

# Stop making nonSense: the *C. elegans smg* genes

Susan E. Mango

**Cells monitor the quality of their mRNAs and degrade any transcripts that are poorly or incompletely translated. In the nematode *Caenorhabditis elegans*, degradation by the mRNA surveillance pathway depends on seven *smg* genes. Three of these genes also have a role in a second mRNA degradation pathway called RNA interference (RNAi), which is triggered by double-stranded RNA (dsRNA). Here I describe what is known about the *smg* genes and their potential functions in these two mRNA degradation pathways.**

When Alexander Pope noted that 'to err is human', he did not have in mind the mistakes that we make at the subcellular level. But since the days of Pope, we have learned that mRNA biosynthesis can produce RNAs with premature stop codons, unspliced introns and other errors. To minimize the potentially harmful effects of these mRNAs a process called mRNA surveillance has evolved to recognize and destroy defective mRNAs in organisms from yeast to humans. In this article I discuss what is known about mRNA surveillance in *Caenorhabditis elegans*, and relate these findings to processes in other animals. I also describe a genetic link between components of the mRNA surveillance machinery and another mRNA degradation pathway called RNA interference. Several recent reviews deal with RNA interference<sup>1</sup> or mRNA surveillance<sup>2-5</sup>.

## mRNA surveillance

### *Nonsense codons and mRNA marking*

Mutations or biosynthetic errors that produce nonsense codons, retained introns, extended untranslated regions (UTRs) or skipped exons often destine an mRNA transcript for degradation by mRNA surveillance. Many of these defects destabilize an mRNA by introducing a premature stop codon. How does mRNA surveillance discriminate between natural termination events and premature ones?

In mammals, one hallmark that distinguishes a natural from a premature stop codon is its location relative to the last exon-exon boundary. The natural site for a stop codon is usually within the terminal exon, whereas destabilizing nonsense codons occur more than 50 nucleotides upstream of the last exon-exon boundary. This distinction suggested that exon boundaries are 'marked' during splicing, by binding factors associated with the splicing machinery, and that the mark remains associated with the mRNA as it exits the nucleus. According to the model, exon boundary marks are remodeled by ribosomes as they traverse the mRNA during the first round of translation. The nature of remodeling is unclear, but

could involve removing some or all of the proteins that comprise the mark. For an aberrant RNA, premature translation termination in an upstream exon leaves downstream marks intact and competent to trigger mRNA degradation (reviewed in Refs 4 and 6).

Most yeast genes lack introns, so presumably exon boundary marking cannot be the means by which natural and premature stop codons are distinguished in yeast cells. Instead, yeast genes carry within their coding domains loosely defined downstream sequence elements (DSEs) that might have a function analogous to that of marks at mammalian exon boundaries<sup>5,6</sup>.

### *mRNA marking in C. elegans*

No rigorous tests have yet determined whether *C. elegans* mRNAs are marked by DSEs, exon boundaries or both. However, most *C. elegans* genes have introns, suggesting that an exon boundary mechanism could exist, and I shall discuss the circumstantial evidence that supports this idea.

The link between the location of a stop codon and its effects on mRNA stability has been studied systematically using the gene *unc-54* (Ref. 7). The *unc-54* locus codes for myosin heavy chain B, and its mRNA normally accumulates to high concentrations. Most nonsense alleles of *unc-54* destabilize transcripts dramatically, with a 20-fold effect on mRNA accumulation. The *unc-54(e1328)* and *unc-54(e1300)* alleles apparently produce less effective targets, with only a fourfold effect on mRNA accumulation for *unc-54(e1328)* and a threefold effect for *unc-54(e1300)*. The location of the different *unc-54* nonsense codons might explain why they produce mRNAs with different stabilities. *unc-54(e1300)* and *unc-54(e1328)* are located in the ultimate and penultimate exons, respectively, whereas the other mutations are located two to three exons further upstream<sup>8,9</sup>. Hence one model to explain the data is that exon boundary marking occurs in *C. elegans*, but that at least two unremodelled marks are needed to destabilize an RNA dramatically.

Could the data also be explained by exonic destabilizing sequences? Yes, they could, if we assume there are many DSEs scattered throughout *unc-54* mRNA or that the DSEs have very long-range effects. Twelve *unc-54* nonsense alleles, occurring over 4 kb of mRNA, lead to equivalent low levels of mRNA accumulation. This observation suggests that RNAs from these 12 alleles are equally potent mRNA

Susan E. Mango  
Huntsman Cancer  
Institute, 2000 Circle of  
Hope, University of Utah,  
Salt Lake City, UT 84112,  
USA.  
e-mail: susan.mango@  
hci.utah.edu

**Table 1. Known or predicted targets of the mRNA surveillance machinery**

Target of NMD?	Gene	Number of examples	Type of lesion codon within exon	Location of stop	Refs
Yes	<i>unc-54</i> <sup>a</sup>	13	Nonsense codons; deletion in 3' UTR	Upstream	7
	<i>rpl</i> (-3, -7a, 10a, -12) <sup>a</sup>	4	Alternative splicing	Upstream	18
	<i>SRP</i> <sup>a</sup>	2	Alternative splicing	Upstream	17
	<i>pha-4</i> <sup>b</sup>	2	Nonsense <i>q500</i> , <i>zu225</i>	Upstream	<sup>d</sup>
	<i>unc-76</i> <sup>b</sup>	1	11-bp deletion <i>e911</i>	Upstream	54
	<i>lin-29</i> <sup>b</sup>	1	Nonsense <i>n546</i>	Upstream	55
Partial	<i>emb-30</i> <sup>b</sup>	1	Nonsense <i>tn477</i>	Ultimate	56
	<i>unc-54</i> <sup>a</sup>	2	Nonsense <i>e1328</i> , <i>e1300</i>	Penultimate, ultimate	7
	<i>glp-1</i> <sup>b</sup>	1	Nonsense <i>q35</i>	Penultimate	10
	<i>emb-30</i> <sup>b</sup>	1	Nonsense <i>tn474</i>	Last exon	56
	<i>dpy-5</i> <sup>b</sup>	1	Nonsense <i>e61</i>	Last exon	11, <sup>e</sup>
	<i>fem-1</i> <sup>b</sup>	1	Nonsense <i>e1991</i>	Penultimate	<sup>f</sup>
No	<i>uaf-1</i> <sup>a,c</sup>	1	Alternative splicing	Penultimate	15
	<i>CeCAT</i> <sup>a</sup>	1	Alternative splicing	Penultimate	17
	<i>unc-73</i> <sup>a,b</sup>	1	Splicing mutation <i>e936</i>	Upstream	14
	<i>unc-49</i> <sup>b</sup>	1	Alternative splicing	Upstream	16
	<i>avr-14</i> <sup>b</sup>	1	Alternative splicing	Upstream	57

