

Nonsense-mediated mRNA decay: from vacuum cleaner to Swiss army knife

Gabriele Neu-Yilik^{*†}, Niels H Gehring^{*†}, Matthias W Hentze^{†‡} and Andreas E Kulozik^{*†}

Addresses: ^{*}Department of Pediatric Oncology, Hematology and Immunology, University of Heidelberg, Im Neuenheimer Feld 150, 69120 Heidelberg, Germany. [†]Molecular Medicine Partnership Unit, EMBL/University of Heidelberg, Im Neuenheimer Feld 156, 69120 Heidelberg, Germany. [‡]EMBL Heidelberg, Gene Expression Programme, Meyerhofstrasse 1, 69117 Heidelberg, Germany.

Correspondence: Gabriele Neu-Yilik. E-mail: gaby.neu-yilik@med.uni-heidelberg.de

Published: 30 March 2004

Genome Biology 2004, **5**:218

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2004/5/4/218>

© 2004 BioMed Central Ltd

Abstract

Nonsense-mediated mRNA decay (NMD) downmodulates mRNAs that have in-frame premature termination codons and prevents translation of potentially harmful truncated proteins from aberrant mRNAs. Two new approaches have identified physiological NMD substrates, and suggest that NMD functions as a multipurpose tool in the modulation of gene expression.

Nonsense-mediated mRNA decay (NMD) is a specific pathway for the degradation of mRNAs that have premature termination codons (PTCs) in their open reading frames (ORFs). Its importance is highlighted by its conservation in all eukaryotes. NMD counteracts the potentially harmful impact of mRNAs that have PTCs as a result of errors at various levels of gene expression, such as nonsense and frameshift mutations, transcriptional errors and faulty splicing. Thus, NMD serves as a 'cellular vacuum cleaner' that protects the cell from the potentially harmful effects of truncated proteins by eliminating mRNAs with PTCs in a sequence of events that is not yet fully understood. In recent years numerous biochemical and cell-biological investigations in *Saccharomyces cerevisiae* [1], *Drosophila melanogaster* [2], *Caenorhabditis elegans* [3] and human [4,5] cells have helped to elucidate some of the mechanistic details underlying the NMD pathway. A role for NMD in the regulation of mRNA metabolism beyond the mere vacuum cleaner function for faulty mRNAs has been suspected, and was foreshadowed by work on the splicing factor SC35 and some ribosomal proteins [6,7]. Now, 'genome-wide' approaches - one in yeast using microarrays and another *in silico*, analyzing information mined from mRNA and protein

databases - have added powerful evidence to suggest that NMD may serve multiple purposes in gene expression [8-11].

Inspirations from yeast

In yeast, NMD depends on the expression of the Upf1, Upf2 (Nmd2) and Upf3 proteins. Single or simultaneous inactivation of the UPF genes stabilizes nonsense-containing mRNAs, indicating that their protein products interact functionally in the same pathway. He *et al.* [8] used high-density oligonucleotide arrays to analyze genome-wide expression profiles of yeast strains containing single deletions of the *UPF1*, *UPF2* or *UPF3* genes, as well as of the *DCP1* and *XRN1* genes which encode proteins with activities thought to be involved in the NMD pathway - an essential component of the mRNA decapping enzyme and the 5'-3' exonuclease, respectively. They also tested double deletions of the *XRN1* gene in combination with each of the UPF genes. Two-dimensional clustering analysis of the expressed genes for the *Δupf1*, *Δupf2* and *Δupf3* strains yielded several interesting results.

