"Nonsense-mediated mRNA decay involves two distinct Upf1-bound complexes" by M. Dehecq *et al*.

Contents

Supplementary Figures :

Appendix Fig. S1 - Workflow for quantitative analysis of MS/MS results from affinity purified complexes.

Appendix Fig. S2 - Controls of total tagged protein levels in the presence or absence of other NMD components.

Appendix Fig. S3 - Alignment of hSmg6, hSmg5, hSmg7, Ebs1 and Nmd4 domains sequences.

Appendix Fig. S4 - Effect of NMD4 and EBS1 in the destabilization of an NMD substrate by the Upf1-HD-Cter domain.

Appendix Fig. S5 - Comparison between the canonical SURF/DECID model and our extended yeast-based model.

Appendix Fig. S6 - Similarities and differences between yeast Nmd4 PIN and PIN domains of human SMG6 and SMG5.

Supplementary Tables:

Table S1 - Summary of the number of replicates, the number of proteins robustly quantified and the efficiency of RNase treatment for each purification type.

Table S2 - RNASeq raw data analysis results summary.
Table S3 - Strains used in this study
Table S4 - Plasmids used in this study
Table S5 - Oligonucleotides used in this study

Access to primary data (mass-spectrometry and RNASeq):
The accession number for the sequencing data reported in this paper is **GEO: GSE102099**.
The MS proteomics data that support the findings of this study have been deposited in the ProteomeXchange Consortium via the PRIDE repository (Vizcaíno et al, 2016) with the dataset identifier **PXD007159**.
In addition to the access to raw data, enrichment and intensity data are available as an interactive web application at: https://hub05.hosting.pasteur.fr/NMD_complexes.

Appendix Material and Methods

Media and growth conditions

Yeast cells were grown in YPD (20g·L^{-1} glucose, 10g·L^{-1} yeast extract, 20g·L^{-1} bacto-peptone, 20g·L^{-1} bacto-agar for plates only) and in synthetic media without uracil to select transformants and maintain URA3 plasmids. All yeast strains were freshly thawed from frozen stocks and grown at 30°C. Bacterial strains were grown in LB media, supplemented with antibiotics when necessary, at 37°C.

Yeast strains construction

Yeast strains used in this paper are listed in **Appendix Table S3**. C-terminal TAP-tagged strains originated from the collection of systematically built strain (Ghaemmaghami *et al*, 2003). For some strains, we modified the classical TAP tag (CBP-TEV cleavage site-protein A) and added an additional 6-His tag between CBP and the TEV cleavage site. To do that we have used the pCRBlunt-CRAP(6-HisTAP) plasmid that can replace the original tag by homologous recombination. TAP or CRAP (CBP-6His-TEV-proteinA) tagged versions of proteins were used for our experiments and gave similar results; both versions bear the "TAP" name in the text, as only the protein A part of the tag and the TEV cleavage site were used in our experiments.

Plasmids construction

For Gibson assembly, we mixed 30 to 50ng of linearized vector with 60 to 300ng of amplified PCR insert(s) in 200µL microtubes containing 10µL of 2x Hot Fusion Buffer: 2x pre-assembly buffer 5x (0.5M Tris pH 7.5, 50mM MgCl2, 1mM each dNTP, 50mM DTT, 25% PEG-8000) with 0.0075u·µL−1 of T5 exonuclease and 0.05u·µL^{-1} of Phusion Hot Start DNA polymerase. Distilled water was added to the tube for a final volume of 20µL. Tubes were incubated in a thermocycler for 1 hour at 50°C, then slowly ramped down to 20°C in 5 minutes (0.1°C per second), and held at 10°C. The Hot Fusion reaction was used for transformation or stored at -20°C if not used immediately. Otherwise, 1µL of the reaction was transformed in *E. coli* NEB 10-beta competent cells. Insertion of *UPF1* fragments into vector was verified by restriction enzyme digestion and sequencing.

Coding sequences of yeast full length *NMD4*, yeast *UPF1* helicase domain, and human *UPF1* helicase domains (Uniprot accession codes Q12129, P30771, and Q92900-2 respectively) were cloned between NheI/NotI, NheI/XhoI, and NdeI/XhoI in variants of pET28a (Novagen). In these vectors the NcoI–NdeI cassette was either deleted by mutating the NcoI site to an NdeI site or replaced by the coding sequence for a His-tag or a CBP tag followed by a TEV protease cleavage site. In the vector without the NcoI-NdeI cassette, a TEV protease cleavage site was engineered in front of the C-terminal His-tag. Fragments of human *UPF1* (helicase domain, 195-914), yeast *UPF1* (helicase domain, residues 220–851), and *NMD4* (Full length, residues 1–883) were amplified by PCR with the appropriate restriction sites, using oligonucleotide pairs

HLH725/HLH726, HLH2705/HLH2706 and MD99/MD100 respectively. PCR products were purified using PCR cleanup Qiaquick (*NMD4*) or Promega (human and yeast *UPF1*) kits, digested for 1 hour at 37°C using NEB Cutsmart buffer and the corresponding enzyme couples, then further gel purified using Quiaquick kit (*NMD4*) or Promega gel purification kit (human and yeast *UPF1*). Plasmids pHL5 and pHL4 were digested in parallel using NheI/NotI (pHL5), NheI/XhoI (pHL4) or NdeI/XhoI (pHL4) and purified on 0.8% agarose gels. Digested PCR products were mixed with corresponding plasmids in a 1:3 molar ratio and ligated overnight at 16°C in a 15µl reaction using T4 DNA ligase. Ligase was heat-inactivated 10 minutes at 65°C, then ligation products were used to transform *E. coli* NEB 5-alpha competent cells. Insertion of *NMD4* and *UPF1* fragments into vectors was verified by restriction enzyme digestion and sequencing.

Total protein extracts and western blots

Total protein extracts were prepared from 5 A_{600} of exponential culture with a fast method using alkaline treatment (Kushnirov, 2000). Cells were incubated in 200µL of 0.1M NaOH for 5 min at room temperature, collected by 3 min centrifugation and resuspended in 50µL of sample buffer containing DTT (0.1M). Proteins were denatured for 3 min at 95°C, and cellular debris were pelleted by centrifugation. 10µL of supernatant or diluted supernatant (for quantification scale) were loaded on acrylamide NuPAGE Novex 4-12% Bis-Tris gels (Life technologies). After transfer to nitrocellulose membrane with a semi-dry fast system (Biorad trans-blot) with discontinuous buffer (BioRad technote 2134), proteins were detected by hybridization with appropriate antibodies.

RNA extraction

For RT-qPCR and RNAseq, cells were first grown in YPD to log phase and collected. Total RNA was extracted using the hot phenol extraction method and precipitated using ammonium acetate and ethanol. The extracted RNA samples were treated with DNase I (Ambion TURBO DNA-free kit) before reverse-transcription (RT) and library preparation.

Affinity purification for RNA analysis

We used 4L yeast culture at exponential phase, A_{600} 0.6 to 0.8 and processed with the same method as TAP immuno-purification but with more precaution to work fast, on ice and in an RNase free environment. In addition, buffers were freshly prepared, lysis buffer contained 5mM $MgCl_2$ to maintain mRNP integrity, washing steps were reduced to three washes with HKI + DTT buffer and elution was done in HKI + DTT + AcTEV. After elution, RNA extraction was done by 3 steps of acid phenol/chloroform followed by overnight ammonium acetate precipitation at -20°C. RNA pellet were resuspended in water and processed for DNase treatment (Ambion TURBO DNA-free kit) and RT-qPCR analysis.

