Long open reading frame transcripts escape nonsense-mediated mRNA decay in yeast
Laurence Decourty, Antonia Doyen, Christophe Malabat, Emmanuel Frachon, Delphine Rispal, Bertrand Sériaphin, Frank Feuerbach, Alain Jacquier, Cosmin Saveanu

To cite this version:
Long Open Reading Frame Transcripts Escape Nonsense-Mediated mRNA Decay in Yeast

Laurence Decourty, 1 Antonia Doyen, 1 Christophe Malabat, 1 Emmanuel Frachon, 2 Delphine Rispal, 3, 4 Bertrand Séraphin, 3 Frank Feuerbach, 1 Alain Jacquier, *, 1 and Cosmin Saveanu 1, *

1 Génétique des Interactions Macromoléculaires, Institut Pasteur, CNRS UMR3525, 25-28 rue du docteur Roux, 75015 Paris, France
2 Production de Protéines Recombinantes et d’Anticorps (PF5), Institut Pasteur, 25-28 rue du docteur Roux, 75015 Paris, France
3 Equipe Labellisée La Ligue, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), INSERM U964/CNRS UMR 7104/Université de Strasbourg, 67404 Illkirch, France
4 Present address: Department of Molecular Biology, University of Geneva, 30 quai Ernest-Ansermet, 1211, Geneva 4, Switzerland
*Correspondence: alain.jacquier@pasteur.fr (A.J.), cosmin.saveanu@pasteur.fr (C.S.)
http://dx.doi.org/10.1016/j.celrep.2014.01.025

This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-No Derivative Works License, which permits non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

SUMMARY

Nonsense-mediated mRNA decay (NMD) destabilizes eukaryotic transcripts with long 3’ UTRs. To investigate whether other transcript features affect NMD, we generated yeast strains expressing chromosomal-derived mRNAs with 979 different promoter and open reading frame (ORF) regions and with the same long, destabilizing 3’ UTR. We developed a barcode-based DNA microarray strategy to compare the levels of each reporter mRNA in strains with or without active NMD. The size of the coding region had a significant negative effect on NMD efficiency. This effect was not specific to the tested 3’ UTR because two other different NMD reporters became less sensitive to NMD when ORF length was increased. Inefficient NMD was not due to a lack of association of Upf1 to long ORF transcripts. In conclusion, in addition to a long 3’ UTR, short translation length is an important feature of NMD substrates in yeast.

INTRODUCTION

Nonsense-mediated decay (NMD) is a translation-dependent mechanism that leads to degradation of eukaryotic RNAs with long 3’ UTRs (reviewed in Kervestin and Jacobson, 2012). Such transcripts are often described as having a premature termination codon (PTC) and can arise from inefficiently spliced pre-mRNAs that were exported to the cytoplasm (Sayani et al., 2008), RNAs generated through alternative splicing (He et al., 2003; Mendell et al., 2004), or cytoplasmic unstable “noncoding” RNAs (Thompson and Parker, 2007). In contrast to the common assumption that NMD substrates are rare in eukaryotes, recent large-scale transcript analyses have uncovered an important fraction of yeast transcripts that contain short open reading frames (ORFs) in the 5’ UTR of annotated coding sequences (upstream ORFs [uORFs]) (Arribere and Gilbert, 2013; Pelechano et al., 2013). Such uORF-containing transcripts are good substrates for NMD in both yeast (He et al., 2003) and mammals (Hurt et al., 2013).

