terminally His-tagged Upf2 [1–310] was amplified from yeast *Saccharomyces cerevisiae* S288C genomic DNA with oligonucleotides oMG27/oMG32 (see Table S1) and inserted into the pET21-a vector to yield plasmid pMG567. The DNA sequence encoding C-terminally His-tagged Upf2 [1–360] from *Saccharomyces cerevisiae* was synthesized chemically by GenScript and cloned into pET21-a, yielding plasmid pMG464. The sequence encoding C-terminally His-tagged Upf2 [1–820] was amplified from yeast *Saccharomyces cerevisiae* S288C genomic DNA with oligonucleotides oMG27/oMG31 (see Table S1) and inserted into pET21-a to yield plasmid pMG422.

The Upf2 [1–310] domain was expressed in *E. coli* BL21(DE3) Codon$^+$ (Novagen) in 1L of 2YT medium supplemented with ampicillin at 100 μg/mL and chloramphenicol at 25 μg/mL. At OD_{600}=0.8, protein expression was induced at 20°C for 20h by adding 50μg/mL IPTG. Cells were harvested by centrifugation and resuspended in 30 mL of Buffer A (20 mM Tris HCl pH 7.5, 200 mM NaCl). Cell lysis was performed by sonication. This domain was purified on a NiNTA column (Qiagen), and then on a Superdex™ 200 16/60 size exclusion column (GE Healthcare) in buffer A (20 mM Tris HCl pH 7.5, 200 mM NaCl).

The Upf2 [1–360] domain was expressed using the same conditions as the Upf2 [1–310] domain. The protein was purified on a Ni-NTA column (Qiagen), followed by a heparin column (GE Healthcare), and finally on a Superdex™ 200 16/60 size exclusion column (GE Healthcare) in buffer B (20 mM Sodium Acetate pH 5.6, 200 mM NaCl).

The Upf2 [1–820] domain was expressed in *E. coli* BL21(DE3) (Novagen) in 1L of 2YT medium supplemented with ampicillin at 100 μg/mL. At OD_{600}=0.8, protein expression was induced at 20°C for 20h by adding 50μg/mL IPTG. The protein was purified on a Ni-NTA column (Qiagen), and then on a Superdex™ 200 16/60 size exclusion column (GE Healthcare) in buffer A.

Crystallization, structure determination and refinement

The purified Upf2 [1–310] domain was concentrated to 20 mg/mL and crystallization trials were carried out in 200 nL drops with a 1:1 protein/precipitant ratio using a Cartesian crystallization robot (Genomics Solutions). Crystals were obtained within 24h to 48h at 18°C. Crystal 1 was obtained in 100mM Tris-HCl pH 8.5 and 30% PEG 400. Crystal 2 was obtained in 100 mM NaCl, 100 mM Bicine pH9 and 30% PEG MME 550. Both crystals were directly flash-frozen in liquid nitrogen prior to data collection.

Diffraction datasets were collected on beamline Proxima-1 at Synchrotron SOLEIL (Saint-Aubin, France). Statistics for data collection are given in Table 1. The structure of the Upf2 [1–310] domain from *S. cerevisiae* could not be solved by molecular replacement (due to the lack of a good homology model), anomalous diffusion (only one Methionine residue present in the sequence of this N-terminal domain), or isomorphous replacement (despite extensive trials). To solve this structure, we then used the ARCIMBOLDO program[37], which is particularly suited for proteins with a high alpha-helical content and high resolution data, two requirements fulfilled by the ScUpf2 mIF4G-1 domain.

Diffraction data in the space group $P2_1$ was available from two isomorphous crystals, diffracting to high resolution limits of 1.65 Å and 1.45 Å, respectively. Data was merged despite a somewhat high R(int) of 22% in order to maximize completeness. The structure was solved *ab initio* with the program ARCIMBOLDO[37] that combines fragment location with PHASER[38] and density modification and auto-tracing with SHELXE[39] in a supercomputing frame, distributing calculations over a condor grid[63] with a pool of 240 cores at the supercomputer CALENDULA in Leon, Spain (http://www.fcsc.es). An ideal helix of 16 alanines was used as a search model for molecular replacement. Locating two such helices proved sufficient to achieve phasing from one of the partial solutions. Further pursued phasing, iterative density modification and main chain auto-tracing with SHELXE improved the starting correlation coefficient from 7.9% to 41.6% and expanded the initial, optimized 31 amino acids to 226 of the 310 residues of the final model (Fig. S1). This subsequently allowed the tracing of missing residues including side chain building and refinement of the entire structure. This model was completed by iterative cycles of manual rebuilding using the molecular modeling program COOT[64] followed by refinement with the program PHENIX[65] in the 40-1.65 Å resolution range. In the final model, the Upf2 regions ^{1}Met-Pro^{24}, ^{29}Lys-Leu^{310} and five His residues from the C-terminal His_6 tag could be modeled in the electron density as well as 167 water molecules and one chloride ion. Structure refinement parameters and geometry are listed in Table 1. The atomic coordinates and structure factors have been deposited into the PDB under the accession code 4LUN.

Flexible fitting of Upf2 mIF4G-1 domain in cryo-EM maps of UPF-EJC complex

To further investigate Upf2 mIF4G-1 domain function in the context of the UPF-EJC complex, a flexible fitting approach was performed in the 16 Å resolution cryo-EM map of the human UPF-EJC complex[61]. First, PDB models of the complexes of UPF1 core domain (helicase + CH) bound to the C-terminus of UPF2[35], UPF3b-EJC complex[26], UPF2 mIF4G-3 domain in complex with the RRM domain of UPF3[20], and the recently deposited models of human mIF4G1 and mIF4G2[19] were independently manually docked in the

electron density of the human UPF1-UPF2-UPF3-EJC complex (EM databank code: 2048;[61]) using CHIMERA[66]. A rigid body fitting of each domain was then applied using CHIMERA[66]. This initial full complex model complies with the model already proposed by Melero et al[61]. Then, the human mIF4G-1 domain structure solved recently was automatically docked in the electron density obtained by cryo-EM with the SITUS package[67]. The same approach was used with the yeast mIF4G-1 domain model obtained in the present study to take into account the possible flexibility of helices h9 and h10 in mIF4G-1 domains. The different solutions obtained were carefully checked to determine if the domain orientation was compatible with the position of the other domains in the complex. In parallel, information from previous mass spectrometry experiments[61] was used as a constraint to sort the different solutions. One of the docking solutions of the yeast model helped to better position the human mIF4G-1 model. Finally, the reconstructed complete human model docked in the electron density was further refined using iModfit[68], an algorithm that can perform a flexible fitting approach with conserved secondary structure elements. The final model fitted in the human UPF1-UPF2-UPF3-EJC complex electron density map is characterized by a correlation coefficient of 0.85.

