Nonsense-mediated mRNA decay in *Drosophila*: at the intersection of the yeast and mammalian pathways

David Gatfield, Leonie Unterholzner, Francesca D.Ciaccarelli, Peer Bork and Elisa Izaurralde

EMBL, Meyerhofstrasse 1, D-69117 Heidelberg, Germany

1Corresponding author
e-mail: izaurralde@embl-heidelberg.de

The nonsense-mediated mRNA decay (NMD) pathway promotes the rapid degradation of mRNAs containing premature stop codons (PTCs). In *Caenorhabditis elegans*, seven genes (*smg*1–7) playing an essential role in NMD have been identified. Only SMG2–4 (known as UPF1–3) have orthologs in *Saccharomyces cerevisiae*. Here we show that the *Drosophila* orthologs of UPF1–3, SMG1, SMG5, SMG5 and SMG6 are required for the degradation of PTC-containing mRNAs, but that there is no SMG7 ortholog in this organism. In contrast, orthologs of SMG5–7 are encoded by the human genome and all three are required for NMD. In human cells, exon boundaries have been shown to play a critical role in defining PTCs. This role is mediated by components of the exon junction complex (EJC). Contrary to expectation, however, we show that the components of the EJC are dispensable for NMD in *Drosophila* cells. Consistently, PTC definition occurs independently of exon boundaries in *Drosophila*. Our findings reveal that despite conservation of the NMD machinery, different mechanisms have evolved to discriminate premature from natural stop codons in metazoa.

Keywords: EJC/mRNA decay/NMD/RNPS1/Y14

Introduction

Nonsense (codon)-mediated mRNA decay (NMD) is an RNA surveillance pathway that degrades mRNAs possessing premature translation termination codons (PTCs) (reviewed in Wagner and Lykke-Andersen, 2002). In mammalian cells, NMD is linked to splicing. Indeed, the position of a stop codon relative to the 3′-most intron has emerged as a critical determinant of whether it will be interpreted as being premature or natural (reviewed in Nagy and Maquat, 1998). Natural stop codons are typically located within the terminal exon, while destabilizing nonsense codons lie >50 nucleotides upstream of the last exon–exon boundary. This observation gave rise to the hypothesis that the process of splicing deposits a mark at exon–exon junctions. These marks are used to discriminate premature stops (upstream of the last mark) from natural stops (downstream of the last mark) (reviewed in Nagy and Maquat, 1998; Wagner and Lykke-Andersen, 2002).

The exon junction complex (EJC) is thought to provide the splicing marks for NMD. This multiprotein complex binds ~20–24 nucleotides upstream of splice junctions in a splicing-dependent manner (Le Hir et al., 2001). Although the precise biochemical composition of the EJC remains to be established, at least five proteins have been identified as components of the EJC assembled *in vitro*. These include REF1 (also known as Aly), SRm160, RNPS1 and Y14:MAGOH heterodimers (Le Hir et al., 2001). Y14:MAGOH dimers, and possibly other EJC proteins, remain bound to the newly exported mRNAs and could be used in the cytoplasm to mark the location of exon–exon boundaries (Kim et al., 2001a,b; Le Hir et al., 2001; Lykke-Andersen et al., 2001). Consistently, both RNPS1 and Y14:MAGOH have been shown to elicit NMD when tethered downstream of a stop codon (Lykke-Andersen et al., 2001; Fribourg et al., 2003; Gehring et al., 2003).

In *Saccharomyces cerevisiae*, most genes lack introns and, apart from REF, none of the known components of the EJC is conserved. Consistently, exon boundaries are not used to distinguish natural from premature stop codons in budding yeast. Instead, yeast mRNAs are marked by loosely defined downstream sequence elements (DSEs) that might have a function analogous to that of mammalian exon junctions (reviewed in Wagner and Lykke-Andersen, 2002).

The genetics of NMD have been studied in *S.cerevisiae* and *Caenorhabditis elegans*. Three yeast (UPF1–3) and seven *C.elegans* genes (*smg*1–7, suppressor with morphogenetic effects on genitalia) that play an essential role in NMD have been identified. The UPF1–3 proteins (also known as SMG2–4 in *C.elegans*) are conserved from yeast to humans, while SMG1 has orthologs in higher eukaryotes but not in *S.cerevisiae* (reviewed in Wagner and Lykke-Andersen, 2002).

In mammalian cells, UPF1 is a nucleocytoplasmic shuttling protein with a predominantly cytoplasmic localization at equilibrium (Lykke-Andersen et al., 2000; Mendell et al., 2002). UPF2 localizes around the nucleus and is known to bind UPF3 and UPF1 (He et al., 1997; Lykke-Andersen et al., 2000). The two human UPF3 proteins (UPF3a and UPF3b) are found predominantly within the nucleus and associate preferentially with spliced RNAs, possibly through interactions with RNPS1 and/or Y14:MAGOH (Lykke-Andersen et al., 2000, 2001; Kim et al., 2001b; Le Hir et al., 2001; Serin et al., 2001; Gehring et al., 2003). The interactions of the translation termination factors eRF1 and eRF3 with UPF1 and of eRF3 with UPF2 and UPF3 provide the link between NMD and translation termination (reviewed in Wagner and Lykke-Andersen, 2002).