The deletion of *UPF1*, *UPF2* or *UPF3* generated nearly identical expression profiles. Thus, all three gene products act on

the same targets, consistent with the function of Upf1, Upf2 and Upf3 in a single, linear pathway in yeast. The abundance of most mRNAs upregulated in Upf-deficient cells was also increased in *Δdcp1* and *Δxrn1* strains, suggesting that these mRNAs are largely degraded by decapping and subsequent 5'-3' exonucleolytic decay. This approach also identified a considerable number of NMD-regulated transcripts (765 out of the 7,839 genes represented on the microarrays) and showed that NMD substrates are generally expressed at below-average levels. In addition, most NMD substrates were found to be upregulated upon NMD inactivation, but some were downregulated, pointing to the existence of higher-order NMD targets (or additional functions of the Upf proteins in alternative gene-regulation pathways). Finally, only one third of the identified transcripts can be classified into structural or functional groups, some of which are surprising and hitherto unrecognized. Representatives of previously described NMD-substrate categories were identified, including mRNAs with nonsense mutations, transcripts resulting from faulty or regulated alternative splicing, mRNAs subject to leaky scanning during translation initiation, and mRNAs with an upstream ORF or with AATGA or ATGAA motifs immediately upstream of their translation initiation codons. More intriguing is the discovery of several new classes of NMD targets, including mRNAs that use translational +1 frameshifting, bicistronic mRNAs and, most interestingly, two classes of noncoding RNAs: pseudogene transcripts and transcripts encoded by transposable elements or their long terminal repeat (LTR) sequences.

A significant fraction of the protein-encoding transcripts upregulated in the strains with mutations in the NMD-pathway genes [8] could be grouped into clusters of proteins that act in similar pathways. Among these are proteins coordinately involved in telomere maintenance, pre-mRNA splicing, peroxisomal function and DNA repair. This suggests the exciting possibility that NMD could orchestrate the expression of functional groups of genes. Several important results of the work by He *et al.* [8] confirm findings obtained by another lab using a similar approach several years ago [12]. Notably, the coordinate upregulation of genes involved in telomere maintenance by NMD inactivation has already sparked interesting follow-up investigations. The yeast NMD pathway has been shown to accelerate the rate of senescence promoted by the loss of the telomerase enzyme or by the erosion of telomeres that results from altering the stoichiometry of telomere-cap components [13-15]. At least one likely primary target of NMD is the mRNA of Stn1p, an essential protein involved in chromosome-end protection. The observation by He *et al.* [8] that 35.9% of all ORFs encoded in the telomere region were upregulated in strains with mutations in components of the NMD pathway further illustrates that NMD controls many genes near telomere ends, although probably indirectly. These genes are usually silenced and may be derepressed when the protection of chromosome ends is disturbed by the loss of NMD. This enlightening

example demonstrates how NMD can affect whole pathways by regulating the expression of one or a few primary target mRNAs, with consequences for groups of downstream secondary effectors. The discovery of additional examples of NMD-mediated control of functional pathways can be expected, promoting NMD to a gene-expression tool with many utilities. The cellular vacuum cleaner has therefore become a Swiss army knife (Figure 1).

In silico veritas?

In a series of three recent publications [9-11] Steve Brenner and colleagues suggested, by data-mining, that one-third of reliably inferred human splice products form a major class of natural targets for NMD. Alternative splicing is thought to occur in 30-60% of human genes, expanding the coding repertoire of the limited number of genes in the genome and modulating tissue-specific and developmental gene functions. Brenner and colleagues [9-11] now suggest that alternative splicing provides a mechanism to generate PTC-containing splice products that are subsequently degraded by NMD and, as a consequence, that cooperation between alternative splicing and NMD pathways offers a major and currently underappreciated way to regulate gene expression.

For their *in silico* analysis, they mapped well-characterized human RefSeq [16] and LocusLink [17] mRNA sequences to genomic sequences, and then performed high-stringency alignments between these 'RefSeq-coding genes' and expressed sequence tags (ESTs). The reliably inferred splice variants were only accepted as likely NMD targets when they conformed to the '50-nucleotide rule': this hallmark of mammalian NMD predicates that stop codons located at least 50 nucleotides upstream of the last exon junction will be interpreted as 'premature' and trigger NMD. This approach leads to an underestimate of potential NMD targets, because some PTCs (for example, in T-cell receptor gene transcripts) will trigger NMD even when they are not followed by a sufficiently distant intron [18]. Moreover, Brenner and colleagues [9-11] also excluded mRNA variants that are indistinguishable from products of faulty splicing. These studies have unearthed several groups of functionally related proteins whose expression appears to be regulated by NMD, including translation factors and ribosomal proteins. This is in remarkable contrast to the yeast data of He *et al.* [8], where proteins with a function in translation were, if anything, underrepresented in the pool of NMD-regulated genes.