Assessment of NMD activity by northern blotting

Plasmids harboring individual HA-tagged or genomic *UPF2* alleles were introduced into a *upf2Δ* strain (HFY115) and the resulting cells were grown in synthetic complete (SC) medium lacking leucine. The ability of each *upf2* allele to promote NMD was determined by analyzing the steady-state levels of the nonsense-containing *CYH2* pre-mRNA and the *ade2-1* mRNA. Total RNA isolation and northern analysis were performed as described previously[10]. Random primed DNA probes made from the 0.6 Kb EcoRI-HindIII *CYH2* fragment, the 2.0 Kb BgIII-BgIII *ADE2* fragment, or the 0.5 Kb EcoRI-EcoRI *SCR1* fragment were used to detect the *CYH2* pre-mRNA, *ade2-1* mRNA, and the *SCR1* RNA, respectively. Transcript-specific signals were determined with a FUJI BAS-2500 analyzer.

Western blotting

Cells representing 4 OD_{600} units were harvested and resuspended in 200 μL of sample buffer as described previously[69]. Samples were resolved by 8% SDS-PAGE, transferred to Immobilon-P (Millipore) membranes, and incubated with anti-HA antibody (Sigma). Blots were stripped and reprobed with monoclonal anti-Pgk1 antibody (Molecular Probes).

Polysome analysis

Cells were grown at 30°C in SC medium lacking leucine to an OD_{600} of 0.7. Cell extracts were prepared in Buffer C (10 mM Tris pH 7.4, 100 mM NaCl, 30 mM $MgCl_2$, 50μg/ml cycloheximide, 200μg/ml heparin) by glass bead lysis and fractionated on 7%–47% sucrose gradients. Protein was precipitated from individual fractions with trichloroacetic acid (TCA), resuspended in Laemmli buffer, and subjected to western blot analysis using anti-HA antibody (Sigma) and anti-Rpl3 antibody (a kind gift from J. Warner).

Ribosomal subunit purification and analysis

Ribosomal subunits were prepared as described previously[13]. Briefly, cell extracts from strains expressing Upf2 and its derivatives were centrifuged at 120,000g for 10 h at 4°C on 34 mL 15%–40% sucrose gradients that contained 10 mM Tris-HCl, pH 7.4, 10 mM $MgCl_2$, 30 mM NH_4Cl, and 1 mM DTT. Fractions corresponding to the 40S and 60S peaks were concentrated separately in Amicon filters (EMD Millipore) and were analyzed by western blotting using anti-HA antibody (Sigma), anti-Rps6 antibody (Cell Signaling), and anti-Rpl3 antibody.

Immunoprecipitation

Cell extracts were prepared from strains harboring HA-tagged Upf2 and its derivatives. Cells (100 mL) were grown at 30°C in SC medium lacking leucine to an OD_{600} of 0.8. Cells were collected by centrifugation, washed with ice-cold water, and resuspended in 1 mL of IP buffer (Pierce). Cells were broken mechanically with glass beads and 100 μL of extract was incubated with 20 μL of anti-HA magnetic beads (Pierce) for 1 hour at 4°C. The beads were washed 3 times with IP buffer containing 150 mM NaCl and washed once with cold water. Bound proteins were eluted under reducing condition as per the manufacturer's instructions (Pierce). Samples were analyzed on SDS-PAGE followed by western blotting using anti-HA antibody (Sigma).

In vitro binding assay

The sequence encoding N-terminally GST-tagged Dbp6 was amplified from yeast *Saccharomyces cerevisiae* S288C genomic DNA with oligonucleotides oMG63/oMG56 (see Table S1) and inserted into the pGEX-6P1 vector using BamHI/XhoI restriction enzymes to yield plasmid pMG601. GST-Dbp6 was expressed in *E. coli* BL21-codon+ (Novagen) in 1L of 2YT medium supplemented with Ampicillin at 100 μg/mL and Chloramphenicol at 25 μg/mL. At OD_{600}=0.8, protein expression was induced at 20°C for 20h by adding 50μg/mL IPTG. Cells were harvested by centrifugation and resuspended in 40 ml of Buffer A supplemented with protease inhibitors and DNase A. Cell lysis was performed by sonication. The protein was purified on a glutathione-Sepharose resin (GE Healthcare), followed by a HiTRAP monoQ column (GE Healthcare) and then a Superdex 200 10–30 size exclusion column.

Cell extracts were prepared from yeast strains harboring HA-tagged Upf2 and its derivatives. Cells (100mL) were grown at 30°C in SC medium lacking leucine to an OD_{600} of 0.8, collected by centrifugation, washed with ice-cold water, and resuspended in 1 mL of IP buffer (Pierce). Cells were broken mechanically with glass beads and 100 μL of extract was incubated with purified GST-Dbp6 for 1 hour at 4°C. Binding was followed by incubation with anti-HA magnetic beads (Pierce) for 1 hour at 4°C. The beads were washed 3 times with IP buffer containing 150 mM NaCl and washed once with cold water. Bound proteins were eluted under reducing condition as per the manufacturer's instructions (Pierce). Samples were analyzed on SDS-PAGE followed by western blotting using anti-HA antibody (Sigma) or anti-GST antibody (Sigma).

