Nonsense-mediated mRNA reduction and pre-mRNA processing in drosophila

Thesis

UNRESTRICTED

Nonsense-mediated mRNA reduction and pre-mRNA processing in Drosophila

Saverio Brogna

Thesis submitted in fulfilment of the requirements of the Open University for the degree of Doctor of Philosophy

Sponsoring Establishment: Addenbrooke's NHS Trust, Cambridge
Collaborating Establishment: Department of Genetics, University of Cambridge

Discipline: Molecular Biology

June 2000

AUTHOR'S NO : T0138S22
DATE OF SUBMISSION: 17 JUNE 2000
DATE OF AWARD: 10 AUGUST 2000

PREFACE

This thesis is based exclusively on my work and none of it has been submitted for any other degree or qualification. Part of it has previously been published in a paper in which I'm the only author (Brogna, 1999 - see bibliography)

The thesis consists of about 27,000 words and 21 figures.

ACKNOWLEDGMENTS

Firstly, I would like to thank Michael Ashburner for his support during the years I have been working in his laboratory. Secondly, I'm grateful to both Michael and David Gubb for their good will when registering for this PhD.

I would also like to thank Daniel Barbash and David Gubb for reading a draft of the thesis, which much improved the clarity of this dissertation. I'm also gratefull to Mike Millward for proofreading part of this thesis and finally thanks to Louise for her continuous help with my chronic inability to proper learn English.

CONTENTS

SUMMARY

From bacteria to mammalian cells, the presence of a nonsense mutation causes a reduction in the level of the mRNA of the corresponding gene. The reduction is not, contrary to initial expectations, due to a passive mechanism by which non translated mRNAs are degraded; rather it is a active process in which active translation, cis-acting sequences and specific trans-acting factors are required. It is generally accepted that this phenomenon is the consequence of an evolutionary conserved mechanism that evolved to protect cells from the potentially deleterious effect of truncated proteins - this is often referred to as the mRNA surveillance system or nonsense mediated mRNA decay (NMD). This phenomenon has been extensively studied in budding yeast and in mammalian systems and to a lesser extent in *C. elegans*. In yeast the recognition of the nonsense codon appears to occur during cytoplasmic translation and premature translation termination is thought to activate a specific protein complex - called the surveillance complex - which in turn triggers an accelerated decay of the aberrant mRNA.

However, contrary to the expectation that the recognition of the nonsense codon should occur during cytoplasmic translation, several studies in mammalian cells indicate that NMD may take place in the nucleus by a mechanism that is independent of cytoplasmic translation. For example, several reports indicate that this reduction occurs while the mRNA is still associated with the nucleus, and that the stability of the cytoplasmic mRNA is unchanged relative to a wild-type allele.

The common view in the field is that these apparently discordant results between NMD in yeast and in mammalian cells will eventually be accommodated in a single model in which translation in the cytoplasm plays a prominent role. For example, a commonly given explanation is that the recognition of the nonsense codon takes place during nuclear export, and it has been implied that the apparent effects on nuclear RNA are in fact triggered by the premature abortion of translation at the cytoplasmic side of the nuclear envelope.

However not all the data from mammalian systems can be so easily explained by the above model. For example, several reports indicate that nonsense mutations affect the splicing of the corresponding pre-mRNA, which makes it difficult to imagine how premature translation in the cytoplasm could effect such an early event in mRNA biogenesis.

In summary, although it is a well established fact that premature termination codons can cause a drastic reduction of the mRNA level, it is still controversial whether the mechanism which causes this reduction is similar or not in all organisms. Furthermore it is unclear whether such mechanism(s) are exclusively cytoplasmic (and always linked to translation), or also nuclear (and independent of cytoplasmic translation).

This thesis represents the work that I have done to characterise the NMD phenomenon in *Drosophila* and to try to understand the underlining mechanism(s).

During this work, I found that nonsense mutations in the *Alcohol dehydrogenase* gene not only reduce the level of mRNA, but also lead to a longer poly(A) tail in both the mature mRNA and in the unspliced transcript. The effects of nonsense mutations on mRNA processing are happening in-*cis*, since in heterozygous flies only the transcript of the allele that carries the nonsense mutation is affected, while the mRNA from the other allele is not.

Since a longer poly(A) tail is usually associated with stable transcripts, I suggest that the reduction of the level of mRNA rather than reflecting a change in cytoplasmic stability may actually be the secondary consequence of an increase in abortive pre-mRNA processing events; this hypothesis is also supported by the observation that nonsense mutations in *Adh* are also associated with an increase of unspliced transcripts.

Collectively my data indicate that the reduction of mRNA level occurs in the nucleus and the lower level of mRNA is probably a consequence of an inhibitory effect of premature stop codons on pre-mRNA processing rather than being the consequence of a reduced stability of the mature transcript. These findings, therefore suggest that the integrity of the coding region may be in-*vivo* required for accurate pre-mRNA processing.

During this work I have also cloned and partially characterized the *Drosophila* homolog of *Upf1* one of the three "master "genes thought to be involved in the reduction of mRNA harboring a premature stop codon in yeast and other organisms. The cloning of *Upf1* in *Drosophila* opens the possibility of testing whether *Upf1* is also involved in this process in *Drosophila*.

Chapter 1

INTRODUCTION

This chapter is intended to provide a background to what is known about the phenomenon of nonsense-mediated mRNA reduction. The first part is a brief outline of the general facts and questions of the field without an in-depth review of the published studies. The second part, which is divided into four paragraphs, is a more detailed review of the studies in S. ceravisiae, C. elegans, *mammalian cells and bacteria. The chapter ends with a brief introduction to the* Alcohol dehydrogenase (Adh) and Adh-related (Adhr) *genes of* Drosophila melanogaster, *which are used as a model system during my study.*

Nonsense mutations are nucleotide substitutions that change a coding triplet into one of the three termination codons, which in turn leads to premature translation termination [1].