At the first RNA extraction step, we also collected proteins that were at the interface between aqueous phase (RNA) and organic phase. Proteins were precipitated by adding 100% ethanol,

incubating at -20°C for 1h and centrifugation. Washed and dried protein pellets were used for western blotting to verify purification efficiency.

Dcp2 depletion strain

The Dcp2-degron strain derives from BY4741, except that the *DCP2* ORF is followed by the polyG-miniAID-KANMX-TIR1 cassette (adapted from Nishimura *et al*, 2009; Kubota *et al*, 2013). The addition of the cassette leads to the inducible poly-ubiquitination of Dcp2 protein and its degradation. The cassette has been added by transformation of the BY4741 strain with two PCR products; a first one amplifying the polyG-miniAID part with half of the kanamycin resistance cassette from the plasmid 1451 and the second one amplifying the Os*TIR1* gene and the second half of the kanamycin resistance cassette from the plasmid 1367. The polyG is used as a spacer between the Dcp2 ORF and the miniAID element. The miniAID comes from *A. thaliana* AID element (AtIAA17 amino acids 65–132), it is the auxin/IAA inducing degron element. Os*TIR1* is the plant-specific gene from *Oryza sativa,* in our construction it is transcribed from the same chromosomal region as Dcp2 under the control of the ADE2 promoter. *TIR1* gene product interacts with the miniAID fused to Dcp2 in one hand and with a group of protein including an ubiquitin ligase in the other hand. The ubiquitin ligase will poly-ubiquitinate the miniAID domain triggering the degradation of Dcp2-miniAID by the proteasome.

The strain has been tested by western blot with or without IAA at 100 µM for 1 hour and by growth tests on YPD+/- IAA plates. Depletion of Dcp2 was complete after 1h of auxin (IAA, Indole-3-acetic acid, ref: SERVA 26181.01) treatment. We also observed a strong growth defect after auxin treatment.

For the qPCR experiment, 50mL of YPD were inoculated with LMA4423. Once the culture has reached an optical density of 0.5 we divided the culture in two equal parts, and added IAA at a concentration of 100µM to one of the cultures. After 1 hour the cells were collected by centrifugation, and we proceeded with RNA extraction, DNase treatment and RT-qPCR.

Protein expression and purification from *E. coli*

Plasmids pHL1484 (CBP-*NMD4* full length-6His), pHL1301 (yeast *UPF1* helicase domain-6His) and pHL201 (human *UPF1* helicase domain-6His) were used to transform BL21 (DE3) Codon Plus competent cells. After plating and overnight cell growth on Kanamycin LB plates, 3-4 colonies were inoculated in 25 ml LB media supplemented with Kanamycine (50mg.L^{-1}). After 6 hours of incubation at 37°C, each starter culture was used to inoculate a 1L LB culture supplemented with Kanamycine (50mg.L^{-1}) and Chloramphenicol (34mg.L^{-1}). Cultures were first incubated at 37°C to a 0.6 optical density at 600nm wavelength, then transferred and incubated at 16°C. When optical density reached 0.8, protein expression was induced using IPTG (0.5mM).

After overnight incubation at 16°C, cells were harvested at 6000 rpm for 10 minutes, washed once with cold 1x PBS, then collected at 5000 rpm for 10 minutes.

Cells were mixed in a lysis buffer (1.5x PBS, 20mM Imidazole, 0.1% Igepal, 10% Glycerol, 1mM $MgCl_2$) supplemented with protease inhibitors (Aprotinin 2µg.mL^{-1}, Leupeptin 1µg.mL^{-1}, Pepstatin 1µg.mL^{-1}, PMSF 50µg.mL^{-1}) then lysed using sonication for 4 minutes on ice, and centrifuged at 18000 rpm for 25 minutes at 4°C.

Supernatants were mixed with 500µl Ni-NTA agarose resin (=500µl of 50% slurry, Clontech) pre-equilibrated in lysis buffer, and incubated for 2 hours in 50ml falcons on a rotator at 4°C. Beads were collected and washed with lysis buffer then transferred on Bio-Spin columns (Biorad) pre-washed in lysis buffer. Columns were further washed with lysis buffer followed by a wash buffer (1.5x PBS, 50mM Imidazole, 250mM NaCl, 0.1% Igepal, 10% Glycerol, 1mM $MgCl_2$), before protein elution in 800 ul fractions (1.5x PBS, 150mM Imidazole, 10% Glycerol, 1mM $MgCl_2$).

Nmd4 was further dialyzed against calmodulin binding buffer (1x PBS, 100mM NaCl, 0.1% Igepal, 10% Glycerol, 1mM MgCl2, 4mM CaCl2, 1mM DTT) overnight at 4°C in Spectrapor-4 (12-14 MWCO), then mixed with 500µl Calmodulin Affinity Resin (= 1ml of 50% slurry, Agilent) pre-equilibrated in binding buffer. After 1 hour incubation into a Bio-Spin column on a rotator at 4°C, beads were washed twice with binding buffer, before protein elution (1x PBS, 100mM NaCl, 0.05% Igepal, 10% Glycerol, 1mM $MgCl_2$, 20mM EGTA, 1mM DTT).

Proteins were finally dialyzed against 1.5x PBS, 150mM NaCl, 10% (w/v) glycerol, 1mM DTT and 1mM $MgCl_2$ in Spectrapor-4 (12-14 MWCO) then stored at -80°C.

In vitro **pull-down assays**

Pull-down was performed using preblocked calmodulin affinity beads (Agilent).

Preblocking beads. Briefly, in order to preblock beads, 1 ml calmodulin sepharose beads (50% Slurry) were spun, and resuspended in 20mM Hepes, 500mM NaCl, 0.1% Igepal, 0.08mg.ml^{-1} glycogen carrier, 0.08mg.ml^{-1} tRNA and 0.8mg.ml^{-1} BSA. After 2 hours at 4°C, beads were washed 3 times (20mM Hepes, 150mM NaCl, 0.1% Igepal) then resuspended in 500µl 1x binding buffer 250/10 (20mM Hepes, 250mM NaCl, 0.1% Igepal, 2mM $MgAc_2$, 2mM $CaCl_2$, 10% glycerol, 1mM DTT).

In vitro **pull-down assay**. Recombinant CBP-Nmd4, yeast Upf1 and human Upf1 proteins were thawed on ice. Each mix contained 2µg of each protein in 30µl reaction mixes. NaCl and glycerol concentrations were adjusted to 150mM and 15% respectively. Five microliters (1/6) aliquots were taken out of each mix to load on Input gel. Each mix was further supplemented with 5µl of water, and 30µl of a NaCl/glycerol buffer to reach a 60µl reaction volume containing 125mM NaCl and 12.5 % glycerol.

Mixes were incubated for 20 minutes at 30°C. To perform pull-downs, 12µl of preblocked calmodulin beads were added to each mix, along with 200µl 1x binding buffer 150/10 (20mM Hepes, 150mM NaCl, 0.1% Igepal, 2mM $MgAc_2$, 2mM $CaCl_2$, 10% glycerol, 1mM DTT), then rotated 1 hour at 4°C. Beads were further washed 3 times with 1x binding buffer 150/10, then

dried using a thinned Pasteur pipet. The retained complexes were eluted using 20µl elution buffer (10mM Tris pH=7.5, 150mM NaCl, 1mM $MgAc_2$, 2mM Imidazole, 20mM EGTA, 0.1% Igepal, 14% glycerol, 10mM β-mercaptoethanol) while shaking 5 minutes at 1400 rpm, 30°C.

Eluates were collected after spinning, transferred to fresh tubes, concentrated 30 minutes in a Speed-vac then loaded on 10% SDS-PAGE gels.

Pull-down on streptavidin beads using biotinylated-RNA.