Mass spectrometry

For mass spectrometry, immunoprecipitated, reduced samples were resolved on SDS-PAGE gels for a brief period of time followed by excision of a single gel slice containing all the immunoprecipitated proteins. The gel slice containing the sample was subjected to in-gel trypsin digestion. Chromatographed peptides were then identified on an LTQ Orbitrap Velos mass spectrometer and were searched against the Mascot server (Matrix Science, London; version 2.4.0). Comparative data analyses and filtering were performed using Scaffold 4 (Proteome Software).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are indebted to Manuela Argentini and David Cornu (SICaps, IMAGIF platform, Gif/Yvette, France) and John Leszyk (UMass Medical School Proteomics and Mass Spectrometry Facility) for mass spectrometry and to Vonny Caroline for technical assistance. We thank SOLEIL for provision of synchrotron radiation facilities and Andrew Thompson for assistance with beamline Proxima-1. We acknowledge computing time at the FCSCL. We are indebted to Dr Christiane Schaffitzel and coworkers for sharing with us the coordinates of the domains from human UPF and EJC proteins fitted into the cryo-EM map.

FUNDING

This work was supported by the Centre National pour la Recherche Scientifique ATIP-AVENIR program [to M.G.], the Agence Nationale pour la Recherche [grant ANR-06-BLAN-0075-02 to H.vT., S.K. and M.G.], the Association Française contre les Myopathies [to H.vT., S.K. and M.G.], the Fondation pour la Recherche Médicale [to Z.F.], the European Union Sixth Framework program “3D-Repertoire” [LSHG-CT-2005-512028, to H.vT. and M.G.], the National Institutes of Health [R37 GM27757 to A.J.]; the Ministerio de Economía y Competitividad [BFU2012-35367 to I.U.], the Centro para el Desarrollo Tecnológico Industrial [IDC-2010-1173 to I.U.] and the Human Frontiers Science Program [grant RGP0018, to H.vT, S.K., A.J. and M.G.].

References

1. Graille M, Seraphin B. Surveillance pathways rescuing eukaryotic ribosomes lost in translation. Nat Rev Mol Cell Biol. 2012; 13:727–35. [PubMed: 23072885]
2. Kervestin S, Jacobson A. NMD: a multifaceted response to premature translational termination. Nat Rev Mol Cell Biol. 2012; 13:700–12. [PubMed: 23072888]
3. Shoemaker CJ, Green R. Translation drives mRNA quality control. Nat Struct Mol Biol. 2012; 19:594–601. [PubMed: 22664987]
4. Tsuboi T, Kuroha K, Kudo K, Makino S, Inoue E, Kashima I, Inada T. Dom34:hbs1 plays a general role in quality-control systems by dissociation of a stalled ribosome at the 3′ end of aberrant mRNA. Mol Cell. 2012; 46:518–29. [PubMed: 22503425]
5. Behm-Ansmant I, Kashima I, Rehwinkel J, Sauliere J, Wittkopp N, Izaurralde E. mRNA quality control: an ancient machinery recognizes and degrades mRNAs with nonsense codons. FEBS Lett. 2007; 581:2845–53. [PubMed: 17531985]
6. Cui Y, Hagan KW, Zhang S, Peltz SW. Identification and characterization of genes that are required for the accelerated degradation of mRNAs containing a premature translational termination codon. Genes Dev. 1995; 9:423–36. [PubMed: 7883167]
7. Leeds P, Peltz SW, Jacobson A, Culbertson MR. The product of the yeast UPF1 gene is required for rapid turnover of mRNAs containing a premature translational termination codon. Genes Dev. 1991; 5:2303–14. [PubMed: 1748286]

8. Cho H, Kim KM, Kim YK. Human proline-rich nuclear receptor coregulatory protein 2 mediates an interaction between mRNA surveillance machinery and decapping complex. Mol Cell. 2009; 33:75–86. [PubMed: 19150429]
9. Czaplinski K, Ruiz-Echevarria MJ, Paushkin SV, Han X, Weng Y, Perlick HA, Dietz HC, Ter-Avanesyan MD, Peltz SW. The surveillance complex interacts with the translation release factors to enhance termination and degrade aberrant mRNAs. Genes Dev. 1998; 12:1665–77. [PubMed: 9620853]
10. He F, Jacobson A. Identification of a novel component of the nonsense-mediated mRNA decay pathway by use of an interacting protein screen. Genes Dev. 1995; 9:437–54. [PubMed: 7883168]
11. Ivanov PV, Gehring NH, Kunz JB, Hentze MW, Kulozik AE. Interactions between UPF1, eRFs, PABP and the exon junction complex suggest an integrated model for mammalian NMD pathways. EMBO J. 2008; 27:736–47. [PubMed: 18256688]
12. Lejeune F, Li X, Maquat LE. Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylating, and exonucleolytic activities. Mol Cell. 2003; 12:675–87. [PubMed: 14527413]
13. Min EE, Roy B, Amrani N, He F, Jacobson A. Yeast Upf1 CH domain interacts with Rps26 of the 40S ribosomal subunit. RNA. 2013
14. Swisher KD, Parker R. Interactions between Upf1 and the decapping factors Edc3 and Pat1 in Saccharomyces cerevisiae. PLoS One. 2011; 6:e26547. [PubMed: 22065998]
15. 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. 2008; 15:85–93. [PubMed: 18066079]
16. Czaplinski K, Weng Y, Hagan KW, Peltz SW. Purification and characterization of the Upf1 protein: a factor involved in translation and mRNA degradation. Rna. 1995; 1:610–23. [PubMed: 7489520]
17. He F, Brown AH, Jacobson A. Upf1p, Nmd2p, and Upf3p are interacting components of the yeast nonsense-mediated mRNA decay pathway. Mol Cell Biol. 1997; 17:1580–94. [PubMed: 9032286]
18. Aravind L, Koonin EV. Eukaryote-specific domains in translation initiation factors: implications for translation regulation and evolution of the translation system. Genome Res. 2000; 10:1172–84. [PubMed: 10958635]
19. Clerici M, Deniaud A, Boehm V, Gehring NH, Schaffitzel C, Cusack S. Structural and functional analysis of the three MIF4G domains of nonsense-mediated decay factor UPF2. Nucleic Acids Res. 2014; 42:2673–86. [PubMed: 24271394]
20. 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. 2004; 11:330–7. [PubMed: 15004547]
21. Ponting CP. Novel eIF4G domain homologues linking mRNA translation with nonsense-mediated mRNA decay. Trends Biochem Sci. 2000; 25:423–6. [PubMed: 10973054]
22. Serin G, Gersappe A, Black JD, Aronoff R, Maquat LE. Identification and characterization of human orthologues to Saccharomyces cerevisiae Upf2 protein and Upf3 protein (Caenorhabditis elegans SMG-4). Mol Cell Biol. 2001; 21:209–23. [PubMed: 11113196]
23. Grimson A, O'Connor S, Newman CL, Anderson P. SMG-1 is a phosphatidylinositol kinase-related protein kinase required for nonsense-mediated mRNA Decay in Caenorhabditis elegans. Mol Cell Biol. 2004; 24:7483–90. [PubMed: 15314158]
24. Ohnishi T, Yamashita A, Kashima I, Schell T, Anders KR, Grimson A, Hachiya T, Hentze MW, Anderson P, Ohno S. Phosphorylation of hUPF1 induces formation of mRNA surveillance complexes containing hSMG-5 and hSMG-7. Mol Cell. 2003; 12:1187–200. [PubMed: 14636577]
25. Yamashita A, Izumi N, Kashima I, Ohnishi T, Saari B, Katsuhata Y, Muramatsu R, Morita T, Iwamatsu A, Hachiya T, Kurata R, Hirano H, Anderson P, Ohno S. 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. 2009; 23:1091–105. [PubMed: 19417104]
26. Buchwald G, Ebert J, Basquin C, Sauliere J, Jayachandran U, Bono F, Le Hir H, Conti E. 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. 2010; 107:10050–5. [PubMed: 20479275]