[1] Nonsense mutations were first found in bacteria and originally defined as a class of mutations in many different genes that could be suppressed by the same suppressing strain (*su*$^+$, later found to carry a mutation in a tRNA). The first mutation of this class to be

Premature translation termination can also be the secondary consequence of insertions or deletions that cause a frame-shift. Mutations that prevent pre-mRNA splicing can also cause premature termination within intronic sequences, which tend to be A/T rich and therefore have a high frequency of stop-codons.

Nonsense mutations are often associated with a complete lack of gene function and this is usually thought be the direct consequence of producing a truncated polypeptide.

---

identified was called *amber*. It was proposed that amber mutations may give rise to a nonsense codon which interrupts the translation of the genetic message, and that in the su^+ strain this codon can be misread so that the whole cistron is translated (Benzer and Champe 1962). Later it was indeed demonstrated that *amber* mutants lead to either no detectable protein or to a truncated polypeptide (Garen and Siddiqi 1962; Sarabhai et al. 1964). The three triplets that serve as termination codons were first identified by genetical analysis of nonsense mutations in the bacteriophage T4 (Stretton et al. 1966; Brenner et al. 1967). The"nonsense codons" are traditionally known as *amber* (UAG), *ochre* (UAA) and *opal* (UGA) mutations; a nomenclature dating back to the early days of *E. coli* genetics, *amber* being the name of the German aunt of the person who first discovered such mutations. Following this style the other two types were named after semi-precious stones. Nonsense mutations can be suppressed in a given suppressor strain and on the basis of which strain is required for suppression they can be distinguished into the three phenotypic classes.

However, although nonsense mutations can produce truncated polypeptides, frequently the peptide is completely absent. The lack of peptide may be due to the fact that the truncated peptide is unstable. However in several instances, the primary reason for the lack of peptide is that the level of the corresponding mRNA is drastically reduced.

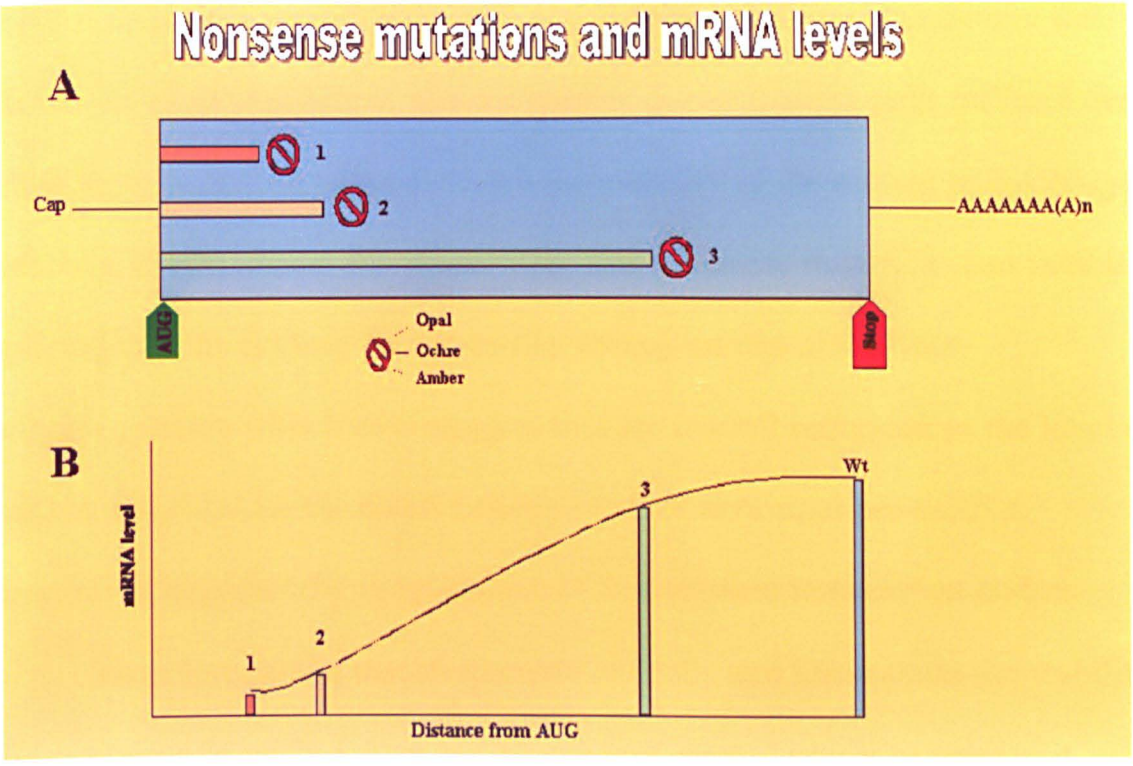
In general, the magnitude of mRNA reduction depends on the position of the mutation within the open reading frame: mutations in the 5' portion of the ORF usually lead to a severe reduction while mutations close to the 3' end portion of the ORF have only minor effects. This "polarity effect" is more apparent in yeast and in Drosophila (this work); Figure 1 is a cartoon of a typical example of this 'polarity' effect.

Nonsense-mediated mRNA reduction has been observed in essentially all organisms where it has been looked for, and so far a considerable amount of data has been produced from studies in *Saccharomyces cerevisiae*, mammalian cells and *Caenorhabditis elegans*.