<sup>a</sup>mRNA examined in *smg*(+) and *smg*(-) background.  
<sup>b</sup>The likelihood of being a target is predicted based on phenotype. RNA analysis was performed in either a *smg*(+) or a *smg*(-) background (but not both).  
<sup>c</sup>The mRNA is retained in the nucleus<sup>15</sup>.  
<sup>d</sup>M. Horner and S.E. Mango, unpublished.  
<sup>e</sup>C. Thacker and A.M. Rose, pers. commun.  
<sup>f</sup>I.D. Chin-Sang and A.M. Spence, pers. commun.

surveillance targets. *unc-54(e1300)* and *unc-54(e1328)* mRNAs are apparently weaker targets, which would imply that an additional weak element exists downstream of each of these stop codons.

The correlation between a stop codon's position and its effects on mRNA stability has not yet been examined for other genes. Instead, I determined the location of 33 premature stop codons relative to exon-exon boundaries (Table 1). Some of these are nonsense or frameshift mutations identified by genetic screens. Others result from alternative splicing in which the new exon introduces an upstream stop codon. Although the stability of the mutant mRNAs has not always been assessed, the strength of the mutant phenotype can reveal whether an mRNA is a likely target of mRNA surveillance. For example, an allele that is weak, not null, suggests that at least some mRNA accumulates. Clearly, molecular analyses will be required to establish whether these predictions hold true.

The survey suggests that in *C. elegans* a modified exon boundary marking system might exist in which two or more unremodeled marks signal an mRNA for destruction. Twenty-four of the 33 examples listed in Table 1 produce mRNA surveillance targets, and for 23 of these targets, a premature stop codon has been induced by a nonsense mutation or an alternative splice site. The *unc-54(r293)* allele represents an unusual case in which the 3' UTR and polyadenylation site have been deleted, leading to a strong *Unc-54*-mutant phenotype. The resulting mRNA is polyadenylated using the poly(A) signal of the next gene downstream and includes both protein-coding domains. Hence, although the normal *unc-54* stop codon is used, it is now placed upstream of a longer

3' UTR that is punctuated by additional introns (S. O'Connor and P. Anderson, pers. commun.).

Six of 33 mutant RNAs listed in Table 1 appear to be intermediate targets of mRNA surveillance. For example, *glp-1(q35)* has a dominant phenotype, suggesting that *glp-1(q35)* mRNA normally accumulates in spite of mRNA surveillance, and the *glp-1(q35)* phenotype is only mildly enhanced by mutations that disrupt mRNA surveillance<sup>10</sup>. Similar observations have been made for *dpy-5(e61)* (Ref. 11). Each of the six intermediate targets bears a stop codon in the last or penultimate exon. Finally, five of the 33 alleles do not produce mRNA surveillance targets. Two of these five have stop codons in the penultimate exon, consistent with the modified marking hypothesis. The three other genes are discussed below.

Together, the survey results reveal that 23 of 24 robust mRNA surveillance targets carry stop codons in upstream exons, whereas eight of 11 mRNAs that are not efficiently targeted are located within the last two exons. These findings are consistent with a model in which exon boundaries are marked during mRNA splicing. For wild-type transcripts, the marks are remodeled or removed during translation. For mutant RNAs, two or more marks located downstream of the premature stop codon trigger mRNA degradation. The testing of these ideas, by altering the intron-exon structure of the target genes, is an exciting prospect.

At first glance, four examples in the survey do not appear to fit the exon boundary model. For example, the model predicts that a nonsense codon introduced into a gene that lacks introns will not destabilize the expressed RNA. However, *dpy-5(e61)* carries a nonsense mutation that is recognized by the mRNA surveillance machinery, at least to some degree, even

**Table 2. Configurations that are targeted by mRNA surveillance but that lack a premature stop codon<sup>a</sup>**

Allele or construct	Mutation or alteration	Expected effect	Refs
<i>gfpx3</i>	216-bp insertion in 3' UTR	Extended 3' UTR	15
+3A	3' splice site mutation for exon 2 of 3	Lose splicing for 1 or both exons	21
-6A-4A	3' splice site mutation for exon 2 of 3	Lose splicing for 1 or both exons	21
-3A	3' splice site mutation for exon 2 of 3	Lose splicing for 1 or both exons	21
GFP	Added 3' UTR sequences	Extended 3' UTR	<sup>b</sup>
<i>mua-6(rh85)</i>	Mutation of AUG initiation codon	Predict initiation of translation at downstream AUG; extended 5' UTR	<sup>c</sup>
<i>unc-17(md1447)</i>	463-bp deletion	Lose natural 3' UTR and poly(A) site	<sup>d</sup>

<sup>a</sup>Note, the loss of the natural ATG in *mua-6* might trigger mRNA surveillance because, in theory, ribosomes could initiate at a downstream, out of frame AUG.

<sup>b</sup>A. Fire, pers. commun.

<sup>c</sup>J. Plenefisch and M. Hresko, pers. commun.

<sup>d</sup>J. Rand, pers. commun.

though *dpy-5* has only one exon<sup>11</sup> (C. Thacker and A. Rose, pers. commun.). A probable explanation is that *dpy-5* is *trans*-spliced; that is, a second RNA provides an upstream exon that is spliced *in trans* onto *dpy-5* RNA, to produce a two-exon mRNA. The *trans*-splicing reaction might recruit marking factors similar to splicing that occurs in *cis*. A second example is *unc-73(e936)*, which contains a splicing defect that is predicted to produce a stop codon in exon 16 of 20 (or 16 of 32, depending on alternative splicing<sup>12</sup>), and yet RNAs from this mutant allele are stable<sup>14</sup>. One possible explanation is that the splicing mutation affects the subcellular localization of the mutant mRNA, thereby blocking degradation. By analogy, alternative splicing of *uaf-1* produces a nonsense-containing mRNA that is retained in the nucleus and refractory to mRNA surveillance<sup>15</sup>. Alternatively, a defective or incomplete splicing reaction might interfere with the marking process itself. Third, some genes behave as mRNA surveillance targets by an indirect mechanism. For example, all alleles of *tra-3* are suppressed by mutations in the mRNA surveillance pathway, including null mutations<sup>11</sup>. This effect probably reflects the existence of a natural target of mRNA surveillance that interacts genetically with *tra-3*. Fourth, some genes produce proteins from mRNAs that normally contain upstream stop codons, and it is unclear the degree to which those mRNAs are targeted by mRNA surveillance. For example, the *unc-49* locus uses alternative splicing to produce two GABA receptor subunits, B and C, which are needed for worm locomotion<sup>16</sup>. It is surprising that the most abundant mRNA coding for the B form carries a stop codon in an upstream exon, suggesting that this mRNA is not an effective target of mRNA surveillance.