NMD depends on three major conserved factors: Upf1 (Nam7 in yeast), an ATP-dependent RNA helicase, Upf2 (Nmd2 in yeast), and Upf3, with additional proteins involved in the process in metazoans and possibly in yeast (He and Jacobson, 1995; Luke et al., 2007). Binding of Upf1 to transcripts is correlated with degradation of these transcripts through NMD in yeast (Johansson et al., 2007). It has been recently postulated that Upf1 binding to long 3’ UTR regions could represent the molecular mark that leads to mammalian mRNA degradation through NMD (Hogg and Goff, 2010). Large-scale Upf1 crosslinking to RNA data indicate that although NMD substrates are more likely to be bound by Upf1 (Hurt et al., 2013), Upf1 binds many transcripts not affected by NMD (Zünd et al., 2013). These studies have shown that blocking translation results in a loss of specificity of Upf1 binding to 3’ UTRs. Whether Upf1 is actively recruited on 3’ UTRs during translation termination (Kurosaki and Maquat, 2013; Shigeoka et al., 2012) or is displaced from ORF regions by translation is unclear.

The presence of Upf1 on a long 3’ UTR thus seems required, but not sufficient, for NMD. Additional molecular events, like an aberrant translation termination, are likely to affect the stability of an Upf1-bound mRNA. Results obtained with NMD reporters with PTCs that are far from the poly(A) tail in S. cerevisiae (Amrani et al., 2004), D. melanogaster (Behm-Ansmant et al., 2007), and mammalian cells (Eberle et al., 2008; Singh et al., 2008) led to the proposal that the long distance between the PTC and the poly(A) tail affects translation termination and thus triggers NMD (Amrani et al., 2004).

Although many studies were focused on the influence of the 3’ region on NMD efficiency, little is known about the influence of the region upstream of the stop codon in the process. To address this issue, and to analyze the importance of the region upstream of the stop codon in NMD, we have set up an experimental system that addresses on a large scale the context in which an otherwise normal termination codon is able to trigger...
RESULTS

Obtaining and Using a Barcoded Collection of NMD Reporters in Yeast

Replacing natural 3’ UTRs with artificially long sequences allows the generation of strains with altered gene expression that are useful for phenotype or genetic studies of essential genes (decreased abundance by mRNA perturbation [DAmP] strategy; Schuldiner et al., 2005). We have built a collection of 979 DAmP-modified strains (Figures 1A and S1A; Table S1) that can be directly used for large-scale genetic interaction mapping (GIM) screens (Decourty et al., 2008). The DAmP strains that we have generated differ from similar strains reported earlier (Breslow et al., 2008; Yan et al., 2008) because they include molecular barcodes at each modified locus. The locus-specific barcodes situated upstream of the inserted KanMX4 cassette are present in the transcribed DAmP mRNAs, which allows the study of NMD efficiency on hundreds of reporter mRNAs that have the same 1.4-kb-long 3’ UTR yet originate from various promoters and have different 5’ UTR and coding regions.

The GIM method allows one to get a large number of haploid yeast cells that contain two genomic modifications tagged by two different antibiotic resistance markers. The populations of double mutants are then used to investigate quantitative outcomes of combining two mutations in a single cell, which most commonly can lead to a growth rate change. We measured the relative growth rate of 807 DAmP and 3,681 deletion strains (Giaever et al., 2002) inactivated for NMD through deletion of either UPF1 or UPF2 as compared with their NMD-competent equivalents (strategy depicted in Figure 1B). Less than 1% of nonessential gene deletion strains displayed an improved growth under these conditions. In contrast, about 8% of DAmP strains showed faster growing rates in combination with NMD mutants (upper right quadrant in Figure 1C; Table S2). This result can be explained by the stabilization of DAmP mRNAs in mutant strains deficient for NMD, followed by increased protein production for the affected gene. We directly tested DAmP mRNA level changes following NMD inactivation in cells expressing the DAmP version of SMT3, a strain among those showing a growth improvement when combined with upf1Δ or upf2Δ (average log2[NMD inactivation/reference] of 1.9). UPF1 deletion increased ten times smt3-DAmP RNA levels (Figures S1B and S1C). Growth defect suppression by NMD inactivation is thus an indirect indication of the extent of mRNA destabilization elicited by DAmP modification.