This phenomenon has also been observed in bacteria, however here the reduction of the level of mRNA is more often attributed to a reduced level of transcription caused by premature RNA polymerase termination. Nonetheless, as shown below, there are data to suggest that in bacteria also, nonsense mutations can lead to a post-transcriptional mRNA reduction.

It is commonly thought that the mRNA reduction is the result of a specialized pathway which degrades aberrant mRNAs that otherwise could encode for potentially deleterious truncated polypeptides. This mechanism, which is often called the *mRNA surveillance system,* is viewed as a kind of cellular quality control system, analogous to that involved in the degradation of abnormally folded proteins (Wickner et al. 1999; Schubert et al. 2000). This mechanism is thought to work during the process of translation and to be able to distinguish premature stop codons from the proper termination codon.

Figure 1



A - A cartoon of a typical eukaryotic mRNA. Rectangles are ORFs; the largest rectangle corresponds to the wild-type ORF, the smaller rectangles are the shorter ORF generated by the nonsense mutations (indicated by the prohibition sign). **B** - The relative level of mRNA in the wild-type and in the corresponding nonsense mutants.

It is thought that a main substrate for the surveillance mechanism may be unspliced transcripts which escape to the cytoplasm, or transcripts with abnormally extended 3' UTR (Muhlrad and Parker 1999; Das et al. 2000).

The phenomenon is commonly called NMD (nonsense-mediated mRNA decay) a yeast terminology that stresses the common view that the mRNA reduction is a consequence of a reduce stability of the mature mRNA. However, as shown below, several studies in mammalian cells indicate that often there is no difference between the stability of the mutant and wild-type mRNAs. Furthermore, the observation that nonsense mutations can inhibit splicing and the finding that nonsense mutations can also affect polyadenylation (this study) suggest that the overall reduction in the level of mRNA may also be the direct consequence of abnormal pre-mRNA processing triggered by the presence of a premature termination codon.

Therefore, given that in mammalian cells and Drosophila the stability of the cytoplasmic mRNA may not be affected, here, I prefer to refer to it, as nonsense-mediated mRNA reduction (NMMR), to emphasize that the mechanism of mRNA reduction is not yet demonstrated and that it may not necessarily involve a rapid decay of the mature mRNA.

Nonsense mutations in yeast.

In yeast, the first observation that nonsense mutations can lead to a reduction of the level of the corresponding mRNA, came from early studies with the *URA3* gene (Losson and Lacroute 1979) (*URA3* encodes for orotidine monophosphate decarboxylase). Losson and Lacroute (1979) mapped several

amber mutations along the *URA3* gene and they found that mutations in the first half of the gene caused a drastic reduction of the steady-state level of *URA3* mRNA; while the mutations further downstream had only minor effects on the level of mRNA. Furthermore, these authors found that a 2-min pulse labeling with [^{3}H]-adenine gave similar RNA levels in the different mutants, but with longer labeling times the level of *URA3* RNA decreased more quickly in a strain with a 5' proximal nonsense mutations than in 3' mutants or wild type. In addition, it was shown that an *amber* tRNA suppressor could prevent the rapid mRNA reduction. From these studies the authors concluded that the nonsense mediated reduction of the steady-state level of mRNA is due to reduced stability of the mature transcript; since a tRNA suppressor could prevent this degradation they assumed that it is the lack of translation that causes the premature decay. Losson and Lacroute (1979) suggested that ribosomes, by simply covering the mRNA, would protected it from degradation by non-specific nucleases. This model was an adaptation of the early model proposed to explain the phenomenon of transcription polarity in bacteria (Morse and Yanofsky 1969) - see the paragraph below on nonsense mutations in bacteria.

More recently, NMMR has been investigated by inserting nonsense mutations at various positions in three other genes: *PGK1*, *HIS4* and *CYC1*

(Peltz et al. 1993; Hagan et al. 1995)[Yun, 1996 #1443 [2]. These studies show that premature termination codons, located approximately in the first half of the coding region, lead to a drastic reduction in the level of mRNA. For example nonsense mutations located within in the fist 2/3 of the coding region of PGK1 result in a 10- to 20-fold reduction in the level of mRNA (Peltz et al. 1993; Hagan et al. 1995). These authors proposed that the reduction in the level of mutant mRNA is the direct consequence of its reduced cytoplasmic stability - as previously did Losson and Lacroute (1979). The half-life of the mRNA was measured in cells in which a temperature-sensitive RNA polymerase II allele had been inactivated at 37 ^{0}C, and they found that the half life of the mutant mRNA was up to 20 times shorter than that of the wild-type (Peltz et al. 1993).

i)cis-acting sequences.

In order to explain the observation that nonsense mutations in the first half of the gene drastically reduce the level of mRNA, while mutations further downstream have little or no effect (Figure 1), it was proposed that in addition to the premature stop codon, NMMR may also require a downstream sequence element (Peltz et al. 1993). The closer the mutation is to the canonical end of

---

[2] PGK1 codes for 3-phosphoglycerate kinase , HIS4 for histidinol dehydrogenase and CYC1 for iso-1-cytochrome c.

the coding region the less likely it is that this putative destabilizing sequence is present in the remaining downstream region; the lack of this sequence in the 3'UTR could in principle provide a means to distinguish between premature termination codons and normal stop codons (Peltz et al 1993, see below). These sequences have been well characterized in the *PGK1* and *HIS4* genes of *S. ceravisiae* (Peltz et al. 1993; Hagan et al. 1995; Zhang et al. 1995). These studies showed that some deletions of the region downstream of a nonsense mutation can abolish the nonsense mediated mRNA decay. It was also shown that the nonsense mediated mRNA reduction could be reactivated by re-inserting a 106-nt long element; this element is named Downstream Element, or DSE. A major determinant of the DSE consists of two copies of the sequence motif UGYYGAUGYYYYY, in which Y stands for U or C (Zhang et al. 1995).