The *smg*1 gene encodes a phosphoinositide-3-kinase-related protein kinase required for the phosphorylation of UPF1 (Page et al., 1999; Denning et al., 2001; Pal et al.,
complex is then targeted for rapid degradation (reviewed in Wagner and Lykke-Andersen, 2002).

Although the NMD pathway is becoming increasingly well understood in yeast, the molecular mechanism of NMD in higher eukaryotes remains unclear. In particular, the degradation route followed by aberrant mRNAs in metazoan is unknown. Moreover, the yeast pathway differs from higher eukaryotic NMD in several aspects. First, phosphorylation of Upf1p does not appear to occur in yeast. Secondly, there are no obvious SMG5 or SMG7 orthologs encoded by the yeast genome. Thirdly, in yeast, premature stop codons are defined independently of exon boundaries. Since most genes in higher eukaryotes have introns, and components of the EJC are conserved, it is assumed that exon boundaries would play a critical role in defining premature stop codons in metazoan.

In order to shed light on the NMD pathway in higher eukaryotes, we have generated Drosophila cell lines constitutively expressing PTC-containing mRNAs. Using double-stranded RNA interference (RNAi), we show that the Drosophila orthologs of UPF1, UPF2, UPF3 and SMG1 are required for degradation of the NMD reporters. Database searches reveal that orthologs of the C. elegans proteins SMG5–7 are encoded by all animal genomes sequenced so far, but not by fungi, indicating that SMG5–7 are a specific feature of the NMD pathway in higher eukaryotes. Interestingly, the Drosophila genome encodes orthologs of SMG5 and SMG6, but not of SMG7. We show that SMG5 and SMG6 are both required for NMD in Drosophila cells. Unexpectedly, the definition of PTCs in Drosophila occurs by a novel mechanism that is independent of exon boundaries and may not rely on specific DSEs. Our findings reveal that despite conservation of the NMD machinery, different mechanisms have evolved to discriminate premature from natural stop codons in metazoan.

Results

Generation of an NMD reporter for Drosophila cells

Mutations in the alcohol dehydrogenase gene (adh) of Drosophila that generate PTCs lead to reduced levels of cytoplasmic and nuclear mRNA, and to mRNAs with poly(A) tails longer than the wild-type transcript (Broagna, 1999). To investigate whether adh alleles carrying PTCs are degraded by the NMD pathway and to gain further insight into the mechanism of this pathway in higher eukaryotic cells, we designed a reporter gene in which the genomic coding region of Drosophila adh is placed downstream of the constitutive actin 5C promoter and upstream of a polyadenylation site derived from SV40 (Figure 1A). In adult flies, transcription of the adh gene produces a pre-mRNA with three introns (Figure 1A). The first intron (IN-1) is located in the 5’-untranslated region (5’-UTR), while introns 2 and 3 (IN-2 and IN-3) lie between codons 33 and 34, and 168 and 169, respectively. The complete coding sequence has 257 codons (Broagna, 1999). In the reporter constructs used in this study, the adh 5’- and 3’-UTRs were deleted (Figure 1A).

Two previously characterized PTCs were introduced in the adh reporter by site-directed mutagenesis. These correspond to adh-nBR114 and adh-n4, which carry
mRNAs degraded through the NMD pathway are stabilized following inhibition of translation (see Zhang et al., 1998b). When cells constitutively expressing the adh reporters were treated with cycloheximide for 45 min, the steady-state levels of n and nBR114 mRNAs were increased 2.5- and 2.8-fold, respectively (Figure 1D).

To assess definitively whether adh-nBR114 and adh-n4 were subjected to NMD, we investigated the effect of depleting endogenous UPF1 or UPF2 on the expression levels of these mRNA species. S2 cells expressing adh-wt or the PTC mutants were transfected with dsRNAs specific for Dm UPF1 or Dm UPF2. To ensure that the proteins were depleted efficiently, all RNAi experiments described in this study were carried out by transfecting cells twice with the corresponding dsRNAs, with the second transfection 4 days after the first on day 0. The expression levels of the adh mRNAs were analyzed on day 7. Depletion of either UPF1 or UPF2 resulted in a 4- to 5-fold increase of adh-n4 or adh-nBR114 mRNA levels (Figure 2B and C). The expression levels of the wild-type mRNA were not significantly changed (Figure 2A). Together, these results indicate that adh-n4 and nBR114 mRNAs are degraded through the NMD pathway.