The experiments were performed as previously described for similar assays (Fiorini *et al*, 2012). To preblock beads, 300 ul of Dynabeads® MyOne Strept (Invitrogen, 10 mg/ml) were spun, rinsed (20 mM Hepes, 150 mM NaCL, 0.1% Igepal) and resuspended in 20 mM Hepes, 500 mM NaCl, 0.1% Igepal, 0.08 mg/ml glycogen carrier, 0.08 mg/ml tRNA and 0.8 mg/ml BSA. After 2 hours at 4°C, beads were washed 2 times (20 mM Hepes, 150 mM NaCl, 0.1% Igepal) then resuspended in 300 ul 1x storage buffer (10 mM Hepes, 250 mM NaCl, 1mM EDTA).

Recombinant NMD4 and yeast UPF1 HD proteins were thawed on ice. Each mix contained 2 ug of each protein in 30 µl reaction mixes. NaCl and glycerol concentrations were adjusted to 150 mM and 15% respectively. 5 µl (1/6) aliquots were taken out of each mix to load on Input gel. Each mix was further supplemented with 1 µl biotinylated RNA (1 µl at 10 µM) and 4 µl of water, and 30 µl of a NaCl/glycerol buffer to reach a 60 µl reaction volume containing 125 mM NaCl and 12.5 % glycerol.

Mixes were incubated for 20 minutes at 30°C. To perform pull-downs, 6 ul of preblocked streptavidin beads were added to each mix, along with 200 ul 1x binding buffer 150/10 (20 mM Hepes, 150 mM NaCL, 0.1% Igepal, 2 mM MgAc2, 10% glycerol, 1 mM DTT), then rotated 1 hour at 4°C. Beads were further washed 3 times with 1x binding buffer 150/10 or 300/10 (same composition, but 300 mM NaCl instead 150 mM). Since the beads are magnetic, supernatant is removed during each wash by placing the tubes on a magnetic tube holder.

The complexes retained on the beads via the biotinylated RNA were eluted using 7.5 µl 1x SDS loading dye added directly on the beads. Tubes were flicked to suspend the beads, then spun 1 min at 3000 rpm. Eluates were carefully collected without the beads and transferred to fresh tubes, then boiled 3 min at 94°C. Eluates and input samples were loaded on 10% SDS-PAGE gels.

NMD efficiency calculation details

We assume that RNA synthesis is constant and that the degradation of RNA occurs by two competitive pathways, with different first-order rate constants: k_{NMD} and k_{base}. Thus, the variation in the RNA levels with time could be expressed as:

$$\frac{d[RNA]}{dt} = T - (k_{NMD} + k_{base}) \times [RNA]$$

where:

T is transcription rate (synthesis, and export),

[RNA] is RNA concentration,

k_{NMD} is the global rate constant for degradation through NMD,

k_{base} is the base rate constant for degradation of the RNA, independent of NMD.

At steady state, the change in RNA concentration is null. A given steady-state level can be obtained either by both high transcription and high degradation or low transcription and low degradation. Since most of the time we do not know the transcription rate, or the degradation constants, we need to remove one of the variables. Let us consider two situations in which transcription is supposed invariable, a wt and a mutant strain devoid of NMD:

$$[RNA]_{wt} = \frac{T}{k_{NMD} + k_{base}}$$

$$[RNA]_{mut} = \frac{T}{k_{base}}$$

Thus, the ratio between the mutant that has no NMD and wild type, N, is:

(1) $$\frac{[RNA]_{mut}}{[RNA]_{wt}} = N = \frac{k_{NMD} + k_{base}}{k_{base}} = \frac{k_{NMD}}{k_{base}} + 1$$

If the RNA is not an NMD substrate, k_{NMD} is 0 and the ratio between the mutant and wt becomes 1.

If the RNA is exclusively degraded through NMD, the ratio would be infinite.

At a fractional NMD efficiency, noted α, the degradation through NMD would be $\alpha \times k_{NMD}$ and the ratio of RNA towards wt in this mutant, P, becomes:

(2) $$\frac{[RNA]_{obs}}{[RNA]_{wt}} = P = \frac{k_{base} + k_{NMD}}{k_{base} + \alpha k_{NMD}} = \frac{1 + \frac{k_{NMD}}{k_{base}}}{1 + \alpha \frac{k_{NMD}}{k_{base}}}$$

N is always larger than P.

We can substitute k_{NMD}/k_{base} in (2) with (N-1), from (1):

$$\frac{[RNA]_{obs}}{[RNA]_{wt}} = P = \frac{k_{base} + k_{NMD}}{k_{base} + \alpha k_{NMD}} = \frac{1 + \frac{k_{NMD}}{k_{base}}}{1 + \alpha \frac{k_{NMD}}{k_{base}}}$$

$$\frac{N}{P} = 1 + \alpha (N - 1)$$

Thus the fraction of NMD, used to define efficiency, can be calculated from:

$$\alpha = \frac{\frac{N}{P} - 1}{N - 1}$$

Where N is the enrichment of RNA in the NMD null mutant and P the enrichment in the tested strain.

Example: if the RNA is enriched 8 fold in an NMD mutant and 4 fold in a "partial" NMD strain over WT, the fractional efficiency of NMD would be 1/7, 0.14 or 14%. This example shows that the correlation between observed ratios and NMD efficiency is not linear. The formula maps values obtained by RT-qPCR to a percentage of "NMD efficiency" and works for extreme values. If a mutant has no NMD defect, P will be equal to 1 and α becomes 1. If the mutant is 100% NMD deficient, P will approach N and, α becomes 0. Due to experimental error, P might be superior to N, but in that case α should be considered 0.

Protein sequence alignment and percentage of identity calculation

We used metaPhOr (Pryszcz *et al*, 2011) to obtain sequences of orthologues for the studied proteins. We aligned entire proteins or fragments using Mafft software called as a web service from Jalview. PIN domain boundaries for Smg6, Smg5 have been determined by sequence homology with orthologous proteins from *D. melanogaster*, *C. elegans*, *M. musculus* (positions 1239-1419 for hSmg6, 849-1016 for hSmg5). For Nmd4, we used all the 1-218 amino acid sequence for the alignment, even if the last 60 amino acids are not part of the PIN domain. Identity was computed as the percentage of identical aligned residues over the total number of aligned residues. To strengthen alignment reliability, we also performed delta-BLAST alignment (Boratyn *et al*, 2012) on the *S. cerevisiae* protein database with the hSmg6 and hSmg5 PIN domains and hSmg7 14-3-3 domain as query, using E-values as significance scores.

GO Term analysis and statistics

We used the GO Term Finder tool associated with SGD database (Cherry *et al*, 2012) to search for common function of protein list extracted from MS analysis; for example, the group of proteins removed by RNase treatment and group of proteins of a given enrichment; default settings were used. We calculated the significance of the accumulation of certain protein classes using hypergeometric distribution test.