27. Le Hir H, Gatfield D, Izaurralde E, Moore MJ. The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. EMBO J. 2001; 20:4987–97. [PubMed: 11532962]
28. Amrani N, Sachs MS, Jacobson A. Early nonsense: mRNA decay solves a translational problem. Nat Rev Mol Cell Biol. 2006; 7:415–25. [PubMed: 16723977]
29. 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. 2007; 26:1591–601. [PubMed: 17318186]
30. Buhler M, Steiner S, Mohn F, Paillusson A, Muhlemann O. EJC-independent degradation of nonsense immunoglobulin-mu mRNA depends on 3′ UTR length. Nat Struct Mol Biol. 2006; 13:462–4. [PubMed: 16622410]
31. Gonzalez CI, Ruiz-Echevarria MJ, Vasudevan S, Henry MF, Peltz SW. The yeast hnRNP-like protein Hrp1/Nab4 marks a transcript for nonsense-mediated mRNA decay. Mol Cell. 2000; 5:489–99. [PubMed: 10882134]
32. Isken O, Maquat LE. Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. Genes Dev. 2007; 21:1833–56. [PubMed: 17671086]
33. Nicholson P, Muhlemann O. Cutting the nonsense: the degradation of PTC-containing mRNAs. Biochem Soc Trans. 2010; 38:1615–20. [PubMed: 21118136]
34. Chakrabarti S, Jayachandran U, Bonneau F, Fiorini F, Basquin C, Domcke S, Le Hir H, Conti E. Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. Mol Cell. 2011; 41:693–703. [PubMed: 21419344]
35. Clerici M, Mourao A, Gutsche I, Gehring NH, Hentze MW, Kulozik A, Kadlec J, Sattler M, Cusack S. Unusual bipartite mode of interaction between the nonsense-mediated decay factors, UPF1 and UPF2. EMBO J. 2009; 28:2293–306. [PubMed: 19556969]
36. Wang W, Cajigas IJ, Peltz SW, Wilkinson MF, Gonzalez CI. Role for Upf2p phosphorylation in Saccharomyces cerevisiae nonsense-mediated mRNA decay. Mol Cell Biol. 2006; 26:3390–400. [PubMed: 16611983]
37. Rodriguez DD, Grosse C, Himmel S, Gonzalez C, de Ilarduya IM, Becker S, Sheldrick GM, Uson I. Crystallographic ab initio protein structure solution below atomic resolution. Nat Methods. 2009; 6:651–3. [PubMed: 19684596]
38. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007; 40:658–674. [PubMed: 19461840]
39. Sheldrick GM. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. Acta Crystallogr D Biol Crystallogr. 2010; 66:479–85. [PubMed: 20383001]
40. Andrade MA, Petosa C, O'Donoghue SI, Muller CW, Bork P. Comparison of ARM and HEAT protein repeats. J Mol Biol. 2001; 309:1–18. [PubMed: 11491282]
41. Nozawa K, Ishitani R, Yoshihisa T, Sato M, Arisaka F, Kanamaru S, Dohmae N, Mangroo D, Senger B, Becker HD, Nureki O. Crystal structure of Cex1p reveals the mechanism of tRNA trafficking between nucleus and cytoplasm. Nucleic Acids Res. 2013; 41:3901–14. [PubMed: 23396276]
42. Atkin AL, Schenkman LR, Eastham M, Dahlseid JN, Lelivelt MJ, Culbertson MR. Relationship between yeast polyribosomes and Upf proteins required for nonsense mRNA decay. J Biol Chem. 1997; 272:22163–72. [PubMed: 9268361]
43. Ghosh S, Ganesan R, Amrani N, Jacobson A. Translational competence of ribosomes released from a premature termination codon is modulated by NMD factors. RNA. 2010; 16:1832–47. [PubMed: 20675403]
44. Schutz P, Bumann M, Oberholzer AE, Bieniossek C, Trachsel H, Altmann M, Baumann U. Crystal structure of the yeast eIF4A–eIF4G complex: an RNA-helicase controlled by protein-protein interactions. Proc Natl Acad Sci U S A. 2008; 105:9564–9. [PubMed: 18606994]
45. Dias SM, Wilson KF, Rojas KS, Ambrosio AL, Cerione RA. The molecular basis for the regulation of the cap-binding complex by the importins. Nat Struct Mol Biol. 2009; 16:930–7. [PubMed: 19668212]
46. Mazza C, Ohno M, Segref A, Mattaj IW, Cusack S. Crystal structure of the human nuclear cap binding complex. Mol Cell. 2001; 8:383–96. [PubMed: 11545740]