If exon boundary marking occurs in *C. elegans*, why two marks instead of one? Two marks imply that the factors associated with the exon-exon boundary are less effective at promoting mRNA instability in a worm than they are in a mammal. This could be true if the mRNA surveillance machinery were inherently less active and the destabilizing effects of unremodelled marks were cumulative. Alternatively, the complex could be sensitive to the distance from the mark to the transcript's stop codon or poly(A) tail.

Because *C. elegans* exons are small (generally fewer than 300 bases) and introns are smaller (~60 bases<sup>13</sup>), at least two exon-exon boundaries might be required to provide enough distance or time for the marking factors to function efficiently. This model suggests that competition could exist between the destabilizing activity of the marking complex and stabilizing events in which the 3' end of the mRNA participates.

#### Natural targets of mRNA surveillance

mRNA surveillance is not used only to eliminate biosynthetic errors, it is also a means of regulating normal gene expression. Two reports have identified natural targets of mRNA surveillance in *C. elegans*. Genes for splicing factors and ribosomal proteins each possess alternatively spliced exons that incorporate an early stop codon in the retained intron<sup>17,18</sup>. These naturally occurring but 'premature' stop codons activate the mRNA surveillance machinery. For the ribosomal genes, the alternative exon is also found in other nematode species, revealing that this pathway is probably a conserved negative-feedback loop. When one ribosomal protein is produced in excess, the free protein might bind its own pre-mRNA, thereby favoring the alternative splice and reducing its own expression<sup>18,19</sup>. The mRNA surveillance machinery ensures that these unproductive RNAs do not overwhelm the cell.

#### Targets of mRNA surveillance that lack premature stop codons

Both the DSE and exon boundary marking models predict that mRNAs without premature stop codons should not be targeted by mRNA surveillance. However, we know that although mRNAs with premature stop codons are the most prevalent class of target, other errors can cause a transcript to be degraded. In *C. elegans*, these errors include extended 5' or 3' UTRs, or unspliced introns (Refs 11,20,21; J. Plenefisch and M. Hresko, pers. commun.; Table 2). In the same way as RNAs with nonsense codons, these mRNAs are degraded, and degradation depends on the mRNA surveillance machinery. One appealing hypothesis is that these mRNAs might be destabilized because they lack the proper configuration for efficient translation<sup>6</sup>. This means that alterations that influence

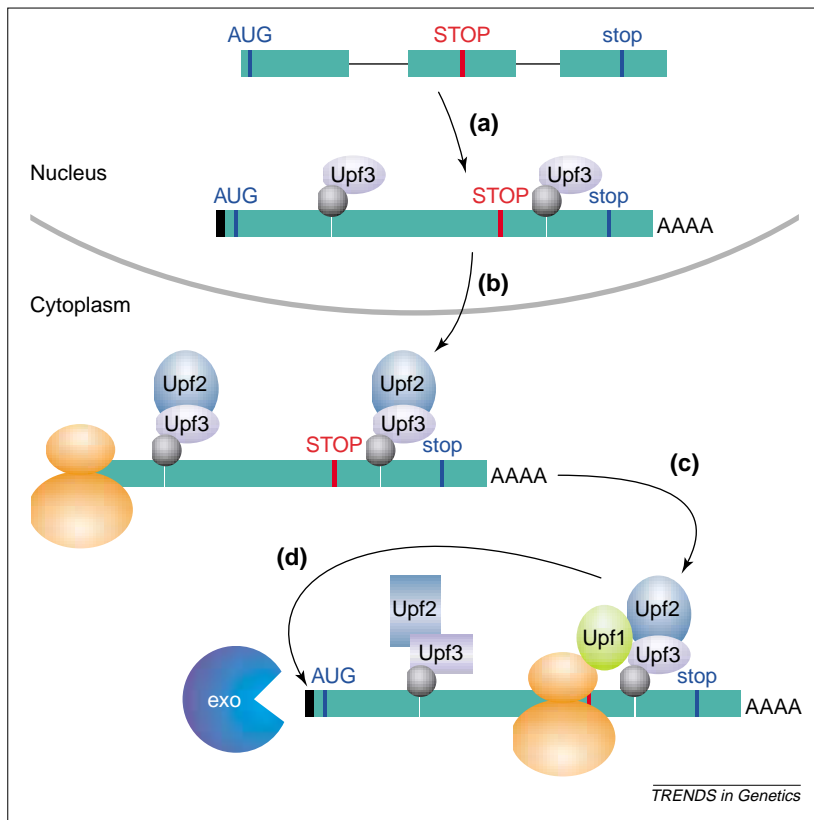


Fig. 1. A model for mRNA surveillance. During mRNA processing (a), Upf3/SMG4 is recruited to exon-exon boundaries by RNPS1 and Y14. The mRNA is exported from the nucleus and associates with Upf2/SMG-3 (b). Translating ribosomes remodel or remove the Upf complex (squares) (c) If the ribosome encounters a premature stop codon (red 'STOP') and terminates, Upf1/SMG-2 is recruited by translation termination factors and forms a functional mRNA surveillance complex at downstream, unremodelled marks (c). In yeast, this complex promotes mRNA decapping and exonucleolytic digestion in the absence of deadenylation (d).

the efficiency of an RNA's translation by any mechanism, not only premature translation termination, could be sufficient to trigger mRNA surveillance in a worm.