Orthologs of SMG5–7 are encoded by all sequenced animal genomes

Having established a reporter assay that reproduces bona fide NMD, we next investigated whether the Drosophila orthologs of UPF3, SMG1 and SMG5–7 are involved in this process. Using BLAST searches (Altschul et al., 1997), we identified single genes encoding SMG1 and UPF3 in Drosophila genomic sequences (Adams et al., 2000). Furthermore, the products of the Drosophila (Dm) genes CG8954 (anony-34Ea) and CG6369 were reported to be related to C.elegans (Ce) SMG7 and SMG5 (Clissold and Ponting, 2000; Chiu et al., 2003). Dm CG8954, Dm CG6369, Ce SMG5 and Ce SMG7 are characterized by the presence of one and a half N-terminal TPRs. With the exception of Ce SMG7, these proteins have, in addition, a C-terminal PIN domain (Clissold and Ponting, 2000).

The orthologous relationships between these proteins had not been assigned in previous studies (Clissold and Ponting, 2000; Chiu et al., 2003); we therefore carried out database searches in order to retrieve from the completely sequenced genomes all proteins with a TPR-like repeat at the N-terminus and a PIN domain at the C-terminus. A multialignment of the retrieved sequences revealed that the conserved region of ~50 residues encompassing the identified TPR repeats (Clissold and Ponting, 2000) could be expanded considerably (Figure 3A and B) and was predicted to be entirely helical over a length of ~150 residues (Figure 3B).

Using the expanded TPR-like helical region as a query, two more proteins were retrieved from the database that did not contain a PIN domain: Ce SMG7 and Hs KIAA0250. A multiple alignment of the family revealed two distinct orthologous groups across the species considered: one includes Ce SMG5, Hs KIAA1089 and Dm CG8954, and the other Hs KIAA0732, Dm CG6369 and an additional Ce protein encoded by the Y54F10AL-2a region. This protein corresponds to Ce SMG6 (P. Anderson, personal communication). Only for the proteins without PIN domains could the orthologous

PTCs at codons 64 and 83, respectively (Broguna, 1999). The reporter constructs were co-transfected with a plasmid encoding puromycin acetyl transferase into Drosophila Schneider cells (S2 cells) to generate polyclonal cell lines constitutively expressing wild-type adh (adh-wt) or one of the two PTC mutants. The steady-state levels of the corresponding transcripts were analyzed by northern blot. In agreement with previous observations (Broguna, 1999), adh-wt mRNA had a higher electrophoretic mobility and was expressed at higher levels than the adh-n4 and adh-nBR114 mRNAs (Figure 1B, lanes 1, 8 and 15). In all experiments performed in Drosophila cells, the levels of the NMD reporters were normalized to those of the endogenous rp49 mRNA (encoding ribosomal protein L32), which is a long-lived mRNA (half-life >8 h). Following this normalization, the levels of PTC-containing transcripts were reduced to 25–30% of wild-type adh levels.

We next determined the stabilities of the adh transcripts. In the experiments shown in Figure 1B, transcription was inhibited by actinomycin D, cells were harvested at the indicated time points and the adh mRNA levels were determined. The half-life of adh-wt was ~3 h, while the half-lives of nBR114 or n4 mRNAs were <15 min (Figure 1B and C). Thus, the overall reduction of the steady-state levels of the PTC-containing mRNAs is a direct consequence of their higher turnover rate.
Fig. 3. Phylogenetic tree and domain organization of the SMG5-7 proteins. (A) The two identified orthologous groups are shown in red. Only one and a half TPR repeats can be identified automatically and are indicated in dark blue within the conserved helical N-terminal domain (cyan). The *Caenorhabditis* sequences are underlined and are known to be involved in NMD (Cali et al., 1999; Page et al., 1999; Anders et al., 2003). Human and *Drosophila* sequences characterized in this study are shown in gray boxes. The vertical black bars on the schematic of the protein domains indicate that part of the protein sequence is not drawn to scale. The number of amino acid residues is shown on the right. Scale bar: 200 amino acids. TPR, tetratricopeptide repeats; PIN, Pit-T-terminus; Ce, *Caenorhabditis elegans*; Ci, *Ciona intestinalis*; Dm, *Drosophila melanogaster*; Fr, Fugu rubripes; Hs, *Homo sapiens*. (B) Multiple sequence alignment of the TPR-like region. The accession number of each sequence is indicated, as well as the species and the start and end points of the sequences. For the sequences of *C. intestinalis* and *F. rubripes*, no accession numbers are shown, as the predictions were performed directly on the genomic sequences using GeneWise (Birney et al., 1998). The consensus in 70% of the sequences is shown below the alignment. H, L, p, +, – represent hydrophobic, aliphatic, polar, positive and negative residues, respectively. Aliphatic residues are highlighted in cyan, hydrophobic in blue, polar in green, negative in pink, and positive in red. The conserved glycines are highlighted in orange. The secondary structure prediction (Sec.Str.Pred.) is taken from the consensus of the multiple alignment (H, helices predicted with an accuracy >82%; h, helices predicted with an accuracy ≤82%). The blue bars above the sequences show the position of the one and a half TPR repeats identified by Clissold and Ponting (2000).
The Drosophila orthologs of UPF3, SMG1, SMG5 and SMG6 are essential for NMD

To assess whether Dm UPF3, SMG1, SMG5 and SMG6 are involved in NMD, we silenced the expression of these genes by RNAi in cells expressing adh-wt or adh-n4 transcripts. The effect of these depletions on the mRNA levels was analyzed by northern blot and compared with the effects observed when UPF1 or UPF2 were depleted. Adh-n4 mRNA levels increased 2.8- to 6-fold in cells depleted of UPF3, SMG1, SMG5 or SMG6 (Figure 4A).