RNA libraries preparation and sequencing

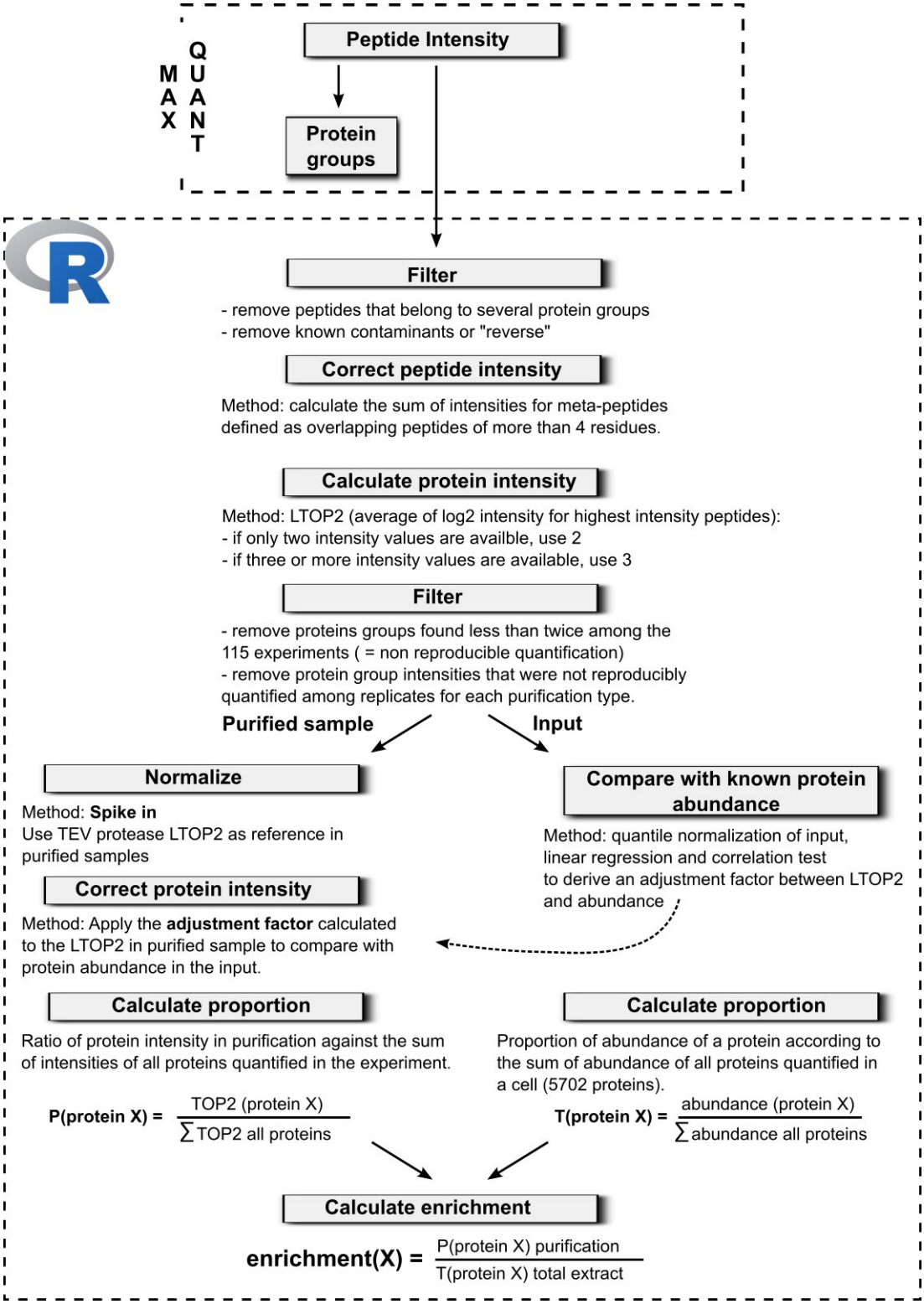
RNA, extracted and treated with DNase, were subjected to a RiboZero treatment to remove ribosomal RNAs. Libraries of mRNA were prepared with the TruSeq Stranded kit (Illumina). Libraries were sequenced on the Illumina HiSeq 2500 for 50 bases (for *upf1*Δ and BY4741) and 65 bases (for *nmd4Δ*, *ebs1Δ*, *nmd4Δ/ebs1Δ* and BY4741). Sequenced reads were aligned to yeast genome version sacCer3 (version R64-2-1) using STAR (Dobin *et al*, 2013) with the default parameters except for -s 0 -0. We used "featureCounts" (Liao *et al*, 2014) from the subread

package (version 1.5.0-p3-Linux-x86_64) to count the number of reads per features. For the features, we used a custom list of coordinates where unstable transcripts, as well as introns could be specifically counted. Three independent experiments were performed for each condition.

RNAseq statistical analysis

Expression of gene in the different samples and replicates were normalized to WT strain using DESeq2 (Love *et al*, 2014). We analyzed the three replicates independently and used the mean for figures. During DESeq2 workflow, we removed features identified by zero reads to avoid problems with the logarithm transformation. We determined NMD substrates from *upf1Δ* RNAseq results and fixed the threshold for these RNA to a minimum of 1.4 fold increase. For the bin analysis, we used a custom script to divide RNA sequenced in *nmd4Δ* and *ebs1Δ* experiments into 5 bins containing the same number of transcript and calculate the percentage of NMD substrates for each bin. Statistical significance of the differences between bins was assessed using a binomial test.

Appendix Figures



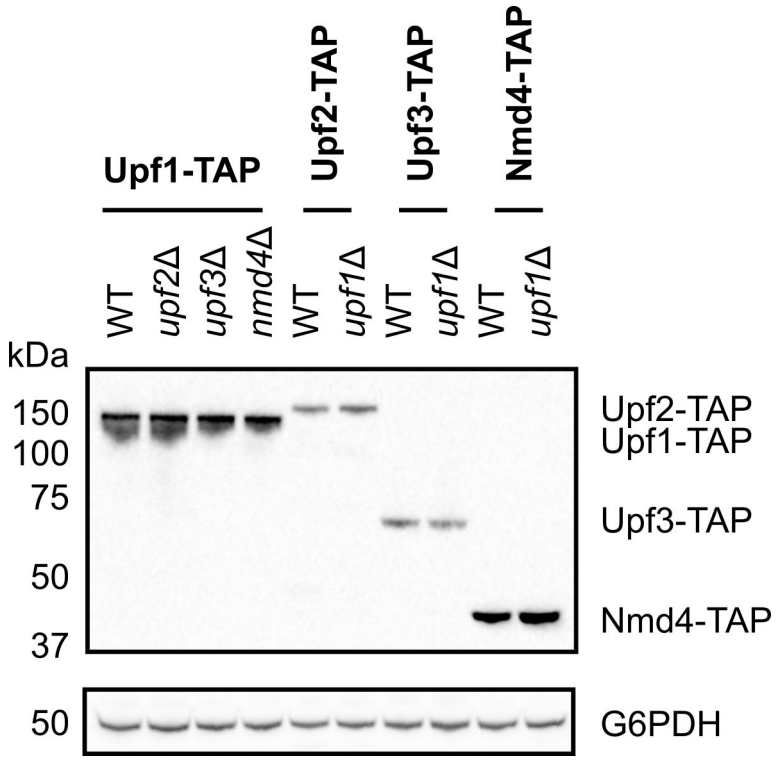
Appendix Fig. S1. Workflow for quantitative analysis of MS/MS results from affinity purified complexes. Notes about the depicted procedure: 1. MaxQuant output for peptide intensities and their association with a given

protein was used as the main input for computing LTOP2 scores and enrichment. To calculate the false discovery rate (FDR) of the MS/MS analysis, MaxQuant builds reverse sequence « artificial » proteins that serve as negative controls for the identification procedure. Reversed sequences and common contaminants (trypsin, keratins) were removed from further analyses in the early steps of the analysis.

