Biochimica et Biophysica Acta 1779 (2008) 538–549

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagrm

Review

Recognition and elimination of nonsense mRNA

Oliver Mühlemann *, Andrea B. Eberle, Lukas Stalder, Rodolfo Zamudio Orozco

Institute of Cell Biology, University of Berne, Baltzerstrasse 4, CH-3012 Bern, Switzerland

ARTICLE INFO

Article history:
Received 3 January 2008
Received in revised form 30 May 2008
Accepted 30 June 2008
Available online 8 July 2008

Keywords:
mRNA surveillance
Nonsense-mediated mRNA decay
Premature translation
Termination
mRNA turnover
mRNA quality control

ABSTRACT

Among the different cellular surveillance mechanisms in charge to prevent production of faulty gene products, nonsense-mediated mRNA decay (NMD) represents a translation-dependent posttranscriptional process that selectively recognizes and degrades mRNAs whose open reading frame (ORF) is truncated by a premature translation termination codon (PTC, also called "nonsense codon"). In doing so, NMD protects the cell from accumulating C-terminally truncated proteins with potentially deleterious functions. Transcriptome profiling of NMD-deficient yeast, *Drosophila*, and human cells revealed that 3–10% of all mRNA levels are regulated (directly or indirectly) by NMD, indicating an important role of NMD in gene regulation that extends beyond quality control [J. Rehwinkel, J. Raes, E. Izaurralde, Nonsense-mediated mRNA decay: Target genes and functional diversification of effectors, Trends Biochem. Sci. 31 (2006) 639-646. [1]]. In this review, we focus on recent results from different model organisms that indicate an evolutionarily conserved mechanism for PTC identification.

1. Sources of PTCs

PTCs arise from mutations in the DNA, but also on the RNA level (Table 1). Many DNA mutations within a gene will truncate the ORF. In addition to nonsense mutations, i.e. base substitutions that directly generate PTCs by changing an amino acid-encoding codon into one of the three termination codons (UAA, UAG, UGA), random nucleotide insertions and deletions shift in two of three cases the reading frame, where within the next 20 codons on average a termination codon will prematurely terminate translation. Often mutations also alter splicing signals and generate alternatively spliced mRNAs, many of which contain a PTC. Overall, it is estimated that about 30% of all known disease-associated mutations generate a PTC-containing (PTC+) mRNA [2,3].

In addition to the sources described above, PTCs arise very frequently in genes belonging to the immunoglobulin superfamily (immunoglobulins, T-cell receptors) as a consequence of programmed V(D)J rearrangements during lymphocyte maturation [4]. During the joining of a V, a D (only in heavy chains), and a J fragment, non-templated nucleotides (N nucleotides) can be added by the enzyme terminal transferase and coding nucleotides from the opposite strand (P nucleotides) can be transferred to the coding strand at the junctions of the segments. The ORF can only be maintained when nucleotides in multiples of three (i.e. three or six) are added, but in two thirds of the rearrangements, a frameshift results in a nonproductive allele that encodes a PTC+ t1ranscript. Interestingly, it was observed that PTCs in genes belonging to the immunoglobulin superfamily cause a much stronger reduction of the steady-state mRNA level by NMD than in other genes [5–8].

On the RNA level, errors during transcription and alternative pre-mRNA splicing generate PTC+ mRNAs. Based on a misincorporation rate for RNA polymerase II in the order of 10^{-5} per nucleotide, and assuming 10^3 to 10^4 coding nucleotides in a typical gene, only 0.05% to 0.5% of all transcripts are estimated to acquire a PTC through transcription errors. In contrast, the fraction of PTC+ transcripts generated by unproductive alternative pre-mRNA splicing is much larger. Computational analysis of human EST databases revealed that among the 60%–70% of human pre-mRNAs that are alternatively spliced, 45% had at least one splice form that was predicted to be a target of NMD [9]. Thus, about one third of all human protein-coding genes produce a PTC+ mRNA, and although their exact abundance is not known, they are likely to represent a significant fraction in the pool of the initially produced mRNAs.

2. *Trans*-acting factors involved in NMD

The first *trans*-acting factors involved in NMD have been identified in genetic screens in *Saccharomyces cerevisiae* and *Caenorhabditis elegans*. In screens for translational suppressors in *S. cerevisiae*, mutations in the three genes UPF1, UPF2/NMD2 and UPF3 (for *Up-f*rameshift) were found to decrease decay rates of PTC+ mRNAs and to promote readthrough of PTCs [10–13] (Table 2). Three labs had identified in independent screens in *C. elegans* loss-of-function mutations in seven genes called SMG1 to SMG7 (for *s*uppressor with *m*orphogenetic effects on genitalia) that several years later were

* Corresponding author.
E-mail address: oliver.muehlemann@izb.unibe.ch (O. Mühlemann).

1874-9399/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbagrm.2008.06.012

Table 1
Features and origins of NMD targets

Aberrant mRNAs	
Problem at DNA level	
Nonsense mutations	Base substitutions that directly generate PTCs.
Insertions and deletions	Random nucleotide insertions and deletions shift the reading frame in two of three cases, resulting in a PTC.
Mutations changing splicing signals	Mutations leading to aberrant splicing often result in a frameshift.
VDJ rearrangement	The immunoglobulin superfamily represents a special class of NMD targets that undergo very efficient NMD. Two of three rearrangements of the V, D, and J segments result in a frameshift.
Problem at RNA level	
Transcription errors	Frequency low, cause premature ORF truncation in < 1% of transcripts.
Unproductive alternative splicing	45% of alternatively spliced mRNAs are predicted to be an NMD target.
Problem at translation level	
Leaky scanning	Observed only in yeast. Ribosomes scan beyond the initiator AUG and initiate at a downstream AUG in a reading frame with a PTC.
Physiological mRNAs	
Programmed translational frameshifting	Programmed + 1 or − 1 frameshifts lead into a PTC, if the ribosome fails to shift the reading frame properly.
mRNAs encoding selenoproteins	UGA can be recognized as codon for selenocysteine or as PTC, depending on endogenous selenium concentration.
mRNAs with uORFs	The termination codon of the uORF is likely to be interpreted as PTC, unless the mRNA harbors stabilizing elements nearby.
mRNAs with long 3′ UTRs	Observed in all eukaryotes (including plants).
mRNAs with introns in the 3′ UTR	Observed in yeast and mammals.
Transposons and retroviruses	Observed in yeast and mammals.
Bicistronic mRNAs	Observed only in yeast.
Transcribed pseudogenes	Observed only in yeast.