Interestingly, the DSE element from the *PGK1* gene triggers a rapid reduction in the level of the *GCN4* mRNA when it is inserted in the leader region of the *GCN4* mRNA (Ruiz-Echevarria and Peltz 1996). The *GCN4* coding region is preceded by four short upstream ORFs (uORF), which regulate the expression of this gene. In optimal growth conditions, these uORFs lead to premature translation termination and it was expected that these uORFs would function in manner analogous to nonsense mutations, promoting a rapid decay of the transcript. Transcripts harboring an uORF were

expected to be rapidly degraded because it was previously reported that a mutation that inserts an ATG codon in the leader region of the *CYC1* gene of yeast, which generates a short uORF, leads to an 100-fold reduction of the steady-state level of the mutant mRNA (Pinto et al. 1992). Similar effects were seen when an uORF was inserted upstream of the coding region of the chloramphenicol acetyltransferase gene of yeast (Oliveira and McCarthy 1995).

However, contrary to the expectations the *GCN4* mRNA is rather stable; therefore, it was hypothesized that the *GCN4* leader sequence lacked a DSE. In agreement with this hypothesis, the *PGK* DSE element destabilized the transcript when inserted downstream of uORF1, but not when it was inserted upstream (Ruiz-Echevarria and Peltz 1996), implying that the DSE element per se does not destabilise the mRNA and that it is only functional after a translation initiation-termination cycle has been completed. More recent studies with the *GCN4* leader region have shown that the unexpected stability of the *GCN4* mRNA depends also on a stabilizer sequence element (STE) which inactivates the NMMR pathway when present between the termination codon and the DSE element (Ruiz-Echevarria et al. 1998).

These results also provide an explanation to the observation that inserting a DSE element downstream of a 3'-proximal premature stop codon does not destabilize the transcript (Peltz et al. 1993). This would suggest that

the 3' end of the coding region and probably also the 3' UTR contain a stabilizer sequence (STE) that is capable of counteracting the negative effects of a DSE element.

Recently, it has been proposed that the DSE element can specifically bind to Hrp1p , and by doing so create a specific RNP structure that can be recognized by the translating ribosome (Gonzalez et al. 2000). Interestingly, these authors also reported that mutations in the *HRP1* gene can prevent NMMR (Gonzalez et al. 2000). The Hrp1p protein is highly similar to the human hnRNP-A1/hnRNP-D proteins and is required for pre-mRNA 3'-end formation (Kessler et al. 1997).

In summary, whether a nonsense mutation can or can not trigger the mRNA reduction seems to depend on a set of cis-acting sequence elements that can have either a negative or positive effect on mRNA stability. The function of these cis-elements seems to depends on translation, but, how these sequences influence NMMR remains uncertain.

ii-Trans-acting factors.

Trans-acting factors involved in nonsense mediated mRNA reduction were first identified in a screen for mutations that suppress the frameshift mutation *his4-38*, in a genetic background containing a tRNA frameshift suppressor. The tRNA frameshift suppressor causes a low level of read-through that is

sufficient to give a His^s phenotype at 30 ^{0}C, but not at 37 ^{0}C. Mutations were selected for the ability to give a His+ phenotype at 37^0. These mutations were called *upf*, for up frameshift (Culbertson et al. 1980). Subsequent analysis of three of these *upf* mutations showed that they cause an increase in the level of the *his4-38* mRNA, and also that of other nonsense-containing mRNAs, without affecting the abundance of either the wild-type *HIS4* mRNAs or the mRNA of other genes (for a review see (Ruiz-Echevarria et al. 1996)[3]. Therefore, it is thought that the *upf* mutations identify genes that are specifically required to degrade mRNAs that carry premature termination codons. These genes are: **1)** *UPF1* (Leeds et al. 1991) - also named *MOF4/IFS2* (Cui et al. 1996) or *NAM7* (Altamura et al. 1992); **2)** *UPF2* (Cui et al. 1995) - also called *NMD2* (He and Jacobson 1995) or *SUA1/IFS1* (Cui et al. 1995; Cui et al. 1996); and **3)** *UPF3* (Lee and Culbertson 1995) - also known as *SUA6* (Cui et al. 1995). Mutations in *UPF1* and *UPF2* were also

---

[3] One exception was found in the *URA3* gene; where Leeds et al. (1991) found that a lack of *UPF1* caused a 1.9-fold increase of also the wild-type form. The authors suggested that this unusual result may be due to an increased transcription of *URA3* gene in a *UPF1*- background caused by the stabilisation of *PPR1* which is a transcriptional activator of *URA3*; they mentioned that there were preliminary data indicating that the level of *PPR1* mRNA is increased in a *UPF1*- background. Data supporting this model have not yet been published.

identified in screens for mutations that increased -1 ribosomal frameshift efficiency; these mutations were called either *mof* - standing for maintenance of frame (Dinman and Wickner 1994), or *ifs* - for increased frameshifting (Lee et al. 1995). *UPF2* and *UPF3* were also hit in a screen for mutations affecting translation initiation; these mutations were called *sua* - for suppressor of upstream ATG (Hampsey et al. 1991; Pinto et al. 1992).

The fact that mutations in *UPF1*, *UPF2* and *UPF3* are phenotypically similar suggests that they are involved in the same cellular process and the suppression phenotypes strongly indicate that they may be directly involved in translation.