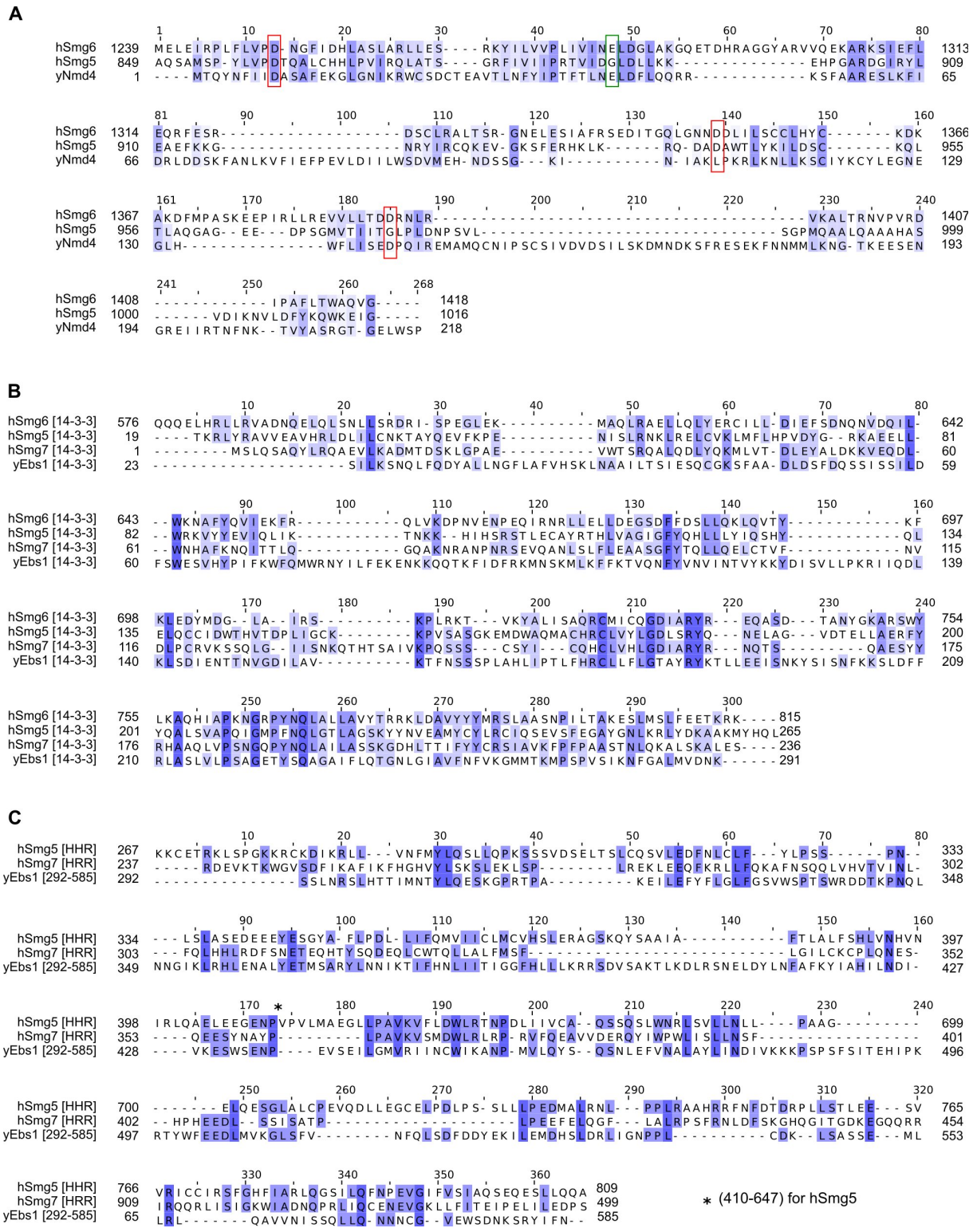
2. A protein group corresponds to a single protein or several proteins with very high sequence similarity that cannot be discriminated by peptide analysis. For further analyses, we used the identity of the protein of the group with most coverage.

3. The TEV protease was added in each purification experiments with the same relative amount to elute complexes from beads. This step is important to be able to compare replicates between them and the different purification types.

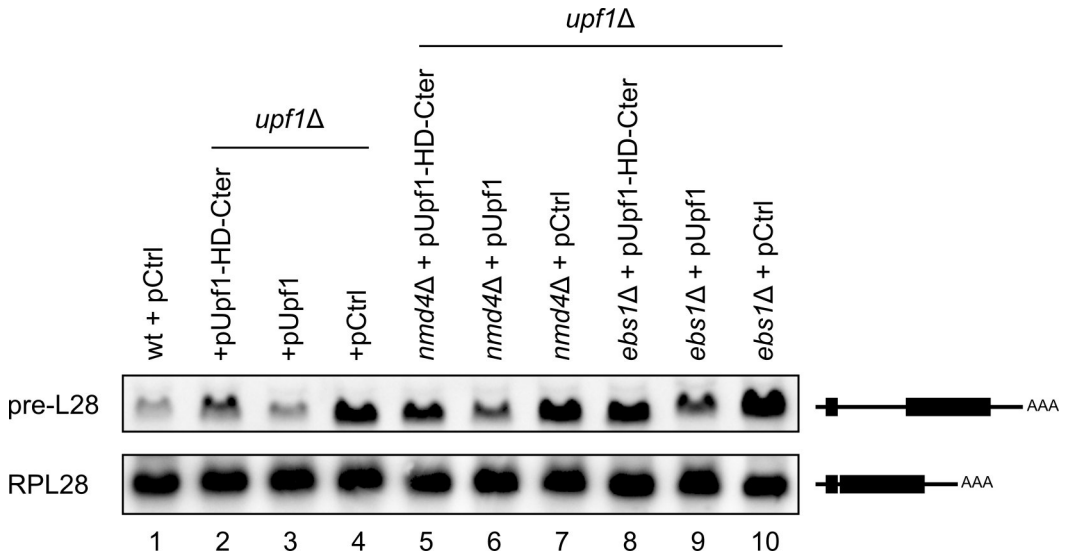
4. The comparison of our input LTOP2 with the abundance data from Ho et al. 2017, allowed to calculate a factor that served to adjust LTOP2 values and make them compatible with the dynamic range and scale of published abundance values.