#### mRNA marking in mammals and yeast

Genetic screens have identified genes required for mRNA surveillance, three of which are evolutionarily conserved. The product of one of these, yeast/vertebrate Upf3, could have a marking function. Upf3 is a nucleocytoplasmic shuttling protein that associates preferentially with spliced mRNAs in vertebrate cells<sup>22,23</sup>. Upf3 does not bind RNA directly, but is recruited to spliced RNAs most probably by RNPS1 and Y14, which have each been shown to co-immunoprecipitate with hUpf3 (Refs 58,59). These two proteins are part of a multiprotein complex that is deposited 20–24 nucleotides upstream of exon boundaries after splicing, and hUpf3 probably joins this complex just before mRNA export. As the mRNA and its associated proteins leave the nucleus, additional components of the mRNA surveillance machinery are recruited (Fig. 1; reviewed in Ref. 4). Vertebrate Upf2 localizes to the periphery of the nucleus and is known to bind Upf3 and a third mRNA surveillance factor called Upf1. Hence, a possible model is that Upf2 joins the complex soon after export and acts as a bridge to bring Upf1 to the mRNA surveillance machinery. Upf1 has RNA binding, RNA-dependent ATPase and RNA helicase capabilities *in vitro*. In addition, yeast Upf1 contacts the translation termination factors eRF1 and eRF3, and influences translation termination<sup>4–6</sup>. These

characteristics, coupled with the observation that Upf1 is associated with polysomes, have led to the hypothesis that Upf1 is recruited by ribosomes to RNA, where it promotes translation termination and mRNA surveillance<sup>5,6</sup>.

In yeast, the mRNA surveillance machinery induces mRNA decapping and degradation. Unlike degradation of RNAs that have natural stop codons, mRNA surveillance does not require the target mRNA to become deadenylated<sup>5,6</sup>. Rather, mRNA decapping might be linked to the surveillance machinery directly because: first, Upf1 associates with Dcp2, a protein that binds the mRNA decapping enzyme Dcp1 and promotes decapping; second, Upf2 contains domains homologous to eIF4G, a translation initiation factor that binds mRNA cap-binding proteins<sup>60</sup>. Recognition of the cap and recruitment of decapping enzymes by the mRNA surveillance machinery are predicted to affect the target mRNA in three ways<sup>4–6</sup>. First, the target mRNA would become susceptible to digestion by 5' to 3' exonucleases such as yeast Xrn1. Second, decapping would impair mRNA recognition and translation initiation by the 40S ribosomal subunit and its associated factors. Third, decapping would block or disrupt the interaction between poly(A) binding protein located at the 3' end with the translation initiation factor eIF4G at the 5' end. A circular RNP configuration, with the 5' and 3' ends closely apposed, is thought to promote efficient translation.

#### mRNA surveillance in *C. elegans*

Do the models devised for mRNA surveillance in yeast and vertebrates hold true for worms? The answer is probably yes, given the sequence conservation between components of the mRNA surveillance machinery. Seven *smg* (for 'suppressor with morphogenetic defects on genitalia') genes have been identified in *C. elegans*, each of which is required for mRNA surveillance. Mutations in these loci were originally isolated as allele-specific suppressors for a wide variety of genes<sup>11</sup>. As *smg*-suppressible alleles included *unc-54(r293)*, which has a normal *unc-54* coding domain, it seemed likely that *smg* mutants were informational suppressors that affected a process other than translation. More recent data have revealed that three of the *smg* genes, *smg-2*, *smg-3* and *smg-4*, appear to encode the *C. elegans* counterparts of Upf1, Upf2 and Upf3 respectively, on the basis of sequence conservation and, for SMG-4, its likely subcellular localization<sup>27–29</sup>.

Biochemical studies in *C. elegans* have centered upon the interaction between SMG-2 (Upf1) and other SMG factors. These studies have shown that *smg-3* (Upf2) and *smg-4* (Upf3) are required for SMG-2 to become phosphorylated<sup>27</sup>. This effect is presumably indirect, as these genes do not encode kinases. The *smg-1* locus, which codes for a phosphoinositide-3-related kinase homolog, is also required for SMG-2 phosphorylation and might function directly<sup>27</sup>. Conversely, *smg-5*, *smg-6* and *smg-7* are required for

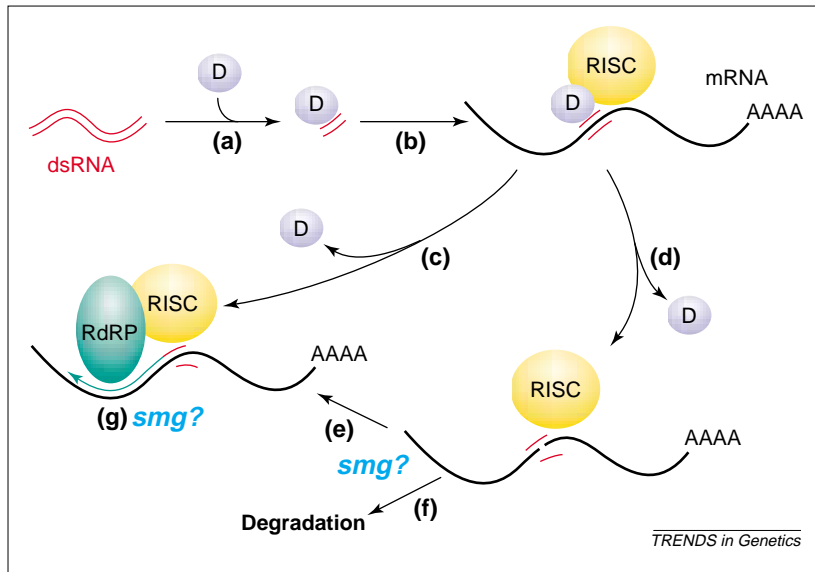


Fig. 2. A model for RNA interference (RNAi). (a) dsRNA (red) is processed into 21–23 nucleotide pieces in an ATP-dependent process that requires dicer/dcr-1 (D). The processed siRNAs, complexed with RISC (RNA-induced silencing complex), bind to homologous endogenous mRNA (b), which is cleaved endonucleolytically (d). The cleaved mRNA might either serve as a template for RdRP or be completely degraded (f). Alternatively, RdRP might use the uncleaved target as a template (c). I propose that *C. elegans smg* genes either influence the choice between amplification (e) and mRNA degradation (f) or directly promote RNA biosynthesis by RdRP (g).

SMG-2 to become dephosphorylated<sup>27</sup>. Homologs of *smg-5* and *smg-7* have not been identified in yeast, suggesting that the phosphorylation cycle of SMG-2/Upf1 might not exist in this organism (cloning of *smg-6* has not yet been reported). By contrast, human Upf1 is phosphorylated by a Wortmannin-sensitive kinase, which is likely to be the *smg-1* ortholog hSMG-1 (Refs 30,31).