The mRNAs most sensitive to the destabilizing effect of a long 3’ UTR, as judged from the growth rate improvement for the corresponding DAmP strains in both upf1Δ and upf2Δ screens (log2[NMD inactivation/reference] over 1; Figure 1C) had significantly (p < 3 x 10^-12) shorter than average ORFs (Figure 1D). This potential effect of ORF size on NMD efficiency was mirrored by the observation that short gene DAmP strains were more affected for growth than longer gene DAmP strains. A correlation between ORF size and growth defect was not present in the nonessential gene deletion collection (Figure S1D; Table S3).

Figure 1. DAmP Modification Preferentially Affects Short Essential Genes Function

(A) Schematics of the DAmP modification. TSS, transcription start site; TER, natural transcription termination region; pTEF, Kanβ, and TerTEF indicate the promoter, G418 resistance ORF, and terminator of the KanMX4 cassette, respectively. Filled rectangles mark barcode regions flanked by universal priming sequences, whereas a black diamond indicates the ORF stop codon.

(B) Strategy for large-scale analysis of the effect of NMD inactivation on DAmP and gene deletion strains based on the GIM method (Decourty et al., 2008). Filled symbols correspond to gene deletion (blue and green) or DAmP modification (orange) alleles.

(C) Correlation between the effects on growth for deletion and DAmP strains when NMD was inactivated via deletion of either UPF1 or UPF2 (n = 4,488, p < 10^-12, Kendall). Each value is the average of results from two independent screens. Orange crosses correspond to DAmP strains and blue dots correspond to deletion strains.

(D) The DAmP strains showing the strongest growth recovery when combined with NMD inactivation (log2 value > 1) correspond to genes that have shorter than average ORFs; p < 3 x 10^-12, Mann-Whitney U test, n = 744 (less than one category), n = 63 (more than one category).

See also Figure S1.
A Strategy for Barcode- and GIM-Based Large-Scale Tests of NMD

Although the observed effects on growth indicated a possible link between NMD efficiency and ORF size, we wanted to directly test this effect by measuring the changes in all DAmp mRNA levels when NMD is functional or not. To this end, we developed an experimental strategy that takes advantage of the barcodes specific to each mutant to estimate the levels of DAmp mRNAs in a complex population of strains (depicted in Figure 2A). The barcode-based strategy led to estimates of RNA abundance that were highly correlated (n = 565, r = 0.77, p < 2.2 × 10⁻¹⁶, Pearson) with published abundance values (Lipson et al., 2009) for the corresponding mRNAs (Figure S2A).

We used the RNA-derived barcode-based method to evaluate the changes in the levels of hundreds of DAmp mRNAs in populations of strains where NMD was inactivated by UPF1 or UPF2 deletion (GIM method). YELO68C deletion, affecting an intergenic region that has no impact on any known yeast process, was used to generate a reference DAmp population. The strong negative correlation (n = 470, r = -0.57, p < 2.2 × 10⁻¹⁶, Pearson) observed between the increase in mRNA levels in the absence of Upf1 and the initial ORF size (Figure 2B; Table S4) suggested that ORF size could be a major determinant of NMD efficiency. Because NMD depends on translation, we also evaluated changes in DAmp mRNA levels after blocking translation for 30 min with 50 μg/ml cycloheximide. A significant inverse correlation (n = 523, r = -0.44, p < 2.2 × 10⁻¹⁶, Pearson) was observed between the size of the affected ORFs and DAmp mRNA accumulation upon translation inhibition (Figure S2B; Table S4).