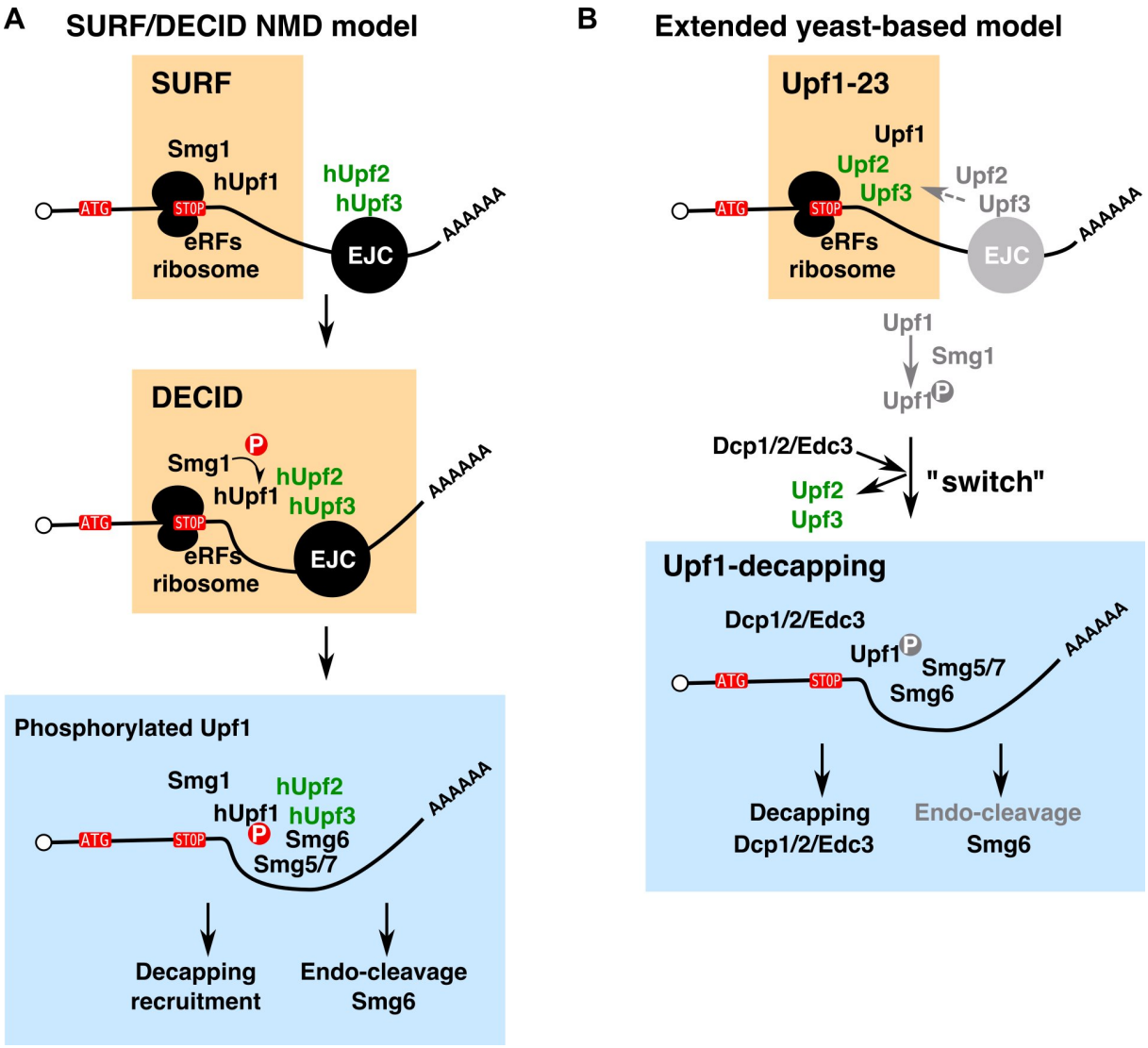
Appendix Fig. S2. Controls of total tagged protein levels in the presence or absence of other NMD components. Total protein extracts from the described strains were tested by immunoblot to detect the protein A part of the TAP tag. G6PDH signal was used as a loading control.



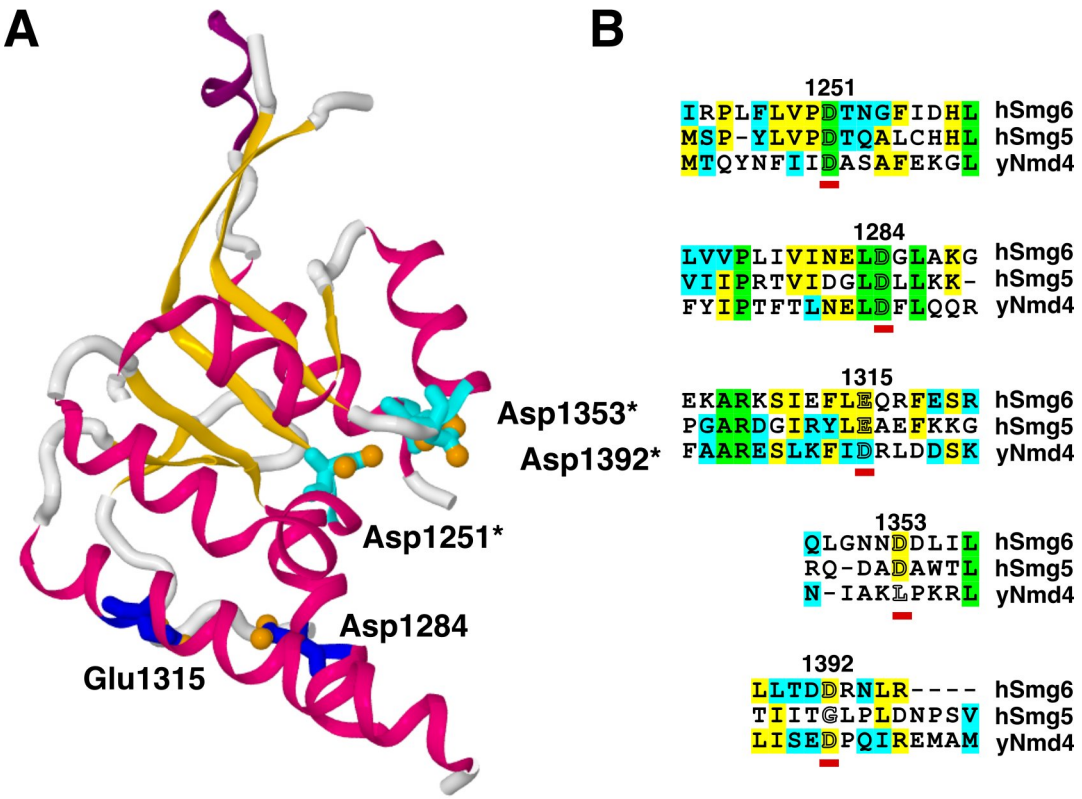
Appendix Fig. S3: Alignment of hSmg6, hSmg5, hSmg7, Ebs1 and Nmd4 domains sequences. Alignment of the PIN domains of hSmg6, hSmg5 and ScNmd4 **(A)**, the 14-3-3 domains of hSmg6, hSmg5, hSmg7 and Ebs1 **(B)** and of the helical hairpin region (HHR) of hSmg5, hSmg7 and Ebs1 (292-585) **(C)**. These alignments were obtained with the algorithm Mafft with default parameters; colour represents percentage of identity or similarity (BLOSUM62).



Appendix Fig. S4: Effect of NMD4 and EBS1 in the destabilization of an NMD substrate by the Upf1-HD-Cter domain. Northern blot estimation of the complementation of *upf1*Δ and *upf1*Δ/*nmd4*Δ and *upf1*Δ/*ebs1*Δ strains with a control plasmid (pCtrl), full-length Upf1 (pUpf1) or the HD-Cter fragment (pUpf1-HD-Cter). Digoxigenin containing PCR probes specific to the intron of RPL28 mRNA precursor and for the mature form were used with anti-digoxigenin peroxidase coupled detection of the two RNA forms.



Appendix Fig. S5: Comparison between the canonical SURF/DECID model features **(A)** and our extended yeast-based model **(B)** for NMD. Orange and blue squares mark equivalent steps in both models. Light grey elements in the revised model represent optional steps that can further enhance the NMD process under certain conditions and in specific organisms.



Appendix Fig. S6. Similarities and differences between yeast Nmd4 PIN and PIN domains of human SMG6 and SMG5. (**A**) Structure of the PIN domain of hSmg6 (Glavan *et al*, 2006, PDB entry 2HWW), highlighting conserved aspartate and glutamate residues. The Asp residues required for catalysis (Asp1251, 1353, 1392) are depicted with an asterisk. (**B**) Zones of sequence alignment between the human sequences and the PIN domain of yeast Nmd4 centered on conserved acidic residues are underlined. Unlike the PIN domain of hSmg5, with no catalytic activity, the PIN domain of Nmd4 retains a conserved Asp in a position equivalent with D1392 in hSmg6. Residue numbers correspond to positions in the human Smg6 protein (Uniprot Q86US8).