47. Montpetit B, Thomsen ND, Helmke KJ, Seeliger MA, Berger JM, Weis K. A conserved mechanism of DEAD-box ATPase activation by nucleoporins and InsP6 in mRNA export. Nature. 2011; 472:238–42. [PubMed: 21441902]
48. Buchwald G, Schussler S, Basquin C, Le Hir H, Conti E. Crystal structure of the human eIF4AIII-CWC22 complex shows how a DEAD-box protein is inhibited by a MIF4G domain. Proc Natl Acad Sci U S A. 2013; 110:E4611–8. [PubMed: 24218557]
49. Basquin J, Roudko VV, Rode M, Basquin C, Seraphin B, Conti E. Architecture of the nuclease module of the yeast Ccr4-not complex: the Not1-Caf1-Ccr4 interaction. Mol Cell. 2012; 48:207–18. [PubMed: 22959269]
50. von Moeller H, Lerner R, Ricciardi A, Basquin C, Marzluff WF, Conti E. Structural and biochemical studies of SLIP1–SLBP identify DBP5 and eIF3g as SLIP1-binding proteins. Nucleic Acids Research. 2013
51. Kashima I, Yamashita A, Izumi N, Kataoka N, Morishita R, Hoshino S, Ohno M, Dreyfuss G, Ohno S. 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. 2006; 20:355–67. [PubMed: 16452507]
52. 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. 2008; 6:e111. [PubMed: 18447585]
53. Wang W, Czaplinski K, Rao Y, Peltz SW. The role of Upf proteins in modulating the translation read-through of nonsense-containing transcripts. EMBO J. 2001; 20:880–90. [PubMed: 11179232]
54. Cheng Z, Muhlrad D, Lim MK, Parker R, Song H. Structural and functional insights into the human Upf1 helicase core. EMBO J. 2007; 26:253–64. [PubMed: 17159905]
55. Fukuhara N, Ebert J, Unterholzner L, Lindner D, Izaurralde E, Conti E. SMG7 is a 14-3-3-like adaptor in the nonsense-mediated mRNA decay pathway. Mol Cell. 2005; 17:537–47. [PubMed: 15721257]
56. Young CL, Khoshnevis S, Karbstein K. Cofactor-dependent specificity of a DEAD-box protein. Proc Natl Acad Sci U S A. 2013; 110:E2668–76. [PubMed: 23630256]
57. Bond AT, Mangus DA, He F, Jacobson A. Absence of Dbp2p alters both nonsense-mediated mRNA decay and rRNA processing. Mol Cell Biol. 2001; 21:7366–79. [PubMed: 11585918]
58. Geissler V, Altmeyer S, Stein B, Uhlmann-Schiffler H, Stahl H. The RNA helicase Ddx5/p68 binds to hUpf3 and enhances NMD of Ddx17/p72 and Smg5 mRNA. Nucleic Acids Res. 2013
59. Mendell JT, Medghalchi SM, Lake RG, Noensie EN, Dietz HC. Novel Upf2p orthologues suggest a functional link between translation initiation and nonsense surveillance complexes. Mol Cell Biol. 2000; 20:8944–57. [PubMed: 11073994]
60. Strunk BS, Novak MN, Young CL, Karbstein K. A translation-like cycle is a quality control checkpoint for maturing 40S ribosome subunits. Cell. 2012; 150:111–21. [PubMed: 22770215]
61. Melero R, Buchwald G, Castano R, Raabe M, Gil D, Lazaro M, Urlaub H, Conti E, Llorca O. The cryo-EM structure of the UPF-EJC complex shows UPF1 poised toward the RNA 3′ end. Nat Struct Mol Biol. 2012; 19:498–505. S1–2. [PubMed: 22522823]
62. Wen J, Brogna S. Splicing-dependent NMD does not require the EJC in Schizosaccharomyces pombe. EMBO J. 2010; 29:1537–51. [PubMed: 20360683]
63. Tannenbaum, T.; Wright, D.; Miller, K.; Livny, M. Condor - A Distributed Job Scheduler. In: Sterling, T., editor. Beowulf Cluster Computing with Linux. 2002.
64. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004; 60:2126–32. [PubMed: 15572765]
65. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC. PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr. 2002; 58:1948–54. [PubMed: 12393927]
66. Goddard TD, Huang CC, Ferrin TE. Visualizing density maps with UCSF Chimera. J Struct Biol. 2007; 157:281–7. [PubMed: 16963278]

67. Wriggers W, Milligan RA, McCammon JA. Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. J Struct Biol. 1999; 125:185–95. [PubMed: 10222274]

68. Lopez-Blanco JR, Garzon JI, Chacon P. iMod: multipurpose normal mode analysis in internal coordinates. Bioinformatics. 2011; 27:2843–50. [PubMed: 21873636]

69. Johansson MJ, Jacobson A. Nonsense-mediated mRNA decay maintains translational fidelity by limiting magnesium uptake. Genes Dev. 2010; 24:1491–5. [PubMed: 20634315]

HIGHLIGHTS

- What is the role of Upf2 N-terminal domain in nonsense mediated mRNA decay pathway?
- Determination of the structure of mIF4G-1 domain from yeast Upf2.
- Upf2 mIF4G-1 domain harbours a highly conserved region mandatory for NMD.
- This region seems to be involved in protein recruitment.