The results obtained with the barcode-based strategy combined with the presence or absence of NMD factors were strongly correlated with changes in the steady-state levels of a set of 17 different mRNAs in individual strains before and after DAmp modification, which were measured by quantitative PCR (Figure 2C). To test the potential influence of cellular phenotypes that could bias the results on the levels of DAmp mRNAs in haploid strains, we measured the relative levels of DAmp mRNAs in heterozygous diploid strains obtained by batch mating of the DAmp strains with a wild-type BY4742 strain. The DAmp mRNA levels in haploid and diploid cells were strongly correlated (p < 1.2 × 10⁻¹³, r² = 0.91) (Figure S2C). Altogether, these data show that addition of a long 3’ UTR to a yeast mRNA elicits translation-dependent transcript instability by NMD only if the ORF is relatively short, whereas longer ORF mRNAs escape NMD.
The inability of NMD to degrade long-ORF-containing DAmp mRNAs could be due to a decreased association of Upf1 with the transcripts. To test this hypothesis, we affinity purified Upf1-TAP-associated mRNAs from strains expressing DAmp mRNAs and found that the addition of a long 3' UTR extension to mRNAs with ORFs longer than 2 kb (RRP5 and ALA1) or mRNAs with ORFs shorter than 1.5 kb (PGK1, PRO3, SHR3, and MMF1) led to the same consistent enrichment in association with Upf1 (Figure 2D). Despite the similar association with Upf1, only short ORF DAmp mRNAs (PRO3, SHR3, and MMF1) showed a marked enrichment in association with the original PTC-containing PGK1 led to stabilization of the mRNA (estimated half-life of 11 min). Filled and empty diamond symbols indicate stop codons.

The inability of NMD to degrade long-ORF-containing DAmp mRNAs could be due to a decreased association of Upf1 with the transcripts. To test this hypothesis, we affinity purified Upf1-TAP-associated mRNAs from strains expressing DAmp mRNAs and found that the addition of a long 3' UTR extension to mRNAs with ORFs longer than 2 kb (RRP5 and ALA1) or to mRNAs with ORFs shorter than 1.5 kb (PGK1, PRO3, SHR3, and MMF1) led to the same consistent enrichment in association with Upf1 (Figure 2D). Despite the similar association with Upf1, only short ORF DAmp mRNAs (PRO3, SHR3, and MMF1) showed a marked enrichment in association with the original PTC-containing PGK1 led to stabilization of the mRNA (estimated half-life of 11 min). Filled and empty diamond symbols indicate stop codons.

**Increasing ORF Length Stabilizes Reporter NMD Substrates**

We wondered if the effect we identified with the large-scale study could be observed on typical NMD reporters with a different long 3' UTR, promoter, and coding sequence. We used the extensively studied PGK1 reporter system to test the effect of translated ORF size on the stability of a reporter mRNA. Placing an additional copy of PGK1 ORF upstream and in frame with a PGK1-derived reporter bearing a PTC led to stabilization of the corresponding NMD-affected mRNA (Figure 3). The increase in ORF size was accompanied by a decrease in the ratio of coding-sequence length to 3' UTR length, and we wondered if this ratio could have an influence on NMD efficiency. To test this hypothesis, we measured the steady-state levels of transcripts containing a long ORF (ADE3; ORF length 2,841 nt) combined with various 3' UTR lengths. Only a modest decrease in mRNA levels was elicited by a 1.7 kb 3' UTR inserted downstream the stop codon of the ADE3 gene, and this decrease was not enhanced when the 3' UTR size was increased from 1.7 to 4.8 and 5.9 kb (Figures S3A and S3B). It is thus likely that it is the ORF size and not its ratio compared with the 3' UTR that plays a role in NMD sensitivity of transcripts.

To establish if ORF size alone, independent of the translated sequence, could affect NMD, we designed a series of NMD reporters consisting of an identical 5' UTR, start and stop codon context, and a long 3' UTR. Transcripts containing three different fragments of ALA1 ORF were tested either alone or in a tandem repeat that created a long coding sequence (Figure 4A). Because mRNA decay for this reporter was too fast to allow precise measurements at 30°C, we measured decay rates of transcripts at
20°C. Increasing the coding sequence length slowed down the NMD-dependent degradation rate of the reporter mRNA (Figures 4B, 4C, and S4B). Because all the sequences contained in the long reporter are translated in one or another of the short transcripts (as verified by western blots; Figure S4C), we concluded that it is ORF size and not the translated sequence that affects NMD efficiency on this reporter. Thus, coding sequence size affected translation-dependent mRNA decay of three different NMD reporters and can potentially affect NMD efficiency on any substrate.