Appendix Tables

Appendix Table S1. Summary of the number of replicates, number of proteins robustly quantified and numbe of proteins lost after RNase treatment. The over-representation of RNA binding factors in the 'lost' fraction was computed using the hypergeometric distribution.

							over-representation (hypergeometric text)			
	strain	number of replicates with/without RNase		number of quantified proteins		number of RNase sensitive proteins	constituents of ribosome (231)		RNA binding (847)	
Experiment		NO	YES	NO	YES		count	p-value	count	p-value
BY4741	WT	5	4	9	0	/	/	/	/	/
Upf1-TAP	WT	6	3	270	45	225	47	5E-27	80	7E-22
Upf2-TAP	WT	6	3	172	62	111	30	2E-21	44	8E-15
Upf3-TAP	WT	5	3	217	70	151	25	1E-12	62	4E-21
Nmd4-TAP	WT	4	3	121	32	89	34	7E-30	37	1E-13
Ebs1-TAP	WT	3	3	195	133	79	19	3E-13	24	2E-06
Dcp1-TAP	WT	3	3	255	84	175	19	7E-07	66	6E-20
Hrr25-TAP	WT	4	3	203	116	99	15	8E-08	38	1E-12
TAP-Upf1-2-971	upf1Δ	3	3	457	214	247	24	3E-07	74	1E-15
TAP-Upf1-2-853	upf1Δ	2	0	223	/	/	/	/	/	/
TAP-Upf1-208-853	upf1Δ	2	0	235	/	/	/	/	/	/
TAP-Upf1-2-208	upf1Δ	3	3	208	59	149	37	2E-24	60	6E-20
TAP-Upf1-208-971	upf1Δ	3	3	429	104	325	41	3E-15	99	6E-21
Upf1-TAP	upf2Δ	3	3	200	42	161	46	4E-33	56	3E-15
Upf1-TAP	upf3Δ	3	3	242	50	192	37	3E-20	71	1E-20
Upf2-TAP	upf1Δ	2	3	242	86	157	20	2E-08	45	2E-09
Upf3-TAP	upf1Δ	2	0	235	/	/	/	/	/	/
Upf1-TAP	nmd4Δ	3	3	140	35	107	44	6E-40	41	2E-13
Nmd4-TAP	upf1Δ	2	2	235	80	165	32	9E-18	64	4E-20

Appendix Table S2. RNA sequencing raw data analysis summary

Experiment	number of replicates	replicates	total reads	number of uniquely mapped reads	reads assigned to known features
BY4741(GB)	3	BY4741_GB_rep1	30,670,583	25,581,713	25,401,415
		BY4741_GB_rep2	25,220,407	20,672,565	20,528,503
		BY4741_GB_rep3	29,245,177	25,083,674	24,903,966
***upf1*Δ(GB)**	3	upf1-delta_rep1	24,224,426	20,728,624	20,501,571
		upf1-delta_rep2	30,152,228	24,816,935	24,548,407
		upf1-delta_rep3	28,395,388	21,711,919	21,468,870
BY4741	3	BY4741_rep1	21,764,071	17,906,929	17,733,252
		BY4741_rep2	29,781,174	13,645,348	13,502,966
		BY4741_rep3	40,044,726	28,578,567	28,287,975
***nmd4*Δ**	3	nmd4-delta_rep1	23,282,316	19,201,253	19,003,405
		nmd4-delta_rep2	33,352,825	27,228,434	26,935,166
		nmd4-delta_rep3	32,569,200	25,866,315	25,600,752
***ebs1*Δ**	3	ebs1-delta_rep1	25,183,381	20,795,767	20,583,050
		ebs1-delta_rep2	30,637,150	25,237,059	24,978,404
		ebs1-delta_rep3	29,713,531	22,452,058	22,197,467
***nmd4*Δ/ *ebs1*Δ**	3	nmd4-ebs1-delta_rep1	23,253,638	19,428,559	19,219,804
		nmd4-ebs1-delta_rep2	23,635,820	20,419,649	20,199,665
		nmd4-ebs1-delta_rep3	37,828,620	30,145,648	29,809,901

Appendix Table S3 (strains)

strain	original strain	Genotype	Reference
LMA2154 (BY4741)		*MATa ura3Δ0 his3Δ1 leu2Δ0 met15Δ0*	Brachmann et al. 1998
LMA2155 (BY4742)		*MATα ura3Δ0 his3Δ1 leu2Δ0 lys2Δ0*	Brachmann et al. 1998
LMA2194	BY4741	UPF1-TAP(HIS3MX)	Ghaemmaghami et al., 2003
LMA3730	BY4741	UPF1-CRAP(URA3)	This study
LMA2192	BY4741	UPF2-TAP(HIS3MX)	Ghaemmaghami et al., 2003
LMA3731	BY4741	UPF2-CRAP(URA3)	This study
LMA2193	BY4741	UPF3-TAP(HIS3MX)	Ghaemmaghami et al., 2003
LMA4263	BY4741	UPF3-CRAP(URA3)	This study
LMA3317	BY4741	NMD4-TAP(HIS3MX)	Ghaemmaghami et al., 2003
LMA3728	BY4741	NMD4-CRAP(URA3)	This study
LMA4264	BY4741	DCP1-TAP(HIS3MX)	Ghaemmaghami et al., 2003
LMA4996	BY4741	EBS1-TAP(HIS5)	This study
LMA3312	BY4741	HRR25-TAP(HIS3MX)	Ghaemmaghami et al., 2003
LMA3849	BY4741	UPF1-CRAP(URA3)/ NMD4-HA(KANMX6)	This study
LMA3851	BY4741	UPF1-CRAP(URA3) / EBS1-HA(KANMX6)	This study
LMA3852	BY4741	UPF1-CRAP(URA3) / EDC3-HA(KANMX6)	This study
LMA1667	BY4741	UPF1::KANMX6	Giaever et al., 2002
LMA1669	BY4741	UPF2::KANMX6	Giaever et al., 2002
LMA1671	BY4741	UPF3::KANMX6	Giaever et al., 2002
LMA3732	BY4741	NMD4::KANMX6	Giaever et al., 2002
LMA4112	BY4741	UPF1::KANMX6 / UPF2-CRAP(URA3)	This study
LMA4113	BY4741	UPF1::KANMX6 / UPF3-CRAP(URA3)	This study
LMA4114	BY4741	UPF1::KANMX6 / NMD4-CRAP(URA3)	This study
LMA3739	BY4741	UPF2::KANMX6 / UPF1-CRAP(URA3)	This study
LMA4701	BY4741	UPF2::KANMX6 / NMD4-CRAP(URA3)	This study
LMA3735	BY4741	UPF3::KANMX6 / UPF1-CRAP(URA3)	This study
LMA3736	BY4741	NMD4::KANMX6 / UPF1-CRAP(URA3)	This study
LMA4523	BY4741	UPF1::HIS3MX / UPF2::KANMX6	This study
LMA4524	BY4741	UPF1::HIS3MX / UPF3::KANMX6	This study
LMA4525	BY4741	UPF1::HIS3MX / EBS1::KANMX6	This study
LMA4678	BY4741	UPF1::HIS3MX / NMD4::KANMX6	This study
LMA3853	BY4742	NMD4::ProMFalpha2NAT / EBS1::KANMX6	This study
LMA5020	BY4741	LSM1-HA(KANMX6)	This study
LMA5022	BY4741	LSM1-HA(KANMX6)/UPF1-TAP(HIS3MX)	This study
LMA5024	BY4741	PAT1-HA(KANMX6)	This study
LMA5026	BY4741	PAT1-HA(KANMX6)/UPF1-TAP(HIS3MX)	This study
LMA4423	BY4741	DCP2-3Flag-AID-kanMX6-TIR1	This study

Appendix Table S4 (plasmids)