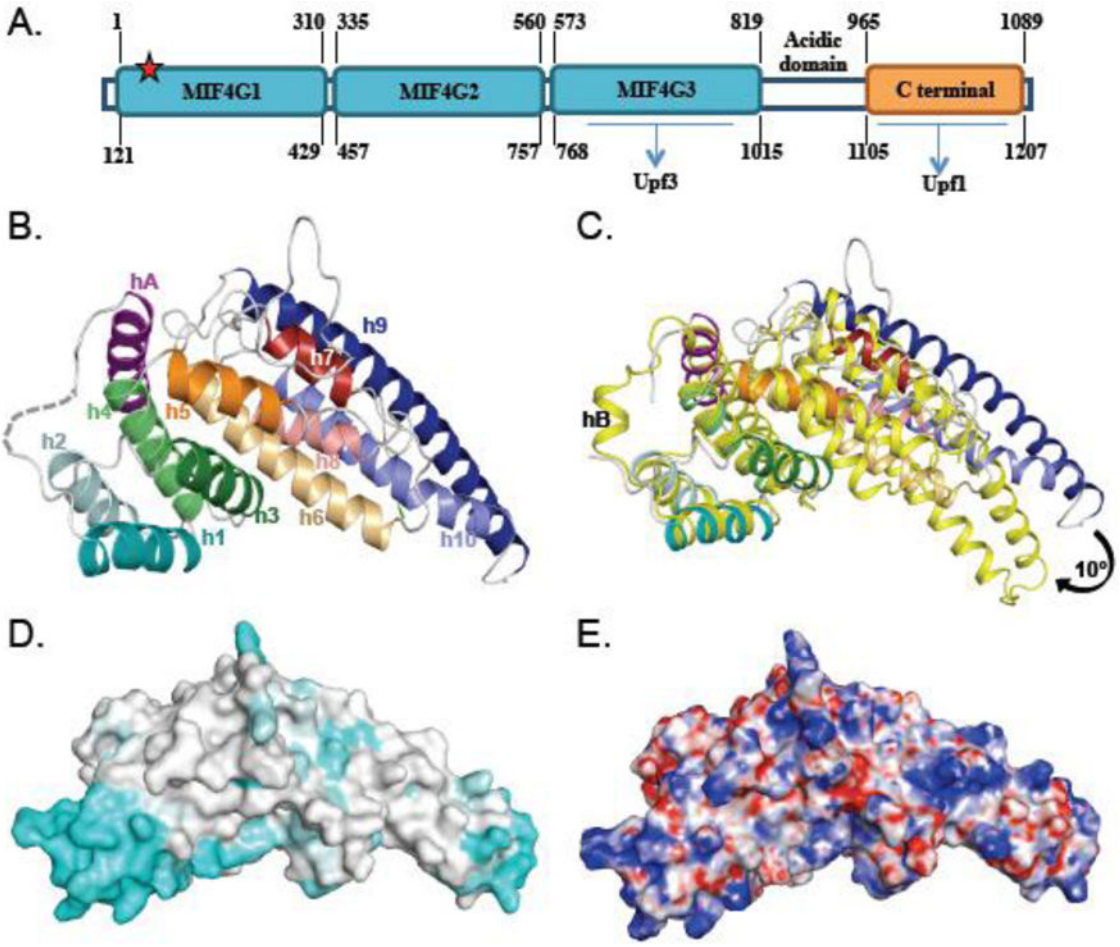


Figure 1. Structure of the *S. cerevisiae* Upf2 mIF4G-1 domain

A. Schematic representation of the Upf2 protein with domain boundaries from yeast or human proteins indicated above or below, respectively. Human UPF2 domain limits are those described by Clerici et al[19]. Approximate location of mutated residues is indicated by a red star. Low complexity regions are indicated by lines. Structured domains are depicted by squares. B. Ribbon representation of Upf2 mIF4G-1 domain crystal structure. Each antiparallel helical hairpin is represented by different colors. The first helix of the hairpin is depicted with dark color while the second helix is in light color. The missing loop connecting helices hA to h1 is depicted by a grey dashed line. C. Superimposition of human UPF2 (yellow) and ScUpf2 (same color code as panel B) mIF4G-1 structures. Only the N-terminal half of these two domains was considered for superimposition. D. Mapping of the sequence conservation at the surface of the ScUpf2 mIF4G-1 domain. Coloring is from grey (low conservation) to cyan (highly conserved). The conservation score was calculated using the CONSURF server[70] and using an alignment made from the sequences of 20 Upf2 orthologues. E. Mapping of the electrostatic potential at the surface of the Upf2 mIF4G-1 domain. Positively (15 k_BT/e^-) and negatively (−15 k_BT/e^-) charged regions are colored in blue and red, respectively. The electrostatic potential was calculated using the PBEQ Solver server[71]. The same orientation is used for all panels.