NMD is not (yet) on everyone's mind. As a consequence, Brenner and colleagues [11] found several entries for truncated proteins in the Swiss-Prot database. In some of these cases the available experimental evidence confirms that the mRNAs that encode these truncated proteins are *bona fide* NMD substrates. We are left with a consolation and a surprise. The consolation is that traces of NMD can be uncovered even though they had been overlooked before. Having

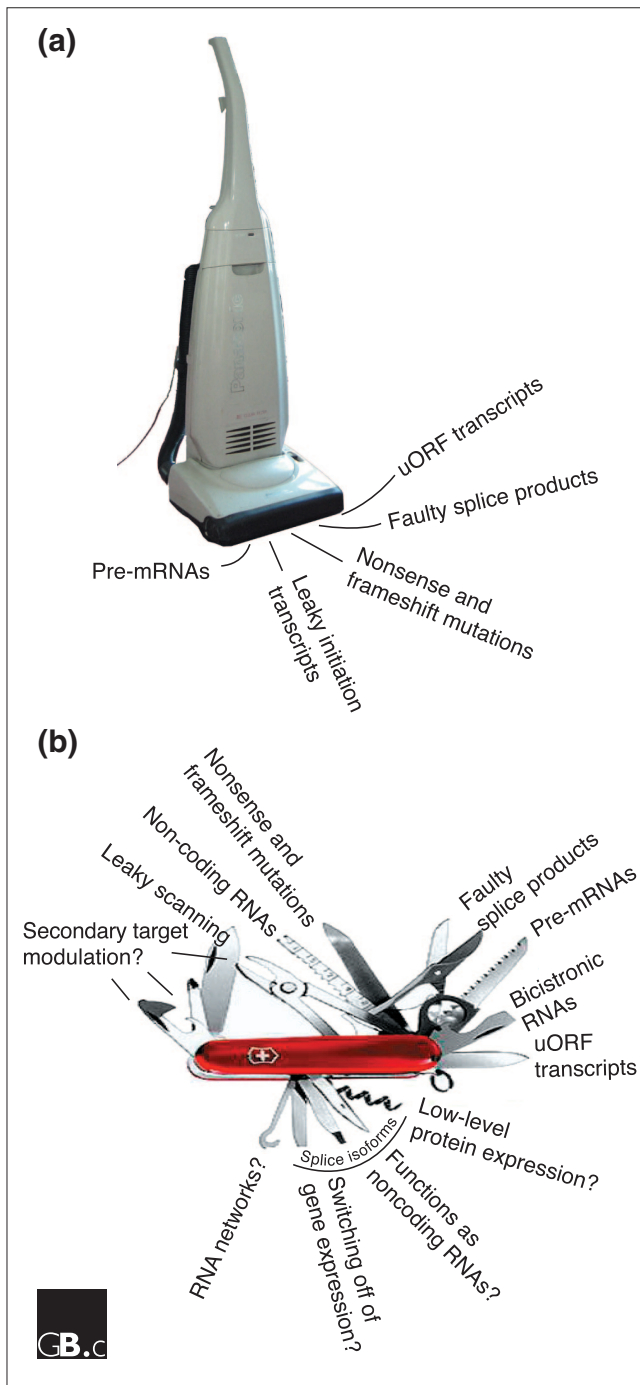


Figure 1
 The role of NMD. (a) Until recently the role of NMD has been predominantly seen as that of a cellular 'vacuum cleaner' that rids the cell of erroneous mRNAs. (b) Now a more sophisticated picture of NMD is emerging - a highly specific and delicate multipurpose tool that contributes at multiple levels to the control and balance of physiological gene expression.

recently had to accept that the number of genes in the human genome is too limited to explain the far higher number of proteins (not to mention other gene products), we

then had to learn that one plausible and elegant explanation lies in alternative splicing, which enables a gene to code for a whole family of related, or sometimes antagonistic, proteins. And now the surprise is that a large portion of this effort is supposedly expended only to direct many of the primary products to decay. Several conclusions can be drawn from these observations. First, databases need to be read and annotated with a full realization of the implications of NMD; and second, NMD seems to serve as a tool for rapidly switching off gene expression. This view extends the idea of NMD as a mechanism for ridding the cell of the potentially harmful products of faulty splicing. But there may be more to it. NMD rarely downregulates the expression of a transcript completely; more commonly, 10-30% of the PTC-containing transcripts survive and may allow the production of physiologically relevant levels of truncated protein products.