Plasmid	Description and ID (yeast marker)	Reference
pRS316	pRS316 (URA3)	Sikorski & Hieter, 1989
pCM189-NTAP	pl.1233 (H1) (URA3)	This study
pCM189-NTAP-Upf1-FL	pl.1442 (TAP-UPF1-FL) (URA3)	This study
pCM189-NTAP-Upf1-CH	pl.1443 (TAP-UPF1-2-208) (URA3)	This study
pCM189-NTAP-Upf1-HD-Cter	pl.1444 (TAP-UPF1-208-971) (URA3)	This study
pCM189-NTAP-Upf1-CH-HD	pl.1521 (TAP-UPF1-2-853) (URA3)	This study
pCM189-NTAP-Upf1-HD	pl.1522 (TAP-UPF1-208-853) (URA3)	This study
pDEST14-UPF1	pl.1350 – Gateway destination vector	This study
pDONR201-UPF1	pl.1330 – Gateway source vector	This study
pCRBlunt-CRAP(6-HisTAP)	pl.1287 – TAP to CRAP cassette vector	This study
p189-HA-ALA1-KAN	pl.1490 - NMD reporter (HA and long 3'UTR)	This study
pFA6a-polyG-3Flag-miniAID-KanMX6	pl 1451 - to build Dcp2 degron strain	This study
BYP7569-promADE2-KanMX6	pl 1367 - to build Dcp2 degron strain	This study, Nishimura et. al 2009

Appendix Table S5 (oligonucleotides)

Oligonucleotide	Use	Sequence
CS887_fw_RPL28intron	real time PCR forward primer	CCATCTCACTGTTGAGACGG
CS888_rv_RPL28intron	real time PCR reverse primer	CTCAGTTTGCGATGGAAGAG
CS889_rv_RPL28exon2	real time PCR reverse primer	ATGTTGACCACCGGCCATAC
CS946_RPL28	real time PCR forward primer	TCACGTCTCAGCCGGTAAAG
CS1076_fw_RIM1Qex1ex2	real time PCR forward primer	GTTAGAAAAGGCGCTTTGGTATATG
CS1077_rv_RIM1QRTex2	real time PCR reverse primer	AACCGTCGTCTCTCTCGAAG
CS1128_rv_HA_Q	real time PCR forward primer	GCATAATCTGGAACATCATATG
CS1429_DAL7_fwQ	real time PCR forward primer	TGAAACTTTGCCAGCGGCCTTC
CS1430_DAL7_rvQ	real time PCR reverse primer	TCCCAACGACCACAGTTCAAACC
CS1359_fw_NAM7_2_pTM189Not	construction of UPF1 plasmid	ttaagaaaatctcatcctccggggcacttGATgcgGTCGGTTCCGGTTCTCACAC
CS1361_rv_NAM7_208	construction of UPF1 plasmid	ATAACTAATTACATGATGCGGCCCTCCTGCAGGGCTTAATTGGATCTCCATTTTGCCTC
CS1362_fw_NAM7_s208	construction of UPF1 plasmid	ttaagaaaatctcatcctccggggcacttGATgcgAATAAAGACGCTACAATTAATGATATTGACG
CS1364_rv_NAM7_971	construction of UPF1 plasmid	ATAACTAATTACATGATGCGGCCCTCCTGCAGGGCTTATATTCCCAAATTGCTGAAGTC
CS1393_rv_UPF1_853STOP	construction of UPF1 plasmid	ATAACTAATTACATGATGCGGCCCTCCTGCAGGGCTTActgaggacgaactaattgaac
CS1379_AscI_fw	construction of ALA1 NMD reporter	ATGTCGTATCCATATGATGTTCCAGATTATGCTGGAGGAGGAGGCGTCGTTTACGTAGGTTACGGAGCGC
CS1380_AscI_rv	construction of ALA1 NMD reporter	AGCGGTCCATTTTTGCTTATCACCGATCGTCGGCGCTCCGTAACCTACGTAAACGACGCCTC
CS1473_Pat1_fw_F2	construction of HA tagged strain	TAAACGTTATGGGGTTGGTGTATCGCGATGGTGAAATATCAGAACTAAAGCGGATCCCCGGGTTAATTAA
CS1474_Pat1_rv_R1	construction of HA tagged strain	GGAGAAAAAAAAAATACATGCGTAAGTACATTAAAATTACAGGAAAAATCTTAGAATTCGAGCTCGTTTAAAC
CS1476_Lsm1_fw_F2	construction of HA tagged strain	AAATGGCCCGCCATGGTATCGTTTACGATTTCCATAAATCTGACATGTACCGGATCCCCGGGTTAATTAA
CS1477_Lsm1_rv_R1	construction of HA tagged strain	GAGAGTTTACTCCAGGATATATGTTGGTAGTATTGTGTTTTTCTTTCTTAGAATTCGAGCTCGTTTAAAC
MD57_fw_EDC3-HA	construction of HA tagged strain	CCAAAACTGTGATCTTTTCGTCACTGACGGGTCCCTGCTATTAGATTTGCGGATCCCCGGGTTAATTAAC
MD58_rv_EDC3-HA	construction of HA tagged strain	CCGTATGCTTATACGTATGTATCCAGTTTAGGCTAAAGTAATTCTTGCATCGATGAATTCGAGCTCG
CS1450_RPL28i_fw	Northern blot probe amplification	GTGTTGTGCAACCAATATGTCG
CS1451_RPL28i_rv	Northern blot probe amplification	CATAATTGCGCTCTCTACAACC
CS1463_RPL28fw_ex2	Northern blot probe amplification	GTAAAGGTCGTATCGGTAAGC
CS1464_RPL28rv_ex2	Northern blot probe amplification	CGATCAATTCAACAACACCACC
MD100_rv_Nmd4_640_NotI	Cloning of NMD4 in pHL5	CTAGTGCGGCCGCCTGTGGGGACCACAATTC
MD99_fw_Nmd4_1_NheI	Cloning of NMD4 in pHL5	GTACTGCTAGCATGACACAATATAATTTCATTATAG

Appendix references

Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ & Madden TL (2012) Domain enhanced lookup time accelerated BLAST. *Biol Direct* **7:** 12

Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hirschman JE, Hitz BC, Karra K, Krieger CJ, Miyasato SR, Nash RS, Park J, Skrzypek MS, Simison M, et al (2012) Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.* **40:** D700-705

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M & Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29:** 15–21

Fiorini F, Bonneau F & Le Hir H (2012) Biochemical characterization of the RNA helicase UPF1 involved in nonsense-mediated mRNA decay. *Meth. Enzymol.* **511:** 255–274

Ghaemmaghami S, Huh W-K, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK & Weissman JS (2003) Global analysis of protein expression in yeast. *Nature* **425:** 737–741

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

Kubota T, Nishimura K, Kanemaki MT & Donaldson AD (2013) The Elg1 replication factor C-like complex functions in PCNA unloading during DNA replication. *Mol. Cell* **50:** 273–280

Kushnirov VV (2000) Rapid and reliable protein extraction from yeast. *Yeast* **16:** 857–860

Liao Y, Smyth GK & Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30:** 923–930

Love MI, Huber W & Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15:** 550

Nishimura K, Fukagawa T, Takisawa H, Kakimoto T & Kanemaki M (2009) An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat. Methods* **6:** 917–922

Pryszcz LP, Huerta-Cepas J & Gabaldón T (2011) MetaPhOrs: orthology and paralogy predictions from multiple phylogenetic evidence using a consistency-based confidence score. *Nucleic Acids Res* **39:** e32

Vizcaíno JA, Csordas A, del-Toro N, Dianes JA, Griss J, Lavidas I, Mayer G, Perez-Riverol Y, Reisinger F, Ternent T, Xu Q-W, Wang R & Hermjakob H (2016) 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **44:** D447-456