Recently, it has been reported that similarly to *UPF1*, *UPF2* and *UPF3*, some mutants of *MOF2/SUI1*, *MOF5*, and *MOF8* also lead to a stabilization of aberrant mRNAs (Cui et al. 1999) - as mentioned above the *mof* mutants were isolated in a screen for suppressors of frameshift mutations; *sui* mutants were isolated in a screen for suppression of a *HIS4* alleles lacking the AUG initiator codon (Castilho-Valavicius et al. 1990). Interestingly, *MOF2/SUI1* encodes for eIF1, a factor that has been recently shown to be essential for cap-mediated initiation of translation in higher eukaryotes (Pestova et al. 1998). More recently a mutation in the *PRT1* gene, which encodes the p90 subunit of eIF3 - another essential translation initiation factor - was also shown to stabilize aberrant mRNAs (Welch and Jacobson 1999).

Among these genes, the first to be characterised and the most intensively studied is *UPF1* (Leeds et al. 1992). The UPF1 gene encodes for 109-kDa protein which contains a region rich in cysteines and histidines near the amino terminal and a sequence domain characteristic of the superfamily I of helicases (Hodgman 1988; Altamura et al. 1992). The yeast Upf1p has been purified and appears to have RNA binding, RNA-dependent ATPase and RNA helicase activities (Czaplinski et al. 1995; Weng et al. 1998).

As mentioned above mutants in the *UPF1* gene (as in *UPF2* and *UPF3*) were first isolated for their ability to efficiently co-suppress the *his4-38* frameshift allele in a strain carrying a tRNA frameshift suppressor. Since mutations or deletions of *UPF1* also caused an increase in the level of *his4-38* mRNA, it was proposed that the His^+ phenotype was a consequence of enhancing the stability of the mRNA, such that the increased level of the *his4-38* mRNA would simply enhance the effect of the tRNA suppressor [Leeds, 1991 #1294]. However, it was recently reported that mutations in the amino terminal cysteine- and histidine-rich domains of Upf1p do not increase the level of the mutant mRNA, but are still able to suppress nonsense alleles. This suppression phenotype does not require the aid of a tRNA suppressor (Weng et al. 1996b). Furthermore mutations in the conserved helicase motif of Upf1p that inactivate its mRNA decay function (such that the level of mutant mRNA is increased) do not have a nonsense suppression phenotype (Weng et al.

1996a). These results strongly indicate a direct role of Upf1p in translation termination and suggest that Upf1p may contain two distinct functional domains; one may be involved in promoting rapid decay of nonsense-mRNAs, while the other may be involved in translation termination. More recently, it has been shown that Upf1p directly interacts with eRF1 and eRF3, two factors required for translation termination in eukaryotes - eRF1 catalyzes peptide hydrolysis and eRF3 stimulate the activity of eRF1 (Czaplinski et al. 1998). In summary, a considerable body of data indicates that Upf1p is involved in translation termination and it is therefore generally assumed that translation termination may be the key event that regulates mRNA stability and NMMR (Czaplinski et al. 1998). Upf1p interacts with Upf2p and Upf3p, which are thought to be other essential components of the surveillance complex. This complex, as I will discuss below, is thought to be able to distinguish between normal and premature translation termination events and trigger a rapid decay of the mRNA in the latter event (He et al. 1997; Czaplinski et al. 1998).

Genes homologous to the yeast *UPF1* gene have been found in mammalian organisms (Perlick et al. 1996; Applequist et al. 1997), *C. elegans* (Page et al. 1999), and *D. melanogaster* (this study). Interestingly a human/yeast Upf1p hybrid has a nonsense suppression phenotype, implying that the role of Upf1p in translation termination is likely to be conserved in evolution (Czaplinski et al. 1998).

iii) Link between translation and decay: the surveillance complex.

Since the only known mechanism for detecting nonsense codons is ribosomal translation, it is expected that the reduction in the level of nonsense-containing mRNAs should be linked to translation.

Initially, it was proposed that premature translation termination would leave the region downstream from the nonsense mutation free of its normal set of ribosomes and therefore accessible to nonspecific nucleases - in other words, translation would stabilize the mRNA by simply preventing the attack of unspecific nucleases (Morse and Yanofsky 1969; Losson and Lacroute 1979). This model agrees with a number of observations in bacteria that indicate that inhibition of translation in *cis*, for example by impairing the function of the ribosome binding site (RBS), also causes mRNA destabilization (see below). In yeast, in agreement with a passive role of translation on mRNA stability, it has been observed that translation inhibition of the *PGK1* mRNA, by inserting a strong secondary structure in the 5' UTR, can destabilize the transcript to an extent similar to that observed when the mRNA is destabilized by premature transaltion termination (Muhlrad et al. 1995).

However, in yeast and other eukaryotic cells, the prevailing view is that NMMR is an active process in which ongoing translation and specific trans-

acting factors are required. The active role of translation in NMMR is suggested by the observation that inhibition of translation can prevent NMMR (see below) and by the genetical studies described above, which indicate that mutations in some translation factors can suppress NMMR.

The effect of translation inhibition in trans , however, may be an indirect one; for example the treatment of cells with translation inhibitors has pleiotropic effects and can stabilise a variety of mRNAs (Jacobson and Peltz 1996) (Ross 1997). The indirect effect of translation inhibition on RNA stability is suggested by the observation that treatment of cells with translation drugs can also stabilise transcripts lacking an ORF (Lopez et al. 1998). However there are examples in which inhibition of translation in cis by inserting a stem-loop structure in the 5' UTR (Belgrader et al. 1993) or by mutating the AUG initiator codon can suppress NMMR (Yun and Sherman 1996).