What is the role of SMG-2 phosphorylation? One speculative suggestion is that phosphorylation is linked to cycles of SMG-2 function. For example, as most SMG-2 is non-phosphorylated, this form could reflect SMG-2 free in the cytoplasm. Once SMG-2 is associated with ribosomes or assembled into a functional mRNA surveillance complex, it might become phosphorylated to regulate its activity, localization, stability or affinity for RNA or protein. If this is the case, one could imagine a scenario where *smg-1*, *smg-3* and *smg-4* are required to assemble the mRNA surveillance machinery onto mRNA (and induce SMG-2 phosphorylation), whereas *smg-5*, *smg-6* and *smg-7* are required for the mRNA surveillance machinery to perform its function and restore SMG-2 to its resting state. This idea fits nicely with the observation that mutants carrying *smg-2* alleles predicted to inactivate its nucleotide-binding site also accumulate phosphorylated SMG-2 (Ref. 27). These mutant proteins might be able to associate with target mRNA and associated mRNA surveillance factors, where they remain trapped because they are inactive.

How do *smg-5*, *smg-6* and *smg-7* function? Two models have been proposed. First, SMG-5 might encode a regulatory subunit of protein phosphatase 2A (PP2A), an abundant Ser/Thr protein phosphatase. Anderson and colleagues have identified SMG-5 bound to SMG-2, SMG-7 and two subunits of PP2A *in vivo* (K. Anders, A. Grimson and P. Anderson, pers. commun.). Normally, PP2A is composed of three subunits: a catalytic C subunit, a regulatory B subunit and a scaffolding A subunit<sup>32</sup>. SMG-5 binds to PP2A<sub>A</sub>

and PP2A<sub>C</sub>, leading Anderson and colleagues to propose SMG-5 is a novel B subunit that targets SMG-2 for PP2A-mediated dephosphorylation.

A second model for *smg-5* and *smg-7* is that these proteins encode nucleases. This model derives from sensitive homology searches suggesting that *smg-5* and *smg-7* belong to the PIN family of proteins (PiIT-amino-terminal), which includes PiIT and yeast Dis3p 3'–5' exonucleases<sup>33</sup>. If this model is correct, then SMG-5 might associate with PP2A as a substrate rather than a regulatory subunit.

#### A second mRNA degradation pathway called RNA interference depends on a subset of *smg* genes

Recently, the mRNA surveillance pathway has been linked to a second mRNA degradation pathway called RNA interference (RNAi). When a cell is exposed to double-stranded RNA (dsRNA), mRNA from the endogenous, homologous gene is selectively degraded<sup>1</sup>. Biochemical and genetic studies have begun to address the mechanism underlying this degradation pathway, and have suggested that one stage of RNAi is amplification or propagation of an RNAi signal. The phenotypes associated with mutations in the mRNA surveillance pathway suggest that the SMG factors might influence RNAi amplification.

#### A model for RNAi

A two-step mechanism has been proposed for dsRNA-mediated degradation, on the basis of genetic and biochemical studies<sup>1</sup> (Fig. 2). In the first stage, dsRNA is bound by cellular factors and processed into 21–23 nucleotide pieces. One of these factors is likely to be Dicer, an RNase III-related protein with helicase and dsRNA-binding motifs<sup>34,35,61</sup>. In the second stage, the cleaved dsRNA functions as a guide RNA or 'siRNA' to recruit a degradation complex called RISC (RNA-induced silencing complex) to the target mRNA, which is then cleaved endonucleolytically. One component of RISC is Argonaute2, which is homologous to genes identified in other organisms as critical for RNAi<sup>62</sup>. Additional genes required for RNAi have been identified genetically, but their biochemical roles are unknown<sup>1</sup>.

#### Amplification of the RNAi signal

Although we have an outline of how mRNAs might be targeted for destruction during RNAi, it is clear that additional processes are at work *in vivo*. One of these is amplification or propagation of the RNAi signal. In worms and mammals, small amounts of dsRNA are effective for silencing<sup>36,37</sup>. For example, Fire and colleagues calculated that two dsRNA molecules per cell were sufficient to inactivate abundant *unc-22* RNAs for the life of a worm. This observation raised the conundrum of how so little dsRNA could inactivate a highly expressed target. One possible explanation is that the dsRNA functions catalytically to target many mRNA molecules. According to this model, one role of the helicase domain of *Dicer* might be to unwind the siRNA from its target mRNA,

thereby releasing the cleaved mRNA and readying the degradation complex for another substrate.

A second explanation is that dsRNA induces transcriptional silencing and, in some organisms, DNA methylation<sup>1,38</sup>. DNA methylation has not been observed in worms, which could explain why dsRNA homologous to promoter sequences fails to elicit silencing in worms, in contrast to other organisms<sup>36</sup>. dsRNA directed to promoter sequences leads to promoter methylation and transcriptional silencing in plants, whereas dsRNA directed to exonic sequences results in post-transcriptional silencing<sup>38</sup>.

A third possible explanation for the potency of RNAi is that the RNAi signal is amplified. Several lines of evidence suggest that amplification occurs *in vivo*, although it has not yet been observed *in vitro*. Perhaps the most compelling is that RNA-dependent RNA polymerases (RdRPs) are important mediators of RNAi in worms<sup>39</sup>, *Neurospora*<sup>40</sup> and *Arabidopsis*<sup>39,41,42</sup>. Surprisingly, RdRP homologues have not been observed in *Drosophila*, raising the question of whether other genes exist to fulfill this function. In plants, the amount of siRNA decreases when RdRP has been inactivated, consistent with an amplification function for RdRP (Ref. 43). RdRPs have been studied most extensively in tomato, where *in vitro* they can synthesize an RNA complement from a single-stranded RNA or DNA template<sup>44</sup>. Double-stranded templates can also be used, provided a helicase is present to separate the template strands. Although the *in vivo* template for RdRPs is unknown, it is unlikely to be the exogenous dsRNA itself. Chemically modified dsRNA molecules have revealed a strand bias for RNAi in which modification of the sense strand is tolerated better than the antisense strand<sup>63</sup>. This bias implies that incoming dsRNA is not replicated to any significant extent, because replication would dilute the modified dsRNA. Instead, an attractive candidate for the RdRP template is the target mRNA bound to the siRNA, which could function as a primer. This model predicts that highly expressed mRNAs might be better targets for RNAi than rare mRNAs, as the amplification pathway could reinforce the primary degradation pathway. This prediction has been borne out in genome-wide RNAi screens in *C. elegans*<sup>45</sup>. At present, nothing is known about amplification, beyond the presumed involvement of RdRPs.