The steady-state levels of adh-wt and adh-n4 mRNAs were quantitated in at least three independent experiments and normalized to those of rp49 mRNA. The normalized values obtained for the PTC-containing mRNA were divided by those obtained for adh-wt in order to correct for potential non-specific effects of the depletions. This ratio was set arbitrarily to a value of 1 in control cells. The up-regulation of adh-n4 mRNA was within the 2- to 4.5-fold range (Figure 4B). Similar values were reported for PTC-containing transcripts in human cells depleted of UPF1 or Y14 (Mendell et al., 2002; Gehring et al., 2003).

Because the steady-state levels of wild-type adh mRNA also increased in Smg6-depleted cells (Figure 4A), it was important to confirm that the increase in adh-n4 mRNA levels was due to a specific inhibition of the NMD pathway. We therefore tested whether the half-life of adh-n4 mRNA was specifically increased in cells depleted of SMG6. As a positive control, we measured the stability of this transcript in cells depleted of UPF1. The half-life of adh-n4 mRNA increased from <15 min in control cells to 100 min in UPF1- or SMG6-depleted cells (Figure 4C), indicating that the overall increase of the steady-state levels of this mRNA reflects a decrease of its turnover rate. The half-life of adh-wt was ~3 h both in control cells and in cells depleted of UPF1 or SMG6 (Figure 4C). These results indicate that SMG1, SMG5 and SMG6 are essential for NMD in Drosophila.

Depletion of SMG5–7 in human cells leads to the stabilization of PTC-containing mRNAs

Although Hs SMG6 (KIAA0732) has been shown to function in the dephosphorylation of UPF1 (Chiu et al., 2003), evidence for a role for Hs SMG6 in telomere maintenance has now been reported (Reichenbach et al., 2003), raising the possibility that not all orthologs of SMG5–7 play a role in NMD. We therefore investigated whether Hs SMG5–7 are directly implicated in this process by analyzing the effect of depleting these proteins on the steady-state levels of a β-globin mRNA carrying a PTC at codon 39 (NS39; Thermann et al., 1998).

HeLa cells were transfected with short interfering RNAs (siRNAs) specific for SMG5, SMG6 or SMG7. An siRNA specific for UPF1 served as a positive control. After 48 h, cells were retransfected with the same siRNAs and plasmids expressing either wild-type β-globin or β-globin-NS39 mRNA. A plasmid expressing a green fluorescent protein (GFP) fusion (GFP–NFX1) served as a transfection control (Figure 5). The levels of the mRNAs targeted by RNAi were reduced to 30% of the levels in control cells, as estimated by RT–PCR (data not shown).

Normalized to the levels of the transfection control, the levels of the NS39 transcript were reduced in abundance to ~10% of the wild-type β-globin levels, as reported before relationships not be assigned. The phylogenetic tree obtained for the N-terminal regions was identical to that obtained by using the PIN domains (data not shown).

Based on the domain organization and on the functional characterization described below, we conclude that KIAA0250 is likely to represent the human ortholog of SMG7, but that there is no SMG7 ortholog in Drosophila. As the phylogenetic tree with the PIN domain indicates that KIAA1089 and CG8954 are closely related to Ce SMG5, we propose to call these proteins Hs and Dm SMG5. KIAA0732, CG6369 and Y54F10AL-2 define another branch of the family, and we propose to call them Hs, Dm and Ce SMG6.
Nonsense-mediated mRNA decay in Drosophila

The Drosophila orthologs of all known components of the vertebrate EJC are dispensable for NMD

The conservation of the NMD machinery in higher eukaryotes prompted us to investigate whether the mechanism by which PTCs are defined is also conserved. To this end, we first investigated whether Drosophila EJC proteins, and in particular Y14 and RNPS1, are implicated in this process. Although it has not yet been demonstrated that the EJC is assembled in Drosophila, orthologs of all known components of the vertebrate EJC are present in this organism (Gatfield and Izaurralde, 2002). S2 cells expressing adh-wt or adh-n4 were treated with dsRNAs corresponding to the coding sequences of all known EJC proteins (i.e. REF1, RNPS1, SRm160 and Y14). The protein DEK was also depleted, although recent studies suggest that DEK may not be a bona fide component of the EJC (Gatfield and Izaurralde, 2002). As positive controls, dsRNAs corresponding to UPF1 and UPF2 were transfected. Because the EJC proteins might have partially redundant functions, we also simultaneously depleted Y14 and RNPS1 or, in addition, REF1 and SRm160. Western blot analysis revealed that 7 days after addition of dsRNAs, the expression levels of EJC proteins were ≤10% of the endogenous levels when they were depleted individually (Figure 6A and B, lanes 4–8). The depletion efficiency was reduced when more than one gene was silenced simultaneously (Figure 6A and B, lanes 9 and 10).