The strongest available evidence for translation being actively required for NMMR rests on the finding that mutations in some translation factors can specifically stabilise nonsense-containing mRNAs but not a variety of other transcripts. For example, the observation that Upf1p interacts with the release factors, more precisely suggests a link between premature translation termination and NMMR. The current model for NMMR is that upon encountering a stop codon in the A site, the ribosome pauses and the

translation termination factors eRF1 and eRF3 are recruited. The interaction of the release factors with Upf1p would promote the assembly of a complex which in addition to Upf1p contains Upf2p, Upf3p and most likely other factors. This complex is called the surveillance complex, and in addition to enhancing translation termination it is thought that, probably in conjunction with the small ribosomal subunit, it would keep scanning the region downstream of the stop codon in search of DSE element - the recognition may involve a DSE-dependent RNP rather than the simple sequence motif (Czaplinski et al. 1999; Hilleren and Parker 1999; Gonzalez et al. 2000). If a DSE element is found the surveillance complex triggers decapping of the transcript and rapid decay (Muhlrad and Parker 1994).

In yeast the majority of mRNA decay proceeds first with the shortening of the poly(A) tail and then decapping. Decapping is followed by a rapid 5' - to- 3' exonucleolytic degradation (Decker and Parker 1993). In contrast nonsense-containing mRNAs seem to be degraded without prior deadenylation (Muhlrad and Parker 1994). This suggests that nonsense-containing mRNAs may be degraded through a pathway somehow different from that followed by normal mRNAs; the surveillance complex could for example enhance decapping independently of deadenylation (Czaplinski et al. 1998). This difference between the decay intermediates in normal and nonsense-

containing mRNAs is indeed often used as evidence that NMMR in yeast is the result of a specialized pathway of mRNA decay.

Nonsense mutations in C. elegans.

The study of NMMR in *C. elegans* originated while investigating the molecular basis of the suppression phenotype of *smg* mutations (Pulak and Anderson 1993). The *smg* genes were first identified as allele specific, but not gene-specific, suppressors of mutations in a variety of genes. *smg* mutants have a mild morphogenetic phenotype of the genitalia - the acronym *smg* stands for suppressor with a morphogenetic effect on genitalia (Hodgkin et al. 1989). *smg* mutations suppress the reduction in the level of mRNAs that carry nonsense mutations (Pulak and Anderson 1993). Therefore, it has been proposed that the *smg* genes may be involved in the rapid decay of aberrant mRNAs, and may correspond to the *Upf* gene system of yeast described above. Interestingly, certain recessive mutations become dominant in a Smg(-) background(Cali and Anderson 1998). Some recessive *unc-54* alleles (*unc-54* encodes for the myosin heavy chain) become dominant in a Smg(-) background, probably by producing a truncated polypeptide that interacts with the wild-type UNC-54 protein and causes abnormal sarcomere assembly (Pulak and Anderson 1993).

Based on these genetic observations it has been proposed that the *smg* genes represent a specialised system of mRNA decay that protects the cell from the potentially toxic polypeptides that could be produced by aberrant transcripts. This mechanism is called the mRNA surveillance system - a teleology first used by Pulak and Anderson (1993) and then has been widely used to justify why NMMR is evolutionary conserved (the concept of the existence of a mRNA surveillance system will be analyzed in the discussion).

Some of the *smg* genes have been molecularly characterised. *smg-2* is the homolog of yeast *UPF1* and the protein is more similar to the mammalian Upf1p homolog than to the yeast protein (Page et al. 1999). The *smg2/UPF1* homology is consistent with the hypothesis that the *smg* genes in *C. elegans* and *Upf* genes in yeast are involved in a similar mechanism.

Interestingly, it has been found that SMG-2 is a phosphorylated protein, and its phosphorylation is affected by mutations in other *smg* genes implying that phosphorylation may regulate SMG-2 function (Page et al. 1999).

Two other *smg* genes have been cloned, *smg-7* and *smg-5*, but are novel proteins without counterparts in the sequence databases (Cali et al., 1999; Page et al., 1999).

The studies in yeast, suggest that the *smg* genes also function in the cytoplasm in concomitance with translation. It has been reported, however, that *smg* mutations can affect the ratio between alternative spliced isoforms in

two SR genes (Morrison et al. 1997). This effect probably involves the stabilization of one of the isoforms in a Smg(-) background, however an abnormal splicing regulation, in the *smg* mutants, can not be excluded .

Nonsense mutations in mammalian cells.

Hundreds of human genetic diseases and inherited cancers are caused by mutations that generate premature termination codons (Atkinson and Martin 1994). In fact the first report of NMMR in mammalian systems was the observation that in a form of β^0-thalassemia caused by a nonsense mutation in the β-globin gene, the level of the corresponding mRNA was drastically reduced (Chang and Kan 1979). Following these early studies NMMR has been studied using several other genes (reviewed by Maquat, 1995).

A characteristic feature of NMMR in mammalian cells is that the level of nucleus-associated mRNA is often reduced to the same extent as the cytoplasmic mRNA fraction (Maquat 1995). A reduction of the mRNA that co-purifies with nuclei has been observed in transcripts coding for β-globin (Baserga and Benz 1992), dihydrofolate reductase (Urlaub et al. 1989), triose phosphate isomerase (Belgrader et al. 1994), the mouse major urinary protein (Belgrader and Maquat 1994), heavy and light chain immunoglobulins (Lozano et al. 1994), and T-cell receptors (TCR) (Carter et al. 1996).

Furthermore, in most of these systems it is apparent that there is no difference between the cytoplasmic half-life of mRNAs carrying premature stop codons relative to the corresponding wild-type form (Cheng and Maquat 1993) (Urlaub et al. 1989; Baserga and Benz 1992).