For NMD researchers it has always been hard to reconcile these observations with the presumed protective role of NMD, especially as very low levels of biological products can sometimes have enormous effects. Given the problems of detecting the low levels of proteins or peptides produced from downregulated transcripts, in addition to some lingering lack of awareness of NMD, examples to prove otherwise are hard to come by. A recent publication [19] describes a PTC-containing transcript of the high-affinity immunoglobulin E (IgE) receptor, FcεRIβ, arising from retention of an intron. This alternative transcript not only conforms to the 50-nucleotide rule but its expression levels are very low compared to those of the full-length transcript, as would be expected for an NMD target. Nonetheless, the truncated protein is not only detectable, it even competes effectively with the full-length protein to control FcεRIβ expression on the cell surface. Thus, even low endogenous expression levels of NMD targets can suffice to generate a product with a biological function. A similar example of the utility of a *bona fide* NMD substrate is illustrated by the *unc-49* locus in *C. elegans*. This locus uses alternative splicing to produce three GABA-receptor subunits, two of which (A and B) undergo several splice events in their 3' UTRs, rendering them predicted NMD substrates [20]. While A-form transcripts are hardly detectable, B-form transcripts represent the most abundant form and code for a protein essential for the worm's locomotion. Either the B-form transcript escapes NMD, or the residual mRNA left after NMD suffices for the necessary protein production.

Among the examples presented by Brenner and colleagues [11] or in other studies [21] are genes with complex alternative splice patterns resulting in multiple transcript isoforms with or without PTCs. The lower abundance of these isoforms when compared to the full-length form supports the notion that they are targeted to the NMD pathway. But is the possibility that cells engage in complex alternative-splicing procedures generating multiple products, just to finally dispose of them, the only conceivable option? Four out of

five PTC-containing isoforms of the mRNA for the LARD death receptor are readily detectable in non-activated lymphocytes, whereas only the full-length form is expressed in activated lymphocytes [11]. While it is entirely possible that the PTC-containing forms are generated to switch off receptor expression in resting lymphocytes, there are attractive alternatives: as in the case of the Fc ϵ RI β receptor [19], the PTC-containing mRNAs may produce proteins or peptides with relevant, although currently unknown, functions. Thus, quantitative control of the expression of low amounts of protein isoforms could represent yet another facet of the function of the NMD pathway.

A role for NMD in controlling the levels of noncoding RNAs (including noncoding alternative splice products) must also be considered. RNA accounts for more than 95% of the human genome's output and there is increasing evidence that noncoding RNAs (including introns, and spliced and polyadenylated transcripts) can have a function, for example as a modulating network or an additional layer of information [22,23]. Importantly, noncoding RNAs have been discovered among natural NMD targets in yeast [8], where only a relatively small portion of the genome is transcribed into noncoding RNAs. What might be the role of NMD in mammals, which transcribe a far higher percentage of their genome into noncoding RNAs [22,24,25]? Clearly, the recent insights into the RNA-substrate spectrum of the NMD system should enhance the appreciation of NMD as a versatile, multipurpose mechanism that controls the transcriptome qualitatively and quantitatively.