#### *A connection between mRNA surveillance and RNAi*

Three *smg* genes are required for persistence of RNAi in *C. elegans*<sup>46</sup>. Animals carrying a mutation in *smg-2*, *smg-5* or *smg-6* initially show silencing similar to that seen in their wild-type counterparts, but rapidly recover from RNAi. The level of target mRNA parallels the phenotypic recovery, whereas stability and modification of dsRNA appears unaffected by *smg* mutations. It is important to note that the recovery from RNAi in *smg* mutants is not the result of a general inactivation of the mRNA surveillance pathway, because only a subset of *smg* genes is involved. In *Arabidopsis*, *SDE3* encodes a

homolog of SMG-2, and is also required for persistence of post-transcriptional silencing. In one assay, juvenile tissue (hypocotyl and cotyledons) was silenced in *SDE3* mutants, whereas more mature material (leaves) was not<sup>43</sup>. In a second assay, silencing was observed only transiently in *SDE3* mutants compared with the wild type<sup>47</sup>. Although it is not yet clear whether *SDE3* has a role in mRNA surveillance, the parallels between SMG-2 and *SDE3* for RNAi are striking.

Recent work with the green alga *Chlamydomonas reinhardtii* has also uncovered a link between post-transcriptional gene silencing and mRNA surveillance. Cerutti and colleagues identified MUT6, a DEAH-box (Asp-Glu-Ala-His box) RNA helicase homolog that is required for both processes<sup>48</sup>. Algae lacking *mut6* activity reactivate a transgene known to be post-transcriptionally silenced in the wild type and accumulate aberrant RNAs. MUT6 belongs to a different class of RNA helicase from SMG-2 and *SDE3*; the latter are superfamily (SF) II members, while MUT6 belongs to the SFI family<sup>49</sup>. SFI family members of the DEAH-box class typically play a part in pre-mRNA splicing. By analogy, MUT6 might have a role in splicing or mRNA marking or both in the nucleus, whereas SMG-2 is predicted to be a cytoplasmic protein, similar to its homologues in other organisms<sup>23,31,50</sup>.

How do *smg* genes affect RNAi? One possibility is that these factors are required for the process of amplification itself. This idea is supported by the phenotypic similarities between *sde1/sgs2* (RdRP) and *sde3* mutants in plants (e.g. virus susceptibility, transient silencing)<sup>43,47</sup>. Alternatively, the *smg* genes could influence whether an mRNA becomes a template for amplification. For example, a target mRNA that has been bound by siRNA can either be degraded or function as a template for RdRP, and SMG-2, SMG-5 and SMG-6 could bias the outcome. One speculative hypothesis is that these three factors function in an RNAi complex that is distinct from the mRNA surveillance complex. This idea could explain why all SMG proteins are required for mRNA surveillance, but only a subset affect RNAi.

#### Comparing RNAi and mRNA surveillance

A comparison of RNAi and mRNA surveillance reveals that each of these processes plays a part in recognizing aberrant RNAs and each relies on a degradation mechanism that is independent of poly(A) shortening. Nevertheless, these two pathways appear to function by very different mechanisms. For example, RNAi *in vivo* and *in vitro* is apparently uncoupled to translation (see, for example, Refs 51,52) whereas mRNA surveillance depends critically on the translation machinery to identify its targets<sup>5,6</sup>. RNAi inactivates transcripts by an endonucleolytic clip whereas mRNA surveillance leads to mRNA decapping followed by 5' to 3' degradation<sup>5,6,53</sup>. These differences raise the question of what mechanisms link the two processes. Although we do not yet have an answer, it is

#### Acknowledgments

I thank P. Anderson, D. Baulcombe, B. Corneliussen, M. Hresko, J. Plenefisch, J. Rand, A. Rose, A. Spence and C. Thacker for communicating data in advance of publication, and D. Ayer, B. Bass, L. Kaltenbach and especially R. Parker for comments on the manuscript. SEM is an assistant investigator at the Huntsman Cancer Institute, and is funded by the National Institutes of Health and the American Cancer Society.

important to note that the mechanism proposed for mRNA surveillance does not account for *smg-5* and *smg-6*, the genes associated with pronounced RNAi phenotypes. Conversely, the *in vitro* studies of RNAi have not yet tackled amplification. Intriguingly, the *in vitro* studies have revealed that RISC is bound to ribosomes in cell-free extracts, suggesting a possible link to translation<sup>64</sup>. Hence, there could be similarities between these two processes yet to be uncovered.

In summary, mRNA surveillance and RNAi have emerged as two fascinating mRNA degradation pathways. Genetic screens have identified factors required for each of these processes, and we are beginning to understand their biochemical roles. The *in vivo* and *in vitro* assays that have been developed over the past few years have provided powerful tools that we can apply to constructing molecular pathways for mRNA surveillance and RNAi.