Co-depletions of the EJC proteins severely impair cell proliferation (see Gatfield and Izaurralde, 2002). Nonetheless, no stabilization of adh-n4 mRNA was observed in cells depleted of single or multiple EJC proteins (Figure 6C–E). A major concern raised by the experiments described above was that the EJC proteins may have additional functions in mRNA metabolism that could mask the stabilization of adh-n4 mRNA. To address this issue, we simultaneously depleted Y14 or RNPS1 together with UPF1. We observed that in cells depleted of Y14 or RNPS1, depletion of UPF1 stabilized adh-n4 mRNA (Figure 6F). Similarly, depletion of RNPS1 and Y14 proteins together with Upf3 resulted in a stabilization of adh-n4 mRNA similar to that observed when only Upf3 was depleted (data not shown). We conclude that the EJC proteins are not essential for the degradation of adh-n4 mRNA.

PTCs can be defined independently of exon boundaries in Drosophila

The observation that the expression levels of adh-n4 mRNA were not affected in cells depleted of EJC proteins raised the question as to whether exon–exon boundaries play a role in defining PTCs in Drosophila. To address this question, we deleted intron 3 (IN-3), which is the 3′-most intron in our reporters (Figure 7A), and generated cell lines expressing adhIN3-wt, adhIN3-n4 or adhIN3-nBR114 mRNAs.

Normalized to the levels of the endogenous rp49 mRNA, the levels of PTC-containing transcripts in which intron 3 had been deleted were reduced to 25–30% of the corresponding wild-type adhIN3 levels (Figure 7B). Moreover, while the half-life of adhIN3-wt was ~90 min, the half-life of adhIN3-nBR114 mRNA was ~15 min (control panels in Figure 7D–F). To assess whether adhIN3-n4 or adhIN3-nBR114 were subjected to NMD, we suppressed the expression of UPF1, UPF2 or UPF3 genes in cells constitutively expressing these reporters. Silencing of either of the UPF genes resulted in a 3- to 5-fold increase of the steady-state levels of the PTC-containing mRNAs lacking intron 3. The expression levels of the wild-type mRNA were not significantly changed (Figure 7B). The stabilization of the PTC-containing mRNAs lacking intron 3 was therefore within the same range as that observed for the corresponding transcripts carrying intron 3 (Figure 7C versus 4B).

The increased steady-state levels of the PTC-containing mRNAs lacking intron 3 in cells depleted of UPF1 is due
to a decrease of their turnover rate. Indeed, the half-life of adhΔIN3-n4 and of adhΔIN3-nBR114 mRNAs increased to ~1 and 2 h, respectively, following UPF1 depletion (Figure 7E and F). The half-life of the corresponding wild-type transcript remained unchanged (Figure 7D). In summary, at least in the case of the adh mRNA, PTCs can be recognized by the *Drosophila* NMD machinery regardless of any downstream exon boundary.

**The Drosophila NMD machinery discriminates PTCs from natural stops on heterologous mRNAs**

The exon boundary-independent recognition of PTCs described above could either be specific for the adh mRNA or could reflect a general feature of the NMD pathway in *Drosophila*. To discriminate between these possibilities, we generated two reporter genes in which either the CAT- (chloramphenicol acetyltransferase) or the GFP-coding regions were placed downstream of the constitutive actin 5C promoter. Both genes have no introns. The NMD reporters were generated by introducing a PTC at codon 72 and 52 of the CAT- and GFP-coding sequences, respectively. Polyclonal cell lines constitutively expressing wild-type or PTC-containing mRNAs were generated. Northern blot analysis showed that the levels of CAT-72 mRNA and GFP-52 mRNA were reduced in abundance to 30 and 15% of the corresponding wild-type levels (Figure 8A and B). For the GFP reporter, we confirmed that the reduction of the steady-state levels of GFP-52 mRNA correlated with an increase in its turnover rate. Indeed, the half-life of GFP-52 mRNA was ~10 min, while the half-life of the wild-type GFP transcript was >4 h (data not shown).