References

- Gonzalez CI, Bhattacharya A, Wang W, Peltz SW: **Nonsense-mediated mRNA decay in *Saccharomyces cerevisiae***. *Gene* 2001, **274**:15-25.
- Gatfield D, Unterholzner L, Ciccarelli FD, Bork P, Izaurralde E: **Nonsense-mediated mRNA decay in *Drosophila*: at the intersection of the yeast and mammalian pathways**. *EMBO J* 2003, **22**:3960-3970.
- Mango SE: **Stop making nonSense: the *C. elegans* smg genes**. *Trends Genet* 2001, **17**:646-653.
- Schell T, Kulozik AE, Hentze MW: **Integration of splicing, transport and translation to achieve mRNA quality control by the nonsense-mediated decay pathway**. *Genome Biol* 2002, **3**:reviews1006.1-1006.5.
- Maquat L: **Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics**. *Nat Rev Mol Cell Biol* 2004, **5**:89-99.
- Sureau A, Gattoni R, Dooghe Y, Stevenin J, Soret J: **SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs**. *EMBO J* 2001, **20**:1785-1796.
- Mitrovich QM, Anderson P: **Unproductively spliced ribosomal protein mRNAs are natural targets of mRNA surveillance in *C. elegans***. *Genes Dev* 2000, **14**:2173-2184.
- He F, Li X, Spatrick P, Casillo R, Dong S, Jacobson A: **Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5' to 3' mRNA decay pathways in yeast**. *Mol Cell* 2003, **12**:1439-1452.
- Lewis BP, Green RE, Brenner SE: **Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans**. *Proc Natl Acad Sci USA* 2003, **100**:189-192.
- Green RE, Lewis BP, Hillman RT, Blanchette M, Lareau LF, Garnett AT, Rio DC, Brenner SE: **Widespread predicted nonsense-mediated mRNA decay of alternatively-spliced transcripts of human normal and disease genes**. *Bioinformatics* 2003, **19 Suppl 1**:i118-i121.
- Hillman RT, Green RE, Brenner SE: **An unappreciated role for RNA surveillance**. *Genome Biol* 2004, **5**:R8.
- Lelivelt MJ, Culbertson MR: **Yeast Upf proteins required for RNA surveillance affect global expression of the yeast transcriptome**. *Mol Cell Biol* 1999, **19**:6710-6719.
- Lew JE, Enomoto S, Berman J: **Telomere length regulation and telomeric chromatin require the nonsense-mediated mRNA decay pathway**. *Mol Cell Biol* 1998, **18**:6121-6130.
- Dahlseid JN, Lew-Smith J, Lelivelt MJ, Enomoto S, Ford A, Desruisseaux M, McClellan M, Lue N, Culbertson MR, Berman J: **mRNAs encoding telomerase components and regulators are controlled by UPF genes in *Saccharomyces cerevisiae***. *Eukaryot Cell* 2003, **2**:134-142.
- Enomoto S, Glowczewski L, Lew-Smith J, Berman JG: **Telomere cap components influence the rate of senescence in telomerase-deficient yeast cells**. *Mol Cell Biol* 2004, **24**:837-845.
- NCBI Reference Sequence (RefSeq)** [<http://www.ncbi.nlm.nih.gov/RefSeq/>]
- LocusLink** [<http://www.ncbi.nlm.nih.gov/LocusLink/>]
- Carter MS, Li S, Wilkinson MF: **A splicing-dependent regulatory mechanism that detects translation signals**. *EMBO J* 1996, **15**:5965-5975.
- Donnadieu E, Jouvin MH, Rana S, Moffatt MF, Mockford EH, Cookson WO, Kinet JP: **Competing functions encoded in the allergy-associated F(c)epsilonR1beta gene**. *Immunity* 2003, **18**:665-674.
- Bamber BA, Beg AA, Twyman RE, Jorgensen EM: **The *Caenorhabditis elegans* unc-49 locus encodes multiple subunits of a heteromultimeric GABA receptor**. *J Neurosci* 1999, **19**:5348-5359.
- Wollerton MC, Gooding C, Wagner EJ, Garcia-Blanco MA, Smith CV: **Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay**. *Mol Cell* 2004, **13**:91-100.
- Mattick JS, Gagen MJ: **The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms**. *Mol Biol Evol* 2001, **18**:1611-1630.
- Mattick JS: **Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms**. *Bioessays* 2003, **25**:930-939.
- Eddy SR: **Noncoding RNA genes**. *Curr Opin Genet Dev* 1999, **9**:695-699.
- Venter JC, Adams MD, Myers EW, Li PV, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, et al.: **The sequence of the human genome**. *Science* 2001, **291**:1304-1351.