DISCUSSION

Molecular barcodes allow multiplexing of DNA based assays in mixtures of cells or plasmids. We show here that inclusion of barcodes in transcribed sequences also allows parallel tests on hundreds of reporter mRNAs. Barcoded mRNA NMD reporters served to identify short ORF length as an important feature of NMD substrates. Contrary to the model in which Upf1 binding to a long 3’ UTR alone is sufficient to trigger NMD (Hogg and Goff, 2010), our data suggest that ORF size modulates NMD for transcripts associated with Upf1. We do not know yet if the ORF-size effect is due to molecular events affecting translation termination (Amrani et al., 2004) or other steps in the degradation pathway.

The number of ribosomes reaching the termination codon could affect the stability of an Upf1-bound transcript. Ribosome density decreases along the transcripts from initiation to the stop codon (Ingolia et al., 2009), and the ORF length could thus influence the number of translation-termination events. More ribosomes at translation termination could lead, in the presence of a long 3’ UTR, to faster degradation rates. However, we could find no correlation between ribosome density in the 50 nt preceding the stop codon of a given mRNA (Ingolia et al., 2009) and the level of destabilization induced by a long 3’ UTR (not shown). Thus, ribosome density changes along mRNAs are unlikely to explain the ORF-size effect on NMD efficiency. It is possible that it is not the number but the quality of the termination events that is different between long ORF and short ORF transcripts. Inefficient translation termination at early times after translation initiation could allow optimal coupling with the action of NMD factors.

Our observations are consistent with the preference for short ORFs among natural NMD substrates. The bulk of NMD substrates in yeast consists of transcripts that have short ORFs upstream of a long untranslated sequence (Arribere and Gilbert, 2013; He et al., 2003). Abundant mRNAs lack such uORFs (Yun et al., 2012). A paucity of ATG in the 5’ UTR of mRNAs characterizes not only yeast but also many studied organisms (Rogozin et al., 2001). Thus, transcripts with short ORFs (probably mostly uORFs) are likely to be both excellent substrates for degradation, as shown here, and the most prevalent type of NMD substrate in eukaryotes.

EXPERIMENTAL PROCEDURES

The collection of barcoded DAmP strains was built by homologous recombination using PCR products. Oligonucleotide sequences and protocols for analyzing RNA as well as the building strategies for the NMD reporter plasmids are detailed in Supplemental Experimental Procedures.

GIM Screens, Growth Rates, and Double-Mutant Strains

GIM was done as previously described (Decourty et al., 2008), with two exceptions: the barcoded 979 DAmP strains were added to the pool of mutants and a custom-made constant turbidity system for haploid selection and culture was used. This system uses a stream of sterile air that allows continuous culture of yeast in 10 ml reaction flasks. Data analysis of the microarray results was done using R to perform loess normalization with marray function of package marray. For expression analysis, two replicate arrays were performed, with each replicate containing at least two biological replicates. For each strain (experiments were performed in technical replication), the median of the log2-transformed ratios was calculated. The median of the log2-transformed ratios to 0, and calculate average values using R to perform loess normalization with marray function of package marray. For expression analysis, two replicate arrays were performed, with each replicate containing at least two biological replicates.
Funding was provided by the Institut Pasteur, the CNRS, the CNRS CERMIB-IGBMC, the Ligue Contre le Cancer (B.S.: Equipe Labellisée 2011), and the French Agence Nationale de la Recherche (ANR-07- BLAN-0093, ANR-11-BSV8-009-02, and ANR-08-JCJC-0019-01/GENO-GIM).

Received: October 12, 2013
Revised: December 19, 2013
Accepted: January 17, 2014
Published: February 13, 2014

REFERENCES