recognized to be defective in NMD [14–16]. Similarity searches revealed that SMG2 is homologous to yeast UPF1, SMG3 is homologous to UPF2, and SMG4 is homologous to UPF3, respectively. As genome sequencing projects progressed, orthologs of these NMD factors were identified in other eukaryotic organisms based on homology searches [17,18]. All seven factors are present in *Homo sapiens*, and *Drosophila melanogaster* has orthologs for SMG1, UPF1, UPF2, UPF3, SMG5, and SMG6, but appears to lack an ortholog for SMG7 [19–27]. It is likely that additional, yet unknown NMD factors exist in vertebrates. Notably, Longman et al. recently identified in *C. elegans* two additional proteins, called SMGL-1/hNAG and SMGL-2/hDHX34, which are required for NMD in worms and humans [28].

UPF1, UPF2, and UPF3 constitute the conserved core of the NMD system. The nucleic acid-dependent ATPase and RNA helicase UPF1 shows the highest sequence conservation among the UPF proteins in different species [10,17,29,30] and understanding its structure, functions and regulation is key to elucidate the molecular mechanism of NMD. In the conserved region, seven group I helicase motifs can be found, and the ATPase activity of UPF1 resides in two of these helicase motifs and is linked to the 5′ to 3′ helicase activity of the protein [30]. The ATPase activity is essential for NMD in all tested species [31–35] and the RNA-binding activity of UPF1 is modulated by ATP [30,36]. UPF1 localizes predominantly to the cytoplasm [22], but has the capacity to shuttle between the nucleus and the cytoplasm [32]. UPF1 associates with the translation release factors eRF1 and eRF3 and with UPF2 [26,31,37,38] (Fig. 1). UPF1 interacts with UPF2 through its cysteine–histidine-rich domain (amino acids 115–272), which forms three Zinc-binding motifs arranged into two tandem modules, resembling the RING-box and U-box domains of ubiquitin ligases [39]. Multiple serine residues in the N- and C-terminal regions of UPF1 are targets for phosphorylation [33,40]. Indeed, UPF1 activity in humans and worms is regulated by cycles of phosphorylation and dephosphorylation that depend on additional NMD factors. Phosphorylation of UPF1 is catalyzed by SMG1 and requires UPF2 and UPF3 [31,41], whereas dephosphorylation of UPF1 is mediated by SMG5, SMG6, and SMG7, which are thought to recruit protein phosphatase 2A (PP2A) [25,31,33,41–44]. Phosphorylation of human UPF1 seems to lead to a remodeling of the UPF1-containing surveillance complex. Overexpression of a SMG1 mutant deficient in its kinase activity strongly increased UPF1 co-precipitation with eRF3, suggesting that phosphorylation of UPF1 induces the dissociation of eRF3 from UPF1 [31].

UPF2, which is also a phosphoprotein [43,45], interacts with both UPF1 and UPF3, thereby serving as a bridge between the two [22,24,26,38]. The interaction between human UPF2 and UPF3b (see below) involves the last of the three middle of eIF4G-like (MIF4G) domains of UPF2 and the RNA-binding domain (RBD) of UPF3b [46]. Human UPF2 interacts with the N-terminal Zinc-finger domain of UPF1 mainly through its C-terminal region (amino acids 1084–1272) [24,26], but amino acids 94–133 from the N-terminal region also contribute to this interaction to some extent [26]. Even though the N-terminus of UPF2 contains several nuclear localization signals (NLS) and the N-terminal 120 amino acids can target a heterologous protein

Table 2
Homologous factors involved in NMD from different species

	S. cerevisiae	*C. elegans*	*D. melanogaster*	Plants	Mammals
NMD factors	UPF1	SMG2 (UPF1)	UPF1	UPF1	UPF1 (RENT1)
	UPF2 (NMD2)	SMG3 (UPF2)	UPF2	UPF2	UPF2
	UPF3	SMG4 (UPF3)	UPF3	UPF3	UPF3a, UPF3b (UPF3X)
	–	SMG1	SMG1	ND[a]	SMG1
	–	SMG5	SMG5	ND[a]	SMG5
	–	SMG6	SMG6	ND[a]	SMG6
	–	SMG7	–	ND[a]	SMG7
	–	SMGL-1	–	SMGL-1[b]	NAG (SMGL-1)
	–	SMGL-2	SMGL-2[b]	–	DHX34 (SMGL-2)
Translation termination	SUP45	T05H4.6	eRF1	ERF1	eRF1
	SUP35	H19N07.1, K07A12.4b	eRF3	ERF3	eRF3a, eRF3b
	PAB1	pab-1	pAbp	PAB	PABPC1
Exon junction complex (EJC)	No homologs	Present, not involved in NMD	Present, not involved in NMD	Present, role in NMD is still not clear	eIF4A3 Y14 MAGOH BARENTSZ (MLN51)

[a] ND = not determined.
[b] Role in NMD is not determined.