## References

- Sharp, P.A. (2001) RNA interference – 2001. *Genes Dev.* 15, 485–490
- Nagy, E. and Maquat, L.E. (1998) A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem. Sci.* 23, 198–199
- Hilleren, P. and Parker, R. (1999) mRNA surveillance in eukaryotes: kinetic proofreading of proper translation termination as assessed by mRNP domain organization? *RNA* 5, 711–719
- Mitchell, P. and Tollervey, D. (2001) mRNA turnover. *Curr. Opin. Cell Biol.* 13, 320–325
- Wilusz, C.J. *et al.* (2001) The cap-to-tail guide to mRNA turnover. *Nat. Rev. Mol. Cell Biol.* 2, 237–246
- Hilleren, P. and Parker, R. (1999) Mechanisms of mRNA surveillance in eukaryotes. *Ann. Rev. Genet.* 33, 229–260
- Pulak, R. and Anderson, R.P. (1993) mRNA surveillance by the *Caenorhabditis elegans smg* genes. *Genes Dev.* 7, 1885–1897
- Dibb, N.J. *et al.* (1985) Sequence analysis of mutations that affect the synthesis, assembly and enzymatic activity of the unc-54 myosin heavy chain of *Caenorhabditis elegans*. *J. Mol. Biol.* 183, 543–551
- Bejsovec, A. and Anderson, P. (1990) Functions of the myosin ATP and actin binding sites are required for *C. elegans* thick filament assembly. *Cell* 60, 133–140
- Mango, S.E. *et al.* (1991) Carboxy-terminal truncation activates *glp-1* protein to specify vulval fates in *Caenorhabditis elegans*. *Nature* 352, 811–815
- Hodgkin, J. *et al.* (1989) A new kind of informational suppression in the nematode *Caenorhabditis elegans*. *Genetics* 123, 301–313
- Steven, R. *et al.* (1998) UNC-73 activates the Rac GTPase and is required for cell and growth cone migrations in *C. elegans*. *Cell* 92, 785–795
- Roller, A.B. *et al.* (2000) The allele-specific suppressor *sup-39* alters use of cryptic splice sites in *Caenorhabditis elegans*. *Genetics* 154, 1169–1179
- Blumenthal, T. and Steward, K. (1997) RNA processing and gene structure. In *C. elegans II* (Riddle, D.L., Blumenthal, T. *et al.*, eds), pp. 117–145, Cold Spring Harbor Laboratory Press
- MacMorris, M.A. *et al.* (1999) An exon that prevents transport of a mature mRNA. *Proc. Natl. Acad. Sci. U. S. A.* 96, 3813–3818
- Bamber, B.A. *et al.* (1999) The *Caenorhabditis elegans unc-49* locus encodes multiple subunits of a heteromultimeric GABA receptor. *J. Neurosci.* 19, 5348–5359
- Morrison, M. *et al.* (1997) *smg* mutants affect the expression of alternatively spliced SR protein mRNAs in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U. S. A.* 94, 9782–9785
- Mitrovich, Q.M. and Anderson, P. (2000) Unproductively spliced ribosomal protein mRNAs are natural targets of mRNA surveillance in *C. elegans*. *Genes Dev.* 14, 2173–2184
- Warner, J.R. (1999) The economics of ribosome biosynthesis in yeast. *Trends Biochem. Sci.* 24, 437–440
- Pulak, R.A. and Anderson, P. (1988) Structures of spontaneous deletions in *Caenorhabditis elegans*. *Mol. Cell. Biol.* 8, 3748–3754
- Zhang, H. and Blumenthal, T. (1996) Functional analysis of an intron 3' splice site in *Caenorhabditis elegans*. *RNA* 2, 380–388
- Shirley, R.L. *et al.* (1998) A factor required for nonsense-mediated mRNA decay in yeast is exported from the nucleus to the cytoplasm by a nuclear export signal sequence. *J. Cell. Sci.* 111, 3129–3143
- Lykke-Andersen, J. *et al.* (2000) Human *Upf* proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon. *Cell* 103, 1121–1131
- Mendell, J.T. *et al.* (2000) Novel *upf2* orthologues suggest a functional link between translation initiation and nonsense surveillance complexes. *Mol. Cell. Biol.* 20, 8944–8957
- Ponting, C.P. (2000) Novel eIF4G domain homologues linking mRNA translation with nonsense-mediated mRNA decay. *Trends Biochem. Sci.* 25, 423–426
- Cui, Y. *et al.* (1999) Mutations in the MOF2/SUI1 gene affect both translation and nonsense-mediated mRNA decay. *RNA* 5, 794–804
- Page, M.F. *et al.* (1999) SMG-2 is a phosphorylated protein required for mRNA surveillance in *C. elegans* and related to Upf1p of yeast. *Mol. Cell. Biol.* 19, 5943–5951
- Aronoff, R. *et al.* (2001) Molecular identification of *smg-4*, required for mRNA surveillance in *C. elegans*. *Gene* 268, 153–164
- Serin, G. *et al.* (2001) Identification and characterization of human orthologues to *Saccharomyces cerevisiae* Upf2 protein and Upf3 protein (*Caenorhabditis elegans* SMG-4). *Mol. Cell. Biol.* 21, 209–223
- Denning, G. *et al.* Cloning of a novel phosphatidylinositol kinase-related kinase: characterization of the human SMG-1 RNA surveillance protein. *J. Biol. Chem.* (in press)
- Pal, B. *et al.* (2001) Evidence that phosphorylation of human Upf1 protein varies with intracellular location and is mediated by a wortmannin-sensitive and rapamycin-sensitive PI 3-kinase-related kinase signaling pathway. *RNA* 7, 5–15
- Virshup, D.M. (2000) Protein phosphatase 2A: a panoply of enzymes. *Curr. Opin. Cell Biol.* 12, 180–185
- Clissold, P.M. and Ponting, C.P. (2000) PIN domains in nonsense-mediated mRNA decay and RNAi. *Curr. Biol.* 10, R888–R890
- Bernstein, E. *et al.* (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409, 363–366
- Grishok, A. *et al.* (2001) Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* 106, 23–34
- Fire, A. *et al.* (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391, 806–811
- Elbashir, S.M. *et al.* (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* 411, 494–498
- Jones, L. *et al.* (2001) RNA-directed transcriptional gene silencing in plants can be inherited independently of the RNA trigger and requires Met1 for maintenance. *Curr. Biol.* 11, 747–757
- Smardon, A. *et al.* (2000) EGO-1 is related to RNA-directed RNA polymerase and function in germ-line development and RNA interference in *C. elegans*. *Curr. Biol.* 10, 169–178
- Cogoni, C. and Macino, G. (1999) Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase. *Nature* 399, 166–169
- Mourrain, P. *et al.* (2000) Arabidopsis *SGS2* and *SGS3* genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell* 101, 533–542
- Fagard, M. *et al.* (2000) AGO1, ODE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals. *Proc. Natl. Acad. Sci. U. S. A.* 97, 11650–11654
- Dalmay, T. *et al.* (2000) An RNA-dependent RNA polymerase gene in *Arabidopsis* is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell* 101, 543–553
- Bosher, J.M. and Labouesse, M. (2000) RNA interference: genetic wand and genetic watchdog. *Nat. Cell Biol.* 2, E31–E36
- Fraser, A.G. *et al.* (2000) Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* 408, 325–330
- Domeier, M.E. *et al.* (2000) A link between RNA interference and nonsense-mediated decay in *Caenorhabditis elegans*. *Science* 289, 1928–1931
- Dalmay, T. *et al.* (2001) SDE3 encodes an RNA helicase required for post-transcriptional gene silencing in *Arabidopsis*. *EMBO J* 20, 2069–2078
- Wu-Scharf, D. *et al.* (2000) Transgene and transposon silencing in *Chlamydomonas reinhardtii* by a DEAH-Box RNA helicase. *Science* 290, 1159–1163
- de la Cruz, J. *et al.* (1999) Unwinding RNA in *Saccharomyces cerevisiae*: DEAD-box proteins and related families. *Trends Biochem. Sci.* 24, 192–198
- Atkin, A.L. *et al.* (1995) The majority of yeast UPF1 co-localizes with polyribosomes in the cytoplasm. *Mol. Biol. Cell* 6, 611–625
- Holtorf, H. *et al.* (1999) Stochastic and nonstochastic post-transcriptional silencing of chitinase and beta-1,3-glucanase genes involves increased RNA turnover – possible role for ribosome-independent RNA degradation. *Plant Cell* 11, 471–784