Depletion of either UPF1, UPF2 or UPF3 resulted in a 3- to 4.5-fold increase of the steady-state levels of the
CAT-72 and GFP-52 mRNAs (Figure 8A–C), in agreement with the results obtained for the adh reporters. Surprisingly, a significant reduction in GFP-52 mRNA levels was consistently observed in UPF1-depleted cells (Figure 8B and C). Although the basis for this decrease is not known, it is possible that this effect reflects an additional function of UPF1 distinct from NMD, as GFP-52 mRNA levels were not significantly changed by UPF2 or UPF3 depletion (Figure 8B–D). Moreover, depletion of SMG1, SMG5 or SMG6 also led to a specific stabilization of GFP-52 mRNA (Figure 8D). We conclude that PTCs can be recognized by the *Drosophila* NMD machinery in heterologous mRNAs lacking a downstream exon boundary. These results provide strong support for the conclusion that the components of the EJC are dispensable for NMD in this organism.

**Discussion**

The NMD pathway in *Drosophila* provides a snapshot into the evolution of this conserved mRNA surveillance mechanism, as it has features in common with both the yeast and the mammalian pathways. In common with the yeast pathway is the observation that PTCs can be defined independently of exon boundaries. In contrast to the yeast pathway, however, NMD in *Drosophila* requires the products of the *SMG1, SMG5* and *SMG6* genes, which are a specific feature of higher eukaryotic NMD.

**SMG5–7 proteins are a specific feature of the metazoan NMD pathway**

Clissold and Ponting (2000) reported the existence of a protein family with N-terminal TRP repeats and a C-terminal PIN domain. Our mutational analysis of all sequences possessing these domains reveals that the conserved region detected by Clissold and Ponting (2000), which encompasses one and a half TRPs, can be extended to a region of ~150 residues which is predicted to be helical over its entire length. We refer to this extended region as the TRP-like domain. TRP repeats are protein–protein interaction modules present in proteins with unrelated functions. As TRP motifs usually require more than one repeat for structural stability (Goebel et al., 1991), we predict that the region of extended similarity might contain several TRP repeats adapted to a specialized function in this protein family, so that standard computer programs may be unable to reveal statistically significant sequence similarities to known TRPs.

Previous studies identified three human proteins exhibiting limited sequence similarity to Ce SMG5 and SMG7 (Clissold and Ponting, 2000; Chiu et al., 2003; Reichenbach et al., 2003), but the orthologous relationships between these proteins had not been assigned. For that reason, Chiu et al. (2003) named KIAA0732, KIAA1089 and KIAA0250, human SMG5/7a, SMG5/7b and SMG5/7c, respectively. The nomenclature proposed by Chiu et al. (2003) is not supported by the phylogenetic trees based on the alignment of the TRP-like domain and of the PIN domain presented in this study. Indeed, the

![Fig. 7](image-url) PTCs are defined independently of downstream exon boundaries in *Drosophila* cells. (A) Schematic representation of the reporters lacking a downstream intron. Symbols are as in Figure 1A. (B) S2 cells expressing adh-wt, adh-n4 or adh-nBR114 reporters lacking intron 3 (adhΔIN3-wt, adhΔIN3-n4, adhΔIN3-nBR114) were treated with UPF1, UPF2 or UPF3 dsRNAs. Total RNA samples were analyzed as described in Figure 2. (C) The steady-state levels of adhΔIN3-n4 and adhΔIN3-nBR114 mRNAs were quantitated in at least three independent experiments and normalized to those of rp49 mRNA. These normalized values (adhΔIN3-PTC norm.) were divided by those obtained for adhΔIN3-wt. The mean values ± SDs are shown. (D–F) S2 cells constitutively expressing adhΔIN3-wt, adhΔIN3-n4 or adhΔIN3-nBR114 were treated with UPF1 dsRNA. Seven days after addition of UPF1 dsRNA, treated or untreated cells (control) were incubated with actinomycin D (5 μg/ml) for the times indicated above the lanes. Total RNA samples were isolated and analyzed by northern blot.
phylogenetic trees clearly define two different orthologous groups, the SMG5 family and the SMG6 family. As one gene of each organism falls into the same branch of the tree, they are likely to perform equivalent functions in these organisms, distinct from those of the second ortholog, but all involved in NMD. The nematode and human genomes encode in addition SMG7, but there is no obvious SMG7 ortholog in Drosophila. We therefore propose to call KIAA0732, KIAA1089 and KIAA0250, human SMG6, SMG5 and SMG7, respectively. Analogously, CG8954 and CG6369 will be named Drosophila SMG5 and SMG6, respectively. Given the distribution of proteins within the SMG5–7 families that have been shown to be involved in NMD (Cali et al., 1999; Page et al., 1999; Chiu et al., 2003; this study), we predict that all orthologs are likely to have a role in NMD, including those identified in Fugu rubripes and Ciona intestinalis (Figure 3).