The observation that nonsense codons affect the level of RNA which co-purifies with the nucleus raises a paradox, since the cytoplasmic ribosome is the only known structure that can recognize the codons of an ORF. However, despite this apparent nuclear localization, several data suggest an involvement of the ribosome in NMRR. For example protein synthesis inhibitors and tRNA suppressors are able to reverse the nonsense-dependent mRNA down regulation (Maquat 1995; Li et al. 1997a; Li et al. 1997b). Furthermore it has been reported that translation inhibition in cis can also prevent NMMR. For example, both a strong secondary structure in the 5' UTR (Belgrader et al. 1993; Kugler et al. 1995) and translation inhibition by binding of a translation repressor in the 5' UTR (Thermann et al. 1998) can prevent NMMR.

Therefore, it may be that the recognition of the nonsense codon may take place during nuclear export, and it has been implied that the effects on nonsense mutations on mRNA and pre-mRNA (see below) are triggered by the premature abortion of translation at the cytoplasmic side of the nuclear envelope(Maquat, 1995). Currently, the most accepted model is that the first

round of mRNA translation occurs near the nuclear envelope while the mRNA is still associated with the nucleus; if a premature termination codons is encountered in a "sensitive region" of the mRNA a rapid decay occurs while the mRNA is still associated with the nucleus (Maquat 1995; Hentze and Kulozik 1999). In addition, according to the observations that NMMR in mammalian cells does not involve a reduced cytoplasmic stability of the mRNA, the model proposes that if the nonsense mutation is not recognised during this first round of translation or if the mRNA skips this first round then the mRNA will escape NMMR and it will have a normal cytoplasmic stability. In contrast, in yeast there are no indications that that NMMR is restricted during the first round of translation or in concomitance to nuclear export; instead the nonsense-containing mRNAs which are substrates for NMMR are thought to be polysome-associated (Zhang et al. 1997).

An alternative explanation to the nucleus-associated mRNA decay is that premature termination codons may affect pre-mRNA processing events other than decay; and that the mRNA reduction may be a consequence of inhibition (or alteration) of pre-mRNA processing. This hypothesis is based on a number of reports which indicate that nonsense mutations can affect splicing as well as reducing the level of mRNA. The most extensively studied is the fibrillin (*FBN1*) pre-mRNA in which it has been observed that nonsense but not missense mutations in exon 51 lead to its skipping (Dietz and Kendzior

1994). Nonsense mutations in FBN1 are associated with some forms of Marfan syndrome (Dietz et al. 1993). A normal splicing pattern of the *FBN1* pre-mRNA was restored when the nonsense mutation was shifted out of frame with the initiation AUG - which implies that the effect of the mutation is likely to be ORF-dependent rather than reflecting changes in cis-acting elements that may have a role in splicing (e.g., exon splicing enhancer, ESE) or that may fortuitously inhibit splicing (e.g., inhibitory secondary structures).

Similar conclusions were reached in studies with the parvovirus minute virus of mice (MVM), where it has been shown that premature termination codons interfere with an exon splicing enhancer in an ORF dependent fashion: Nonsense mutations, but not missense mutations, cause exon skipping, and give rise to a change in the ratio of two alternative spliced transcripts (Gersappe and Pintel 1999).

Further support for a nuclear explanation for NMMR comes from studies with transcripts from Ig genes, which indicate that nonsense codons can inhibit splicing of an upstream intron, both in vivo and in nuclear extracts (Lozano et al. 1994; Aoufouchi et al. 1996). Unfortunately the in vitro studies are tecniqually difficult to repeat and therefore the results by Aoufouchi et al. (1996) remain controversial in the field.

The effects of NMMR of pre-mRNA processing appear to be happening in cis, since in cases where both the mutant RNA and the wild-type

are present only the mutant allele is affected (Maquat, 1995); this appears to exclude a feed-back mechanism between premature translation termination in the cytoplasm and pre-mRNA processing in the nucleus. Taken together these studies therefore suggest that nonsense mutations can be recognised in the nucleus independently of translation.

The existence of such a nuclear scanning machinery was first hypothesised by Urlaub et al. (1989); according to this model, the nuclear mRNA and pre-mRNA can be scanned for premature termination codons by a ribosome-like structure that is able to read the triplets. This later hypothesis is of course very controversial, but it would provide a simple explanation for most of the features of NMMR in mammalian cells.

Another link of NMMR to splicing is suggested by studies with the β-globin and TPI genes, which indicate that an intron downstream of the nonsense codon is necessary for NMMR (Carter et al. 1996; Thermann et al. 1998; Zhang et al. 1998a; Zhang et al. 1998b), and that a wild-type termination codon can be converted to a premature-like termination codon by inserting a downstream intron (Carter et al. 1996; Thermann et al. 1998). The presence of an intron might also provide a mechanism to distinguish a premature termination codon from the canonical stop codon in mammalian mRNAs (Nagy and Maquat 1998); this hypothesis is supported by the

observation that the vast majority of genes do not have introns in the 3' untranslated region (Hawkins 1988).

The splicing requirement for nonsense mediated mRNA reduction can, however, also be explained by a two phase mechanism that involves the tagging of the position of the exon-exon junction in the nucleus and recognition of this tag during translation in the cytoplasm (Hentze and Kulozik 1999). In support of such a "nuclear-tagging" is the observation that a number of proteins known to bind the pre-mRNA and to be involved in splicing, such hnRNPA1 and a specific subset of SR proteins, remain associated with the mature mRNA and shuttle between the nucleus and the cytoplasm (Nakielny and Dreyfuss 1997; Caceres et al. 1998). Interestingly, very recently, it has been shown that two splicing factors, SRm160 and hPrp8p, remain associated with the exon-exon junction after the release of the RNA from the spliceosome (Le Hir, et al. 2000). This kind of tagging of the exon/exon junction could be the functional equivalent to the DSE of yeast; it is possible that both the DSE and the exon/exon junction are bound by protein factors in the nucleus that remain associated with the mRNA during nuclear export; this RNP could be recognised by the translating ribosome .