- 52 Zamore, P.D. *et al.* (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* 101, 25–33
- 53 Plasterk, R.H. and Ketting, R.F. (2000) The silence of the genes. *Curr. Opin. Genet. Dev.* 10, 562–567
- 54 Bloom, L. and Horvitz, H.R. (1997) The *Caenorhabditis elegans* gene *unc-76* and its human homologs define a new gene family involved in axonal outgrowth and fasciculation. *Proc. Natl. Acad. Sci. U. S. A.* 94, 3414–3419
- 55 Rougvie, A.E. and Ambros, V. (1995) The heterochronic gene *lin-29* encodes a zinc finger protein that controls a terminal differentiation event in *Caenorhabditis elegans*. *Development* 121, 2491–2500
- 56 Furuta, T. *et al.* (2000) EMB-30: an APC4 homologue required for metaphase-to-anaphase transitions during meiosis and mitosis in *Caenorhabditis elegans*. *Mol. Biol. Cell* 11, 1401–1419
- 57 Laughton, D.L. *et al.* (1997) Alternative splicing of a *Caenorhabditis elegans* gene produces two novel inhibitory amino acid receptor subunits with identical ligand binding domains but different ion channels. *Gene* 201, 119–125
- 58 Kim, V.N. *et al.* (2001) Role of the nonsense-mediated decay factor hUpf3 in the splicing-dependent exon-exon junction complex. *Science* 293, 1832–1836
- 59 Lykke-Andersen. *et al.* (2001) Communication of the position of exon-exon junctions to the mRNA surveillance machinery by the protein RNPS1. *Science* 293, 1836–1839
- 60 Gingras, A.C. *et al.* (1999). eIF4 initiation factors: effectors of recruitment to ribosomes and regulators of translation. *Annu. Rev. Biochem.* 68, 913–963
- 61 Knight, S.W. and Bass, B.L. (2001) A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. *Science* 293, 2269–2271
- 62 Hammond, S.M. *et al.* (2001) Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* 293, 1146–1150
- 63 Parrish, S. *et al.* (2000) Functional anatomy of a dsRNA trigger. Differential requirement for the two trigger strands in RNA interference. *Mol. Cell* 6, 1077–1087
- 64 Hammond, S.M. *et al.* (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* 404, 293–296

# Making mitochondrial mutants

Howard T. Jacobs

**Mitochondrial DNA (mtDNA) encodes a mere 13 polypeptides, all with well-defined cellular functions in mitochondrial energy metabolism. It was first sequenced over two decades ago, yet our understanding of the wider physiological role of mtDNA is surprisingly sketchy. Partly, this reflects the fact that the mitochondrial gene products are essential for life; that is, most mtDNA mutations are expected to be lethal. The technical difficulty of engineering mtDNA mutations has been a major handicap in furthering our understanding of the mitochondrial genetic system. Recent developments now offer some possibilities for the genetic manipulation of mtDNA and for elucidating its contribution to human development, physiology and disease.**

Mitochondrial DNA (mtDNA) is an independent genome within eukaryotic cells. Although human mtDNA was sequenced over two decades ago<sup>1</sup>, the fact that it has not been possible to manipulate it at the sequence level means that huge gaps remain in our understanding of the mitochondrial genetic system.

The maternally inherited mtDNA is a 16 569-bp circular molecule, organized compactly into large, overlapping, POLYCYSTRONIC (see Glossary) transcription units. The primary transcripts are processed into two rRNAs, 22 tRNAs and 11 mRNAs, the last encoding 13 of the 80 or so polypeptides of the mitochondrial OXPHOS complexes (Fig. 1). The mitochondrial rRNAs and tRNAs contribute to a separate translation system within the organelle that is dedicated to the synthesis of these polypeptides. Mitochondrial protein synthesis requires, in addition, well over 100 dedicated nuclear gene products.

Even though their cellular functions are known, detailed knowledge of the biology of the mtDNA-encoded polypeptides and RNAs is limited. To investigate structure–function relationships in these gene products, we have needed to rely upon indirect

## Glossary

**Atresia:** Oocyte degeneration by programmed cell death. The majority of oocytes are lost to this process during development/ageing.

**Biolistic transformation:** Introduction of DNA using bombardment with microprojectiles coated with DNA (also known as ‘the gene gun’).

**Chloramphenicol:** An antibiotic that inhibits the elongation step of translation on bacterial and mitochondrial ribosomes. Mutations to chloramphenicol resistance occur on the large subunit rRNA.

**Cybrid:** A genetic hybrid containing the nuclear genome from one source but the cytoplasmic (i.e. mitochondrial or chloroplast) genome(s) from another.

**Homoplasmy (homoplasmic):** Genetic uniformity of the cytoplasm; that is, all copies of mtDNA identical in sequence within a single cell or individual.

**Heteroplasmy (heteroplasmic):** Genetic heterogeneity of the cytoplasm; that is, presence of two different types of mtDNA that differ in sequence within a single cell or individual.

**Mitotic segregation:** Shift in relative amounts of heteroplasmy with eventual resolution to homoplasmy, as a result of random partition of mtDNA to daughter cells.

**OXPHOS:** Oxidative phosphorylation, requiring the five multisubunit complexes of the inner mitochondrial membrane.

**Polycistronic:** Containing coding regions of many different genes within a single primary transcript.

**Polymorphy:** Genetic variation within a population of individuals.

**Polyplasmy (polyplasmic):** Multiple heteroplasmy; that is, many different sequence variants of mtDNA present within a single cell or individual.

**Replicon:** An autonomously replicating segment of DNA.

**p<sup>0</sup>:** A cell, cell line or (yeast) strain devoid of mtDNA.

**p<sup>+</sup>:** A yeast strain containing intact mtDNA, although it can carry point mutations.

**p<sup>-</sup>:** A yeast strain containing massively deleted mtDNA, organized in tandem repeats of the retained portion of the mitochondrial genome.

**Synaptosome:** Preparation of subcellular fragments of the synaptic region of neurons, rich in mitochondria.

Howard T. Jacobs  
Institute of Medical  
Technology and Tampere  
University Hospital,  
FIN-33014 University of  
Tampere, Finland.  
e-mail:  
howj.jacobs@uta.fi