Despite the similarities in domain organization, the SMG5–7 proteins do not appear to have redundant functions, as they are all required for the destabilization of PTC-containing transcripts. On the other hand, mutations in Ce smg-5, smg-6 or smg-7 genes lead to the accumulation of the phosphorylated form of UPF1 (Page et al., 1999), suggesting that all three proteins function in the dephosphorylation of UPF1. Consistently, Anders et al. (2003) have shown recently that Ce SMG5 is part of a protein complex including SMG7, UPF1, PP2Ac (the catalytic subunit of protein phosphatase 2A) and PR65 (a structural subunit of protein phosphatase 2A). Similarly, Chiu et al. (2003) have reported that human SMG6 co-immunoprecipitates with UPF1-3, SMG1 and PP2Ac. Together, these observations suggest that SMG5–7 are components of a conserved multiprotein complex, whose role is to direct PP2A to its UPF1 substrate. Depletion of any of the SMG5–7 proteins may lead to the disassembly of the complex and hence a similar phenotype, even though each protein probably has a distinct and specific function. Understanding the precise role of the SMG5–7 proteins in NMD awaits further biochemical characterization. In this context, the recently reported role of Hs SMG6 in telomere regulation (Reichenbach et al., 2003) suggests that some of these proteins have acquired additional functions distinct from NMD. Alternatively, since NMD regulates the expression levels of telomere components in yeast (Dahlseed et al., 2003), the possibility that SMG6 indirectly affects telomere maintenance in human cells cannot be ruled out.

Y14 and RNPS1 are dispensable for NMD in Drosophila

Human Y14-MAGOH and RNPS1 have been shown to elicit NMD when tethered downstream of a stop codon in human cells (Lykke-Andersen et al., 2001; Fribourg et al., 2003; Gehring et al., 2003). Moreover, depletion of Y14 from human cells by RNAi leads to the stabilization of PTC-containing transcripts, suggesting an essential role for Y14 in NMD (Gehring et al., 2003). Therefore, the observation that the Drosophila orthologs of vertebrate EJC proteins, and in particular Y14 and RNPS1, are dispensable for NMD reflects a specific feature of the Drosophila pathway. This observation is fully consistent with the observation that exon boundaries are not required to define PTCs in this organism.

Definition of premature stop codons in Drosophila

The observation that in the absence of a downstream exon boundary, PTC-containing mRNAs are nonetheless targeted for degradation raises the question of what differentiates a premature termination codon from a normal one.
in *Drosophila*. One possibility is that PTCs are defined relative to a DSE, as in yeast. Alternatively, a ‘fail-safe’ sequence may function instead of exon boundaries when the downstream intron is removed, as reported by Zhang et al. (1998a,b) for the β-globin and triosephosphate isomerase mRNAs. The observation that in-frame stop codons in the CAT or GFP mRNAs can be defined as premature by the *Drosophila* NMD machinery makes it unlikely that specific sequence elements mark *Drosophila* transcripts. However, because of the lack of a strong consensus among yeast DSEs or mammalian fail-safe sequences, we cannot rule out that the CAT or GFP mRNAs possess such sequence elements.

It is also possible that a generic feature of the mRNA, such as the poly(A) tail or a mark deposited during the cleavage and polyadenylation reaction, provides the positional information needed to discriminate premature from natural stop codons in *Drosophila*. The latter possibility is consistent with a model in which it is the environment of the translation termination codon that distinguishes a premature from a normal stop codon (Hilleren and Parker, 1999). According to this model, 3′-UTRs would be marked by a specific set of proteins that may become associated with the mRNA during 3′ end formation. If a terminating ribosome is able to interact with these 3′-UTR-bound proteins, proper termination can occur. If the termination process is impaired or too slow, because the terminating ribosome is unable to establish these interactions, the NMD complex could be assembled, leading to the rapid degradation of the mRNA.

**Materials and methods**

*Drosophila* NMD reporter constructs and cell lines

The adh sequence (accession No. X78384; residues 2149-2923) was amplified from a genomic clone and inserted between the EcoRI and XhoI sites of vector pAc5.1b (Invitrogen). The N4 and nBRI14 mutations were inserted by site-directed mutagenesis using the Quick-change site-directed mutagenesis kit (Stratagene). Intron 3 was excised from the corresponding constructs by site-directed mutagenesis. A plasmid to allow the expression of GFP in S2 cells was generated by inserting the corresponding cDNA into a PBS-based vector containing the *Drosophila* actin promoter followed by multiple cloning sites and the BgIII terminator (pB SacGFP). GFP-52 was derived from plasmid pBSacGFP by changing codon 52 (GGC) to a stop codon (TGA) in the GFP sequence. A plasmid to allow the expression of the CAT gene in S2 cells was generated by inserting the corresponding cDNA between the KpnI and XhoI sites of vector pAc5.1b. CAT-72 was generated by changing codon 72 (GAA) to a stop codon (TAA) in the CAT sequence. All constructs were sequenced completely to confirm the presence of the PTCs and the absence of additional mutations. Stable *Drosophila* cell lines expressing wild-type and PTC-containing transcripts were established by co-transfecting the corresponding constructs with a plasmid encoding puromycin acetyl transferase (pBS-Puro; Benton et al., 2000) at a 1:5 ratio. S2 cells were transfected with Lipofectin (Invitrogen) according to the manufacturer’s instructions and selected in medium containing 10 μg/ml puromycin.