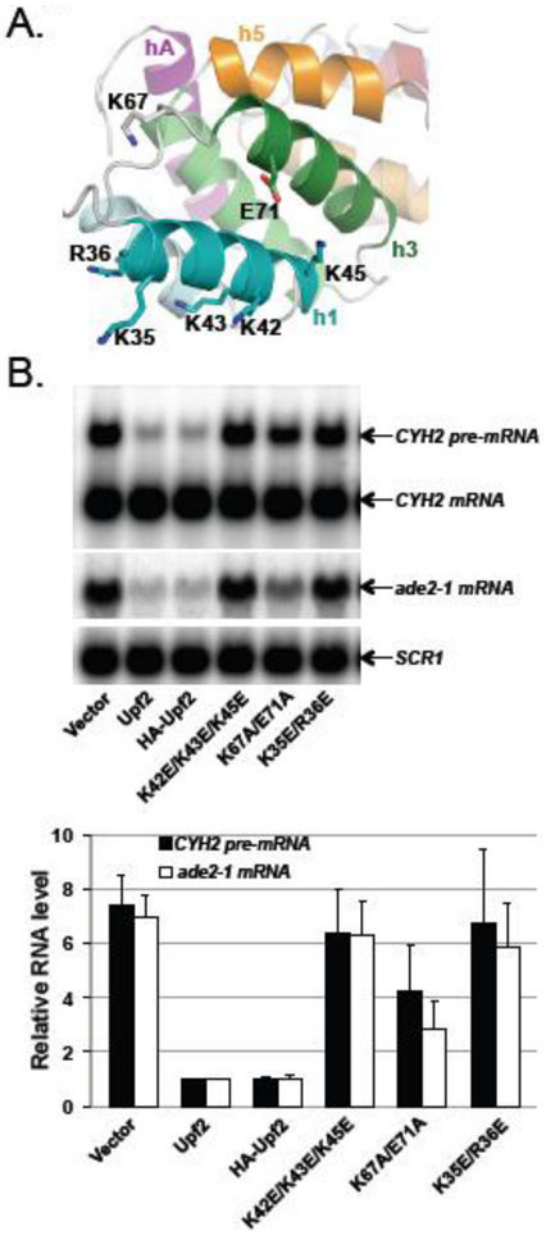


Figure 2. Effect of Upf2 mutants on NMD

A. Residues mutated in this study are shown as sticks on the structure of Upf2 mIF4G-1 domain. The same color code as Figure 1B is used.

B. Plasmids (pRS315) carrying different *UPF2* alleles were individually transformed into *upf2Δ* cells (HFY115) and the resulting strains were analyzed for NMD activity. Total RNA was isolated from each strain and the steady-state levels of the *CYH2* pre-mRNA and the *ade2-1* mRNA were analyzed by northern blotting, using random-primed probes specific for *CYH2*, *ADE2* or *SCR1* transcripts, with the latter serving as a loading control. The relative *CYH2* pre-mRNA and *ade2-1* mRNA levels in each of the *upf2* mutant strains compared to the strain harboring the genomic wild-type *UPF2* gene were determined. Data shown in the

graph were the average of three independent experiments. Mutants were generated from HA-tagged *UPF2*.

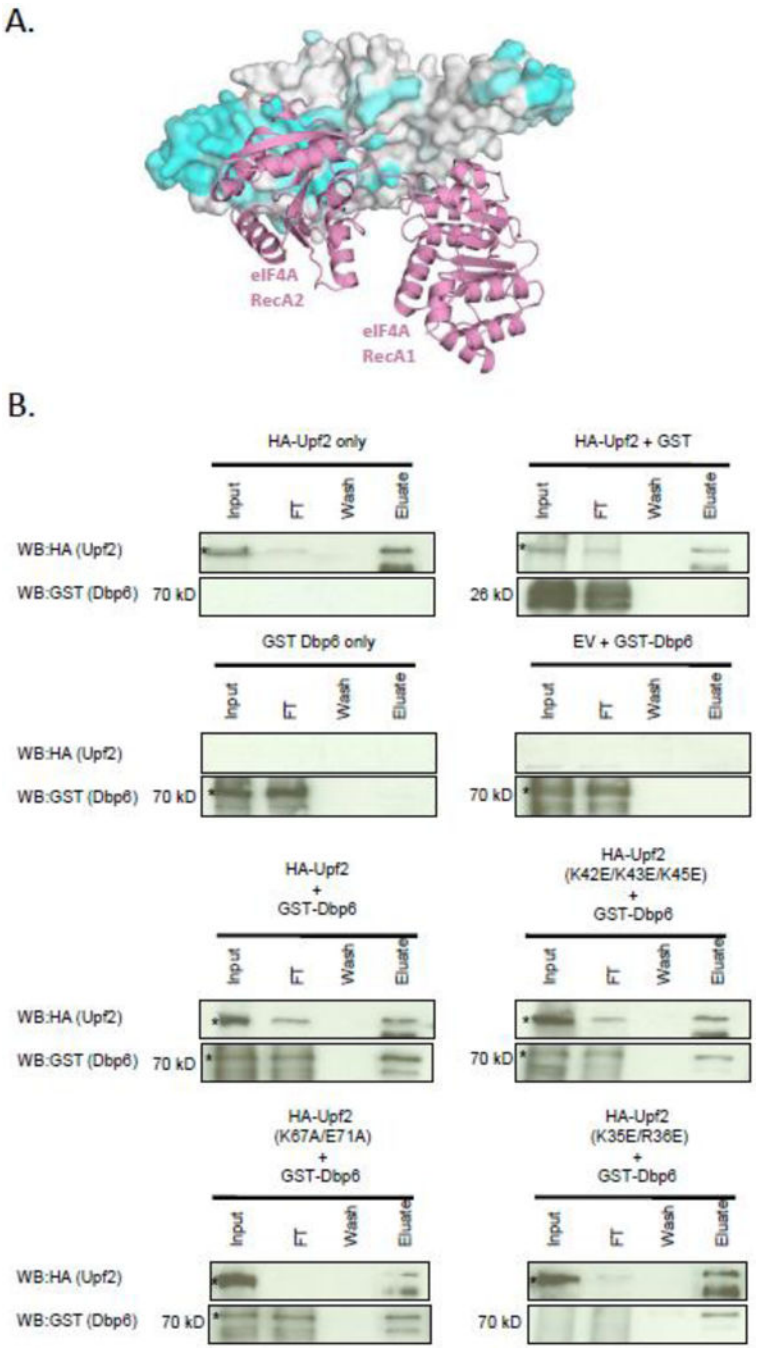


Figure 3. Upf2 interacts with DEAD-box helicase Dbp6 *in vitro*

A. Structure of the eIF4G–eIF4A (pink) complex superimposed onto the structure of the ScUpf2 mIF4G-1 domain. The surface of the ScUpf2 mIF4G-1 domain is colored by conservation score as in Fig. 1D. For clarity, the eIF4G protein is not shown. B. Western blot analyses of anti-HA immunoprecipitation assays utilized anti-HA and anti-GST antibodies on input, flow-through (FT), wash, and eluate fractions. *In vitro* binding assays were performed with GST-DBP6 purified from *E. coli* and yeast cell lysates expressing HA-Upf2

and its variants. Asterisk denotes the HA-Upf2 band. The lower band denotes a cross-reacting protein in the HA-Upf2 eluate.