**Cloning of Drosophila cDNAs**

*Drosophila* genes described in this study correspond to the following genes in FlyBase (http://flybase.bio.indiana.edu): UPF1 (CG1559), UPF2 (CG2253), UPF3 (CG11184), SMG1 (CG4549), SMG3 (CG8954, anon-34Ea) and SMG6 (CG6369). Full-length *Drosophila* UPF1 and UPF3 cDNAs were amplified by PCR using primers introducing unique restriction sites and a random-primed S2 cell cDNA as a template. For UPF2, SMG1, SMG3 and SMG6, cDNA fragments comprising only the first 700 nucleotides were cloned. All PCRs were performed with the Expand high-fidelity PCR system (Roche). The amplified cDNAs were cloned into pGEXCS and sequenced. Cloning of cDNAs encoding *Drosophila* orthologs of EJC proteins has been described before (Galfeild and Izaurralde, 2002).

**Western blotting and double-stranded RNA interference in cultured Drosophila cells**

Preparation of total cell extracts, western blotting and RNAi were performed essentially as described by Galfeild and Izaurralde (2002). DsRNAs corresponding to EJC proteins have also been described (Galfeild and Izaurralde, 2002). Other dsRNAs used in this study correspond to fragments comprising nucleotides 1–700 of the predicted cDNAs. A 30 μg aliquot of dsRNA was used per 6-well dish containing ~2 × 10^6 cells. Depletions described in this study were carried out by transfecting cells with the corresponding dsRNAs on day 0 and day 4. Cells were harvested on day 7.

**Double-stranded RNA interference in HeLa cells**

HeLa cells were grown in Dulbecco’s modified Eagle’s medium (DMEM) supplemented with 10% fetal bovine serum, 100 U/ml penicillin and 100 μg/ml streptomycin. Depletion of endogenous proteins by RNAi was carried out using siRNAs as described by Elbashir et al. (2001), except that cells were transfected with Lipofectamine Plus reagent (Invitrogen) following the manufacturer’s instructions. The following target sequences were used: UPF1, AAGATGAGTTCGCT- CCATT; SMG5, AAGTCCTCGAGATGCTC; SMG6, AAGC- AGTGTACAGGCAAT; SMG7, AACAGCAGCTGTCACAGCCA; and control, AAAGCGGCAACGCAAC. The siRNAs were transfected at a final concentration of 70 nM. At 24 h post-transfection, cells were trypsinized and seeded in 6-well plates. After a further 24 h, the cells were transfected with siRNAs (as above) together with 0.5 μg of each β-globin reporter plasmid (β-globin-wt or β-globin-NS39) and 0.5 μg of pEGFP-C1-NXFI as a transfection control. The β-globin reporters have been described by Thermann et al. (1998). Cells were harvested 48 h after the second transfection.

**RNA isolation and northern blots**

Total RNA was isolated using TRIzol reagent (Life Technologies), separated in denaturing formaldehyde agarose gels (5–20 μg/lane) and blotted onto positively charged nylon membranes (GeneScreen Plus, NEN Life Science). 32P-labeled probes were generated by random priming or linear PCR using standard methods. Hybridizations were carried out overnight at 65°C in Church buffer (0.5 M Na-phosphate pH 7; 7% SDS and 1 mM EDTA). Following hybridization, the membranes were washed four times 15 min at 65°C in 40 mM Na-phosphate pH 7; 1% SDS and 1 mM EDTA.

**Sequence analysis**

Sequence similarity searches were performed using PSI-BLAST (Altschul et al., 1997) with Ce SMG-7 and SMG-5 as input sequences. These searches did not converge due to the presence of widespread domains such as TRP and PIN. Nevertheless, a group of putatively related sequences could be defined on the basis of the presence of both domains. The relationship between these proteins was confirmed by using each of the sequences as an input for PSI-BLAST and retrieving all the other members of the group. The human sequence KIA0250 (BAC53621) was joined to the sequences bearing the two domains, as it was retrieved in all the searches. As the identified proteins show no sequence conservation outside of the TRP-like and PIN domains, two different multialignments and the corresponding bootstrap trees were built with ClustalW (Thompson et al., 1997) using only these domains. Because of the low degree of conservation of the extended TRP-like region, the bootstrap values for this tree were not as good as those obtained for the PIN domain.

**Acknowledgements**

We gratefully acknowledge Saverio Brogna for providing a genomic adh clone, Philip Anderson for providing information on the identity of *C.elegans* SMG6, Philipp Bucher for providing unpublished results on the alignment of the SMG-7 proteins, Kevin Craplinski and Matthias Hentze for helpful and stimulating discussions, and David Thomas for critical reading of the manuscript. This study was supported by the European Molecular Biology Organization (EMBO).
References


Received March 31, 2003; revised May 28, 2003; accepted June 2, 2003