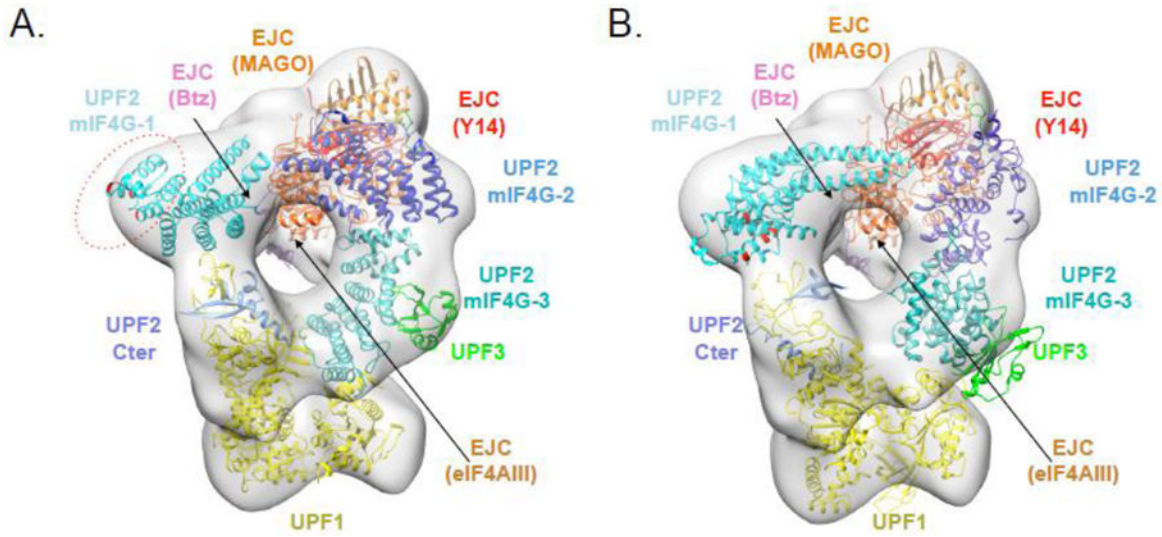


Figure 4. Models of the UPF-EJC complex

A. Fit of the independent structures in the UPF-EJC cryo-EM map (16 Å resolution) generated for this study. The cryo-EM map (contoured at a threshold of 1.5) is colored in grey and shown in transparency. The different domains from UPF2 are indicated with different blue colors. The conserved and functionally important region from the UPF2 mIF4G-1 domain is circled by red dotted lines. Residues corresponding to positions mutated in this study are colored in red. A CHIMERA session allowing the visualization of this fit into the cryo-EM map can be downloaded as supplementary files associated to this paper. B. Representation of the fit proposed by Clerici et al[19]. Same orientation, threshold for cryo-EM map and color code as panel A.

Table 1

Data collection and refinement statistics

	Crystal 1	Crystal 2
Data collection		
Space group	$P2_1$	$P2_1$
Cell dimensions	a=42 Å, b=51.6 Å, c=72.5; β=100°	a=41.9 Å, b=51 Å, c=72.1 Å, β=99.9°
Resolution (Å)	38-1.45 (1.53-1.45)	50-1.65 (1.75-1.65)
Total number of reflections	175,373	107,467
Total number of unique reflections	53,617	36,059
R_{sym} [a]	0.091 (0.467)	0.052 (0.626)
I/σ/	5.6 (2)	14.6 (2.3)
Completeness (%)	99 (98.4)	98.4 (95.9)
Redundancy	3.3	3
Refinement		
Resolution (Å)		40-1.65
R_{work}/R_{free} (%) [b]		19.1/23
R.m.s. deviations		
Bond lengths (Å)		0.006
Bond angles (°)		0.993

[a] $R_{sym}=\Sigma_h\Sigma_i|I_{hi} - \langle I_h\rangle|/\Sigma_h\Sigma_i I_{hi}$, were I_{hi} is the i^{th} observation of the reflection h, while $\langle I_h\rangle$ is the mean intensity of reflection h.

[b] $R_{factor}=\Sigma ||F_o| - |F_c||/|F_o|$. R_{free} was calculated with a small fraction (5%) of randomly selected reflections.

Table 2

Mass spectrometric analysis of proteins that co-immunoprecipitate differentially between wild-type HA-Upf2 and the three HA-Upf2 mutants

A. Co-immunoprecipitation of Upf1 and Upf3 with Upf2

	HA-Upf2 (Wild type)	HA-Upf2 (K42E/K43E/K45E)	HA-Upf2 (K67A/E71A)	HA-Upf2 (K35E/K36E)
Upf1	**44**	**41**	**37**	**30**
Upf3	**14**	**12**	**0**	**7**

B. Proteins co-immunoprecipitating with wild-type Upf2 but not the three Upf2 mutant proteins.

Molecular Function	Protein	Systematic name	Average fold enrichment	Molecular weight
Ribosome biogenesis and assembly	Dbp6	YNR038W	5.2	70,361
Transcription	Rpc19	YNL113W	2.6	16,151

C. Proteins co-immunoprecipitating only with the three Upf2 mutant proteins but not wild-type Upf2.

Molecular Function	Protein	Systematic name	Average fold enrichment	Molecular weight
Ribosome biogenesis and assembly	Dbp8	YHR169W	7.3	47,878

Proteins enriched at least 2 fold over HA-only (control) in three independent experiments have been listed. The average fold enrichment of HA-Upf2 was observed to be 77 fold. In A, numbers denote the fold enrichment of Upf1 or Upf3 over HA-only (control).