However, despite different phenomenologies in yeast and mammalian cells, a mammalian orthologue of Upf1 has been cloned and shown to partially complement the Upf1⁻ deficiency phenotype in yeast (Perlick et al., 1996;

Applequist et al., 1997). A human/yeast UPF1 hybrid protein can prevent nonsense suppression in a *UPF1* deleted yeast strain; but this hybrid does not activate NMMR (Czaplinski et al. 1998). These results suggest that the translation termination role of UPF1 has been conserved, but, interestingly, probably not its role in NMMR. However, it has been reported that a mutated human Upf1 protein has a dominant-negative effect on NMMR in mammalian cells; the mutant protein rescued both the cytoplasmic and nuclear levels of nonsense containing mRNAs (Sun et al. 1998).

In conclusion, the finding of Upf1 homologs in higher organisms and the fact that translation is the only known means of reading the triplets have encouraged several efforts to construct a model which links all the features of NMMR in mammalian cells to the function of UPF1 and, more in general, to cytoplasmic translation (Hentze and Kulozik 1999). However, several features remains enigmatic, in particular the fact that nonsense mutations can affect splicing (see discussion).

Nonsense mutations in bacteria.

The first indication that nonsense mutations may lead to a reduction of mRNA levels come from studying polar mutations in *E. coli*. In bacteria, protein-encoding genes are often organised into polycistronic transcription units, and it was observed that mutations in a cistron not only inactivate that gene

product, but also lead to a parallel reduction in the expression of the downstream cistrons (Franklin and Luria 1961; Jacob and Monod 1961). This phenomenon is known as a polar effect or polarity, and the more distant the mutation is from the next cistron the greater is the effect on the distal genes (Newton et al. 1965). Polar mutations are usually nonsense or frame-shift mutations.

The current model to explain the negative effect of nonsense mutations is based on the fact that in bacteria translation and transcription are functionally linked, and ribosomes bound to the nascent transcript are necessary for efficient transcription elongation. This model proposes that within the upstream cistron there are cryptic transcription termination signals that are normally masked by the ribosomes but become unprotected if translation is prematurely terminated by an upstream nonsense mutation. In agreement with this model is the observation that mutations in Rho, a protein involved in termination of transcription in *E. coli*, reduce or abolish polarity (Das et al. 1976). Therefore the model proposes that genes contain Rho-dependent termination sites that are normally covered by the ribosomes, but if translation is prevented, for example by an upstream nonsense codon, then Rho binds to the resulting naked nascent transcript and terminates transcription (Platt and Bear 1983)

In summary, in bacteria the reduction of the level of mRNA containing premature stop appears to result from a reduced synthesis caused by premature RNA polymerase termination.

However, there are other studies that indicate that nonsense mutations or the use of less efficient RBSs (Ribosome Binding Sequence) also cause a rapid decay of the mRNA without affecting transcription termination (Nilsson et al. 1987; McCormick et al. 1994). These studies seem consistent with an earlier model used to explain polarity in the *trp* operon (Morse and Yanofsky 1969).In this study it was proposed that premature translation termination triggers a rapid decay of the nascent transcript which would be unprotected by the closely-packed complement of ribosomes and therefore accessible to nucleases (Morse and Yanofsky 1969). This rapid decay may be caused by RNase E cleavage of the nascent transcript at sites normally protected by ribosomes (Iost and Dreyfus 1995; Braun et al. 1998).

Studies with the *ompA* gene of *E. coli* showed that nonsense mutations very close to the 5' end triggered a rapid decay of the mRNA, but mutations just 30 codons further downstream resulted in a normally stable transcript (Nilsson et al. 1987). This work suggests that translation of only a relatively short portion at the 5' end is sufficient to stabilize the whole transcript and that depriving the transcript of its usual set of ribosomes does not necessarily destabilise the transcript. These results may be explained by the initiation of

mRNA decay beginning preferentially at, or near to, the 5' end of the transcript, so that blocking or sequestering the 5' end of the substrate might spare it from attack (Bouvet and Belasco 1992; Mackie 1998).

In conclusion, in bacteria, similarly to other organisms, premature termination codons result in a reduced level of the corresponding transcript. This appears to be either a consequence of desynchronizing transcription from translation, which can causes premature transcription termination, or an accelerated mRNA decay due to inefficient translation of the first portion of the ORF. A reduced efficiency of translation initiation and premature termination in the initial portion of the ORF have similar phenotypic consequences - both result in mRNA reduction. Furthermore, mRNA decay in bacteria appears to be a direct consequence of the lack of translation of a significant portion of the transcript, which depleted of ribosomes becomes more sensitive to the attack of RNases. Therefore in bacteria there is no evidence of a mechanism that specifically recognizes premature stop codons and triggers a rapid decay of the aberrant transcript.

In bacteria, however, there is a clear example in which the reduction of the level of mRNA associated with carrying nonsense mutations is the consequence of reducing the rate of splicing, rather than a reduced stability of the mature mRNA. Translation inhibitors like aminoglycoside antibiotics can inhibit the splicing of the group I intron in the thymidylate synthase (*td*) gene

of T4 phage in *E. coli* (Waldsich et al. 1998). Interestingly, premature stop codons introduced into the first exon of the *td* gene were also found to inhibit splicing, but did not affect the stability of the mature mRNA (Semrad and Schroeder 1998).Therefore, these data demonstrate that in vivo splicing of the *td* intron is not directly inhibited by aminoglycosides, but rather indirectly by the interference with its translation.

The *Adh* system.

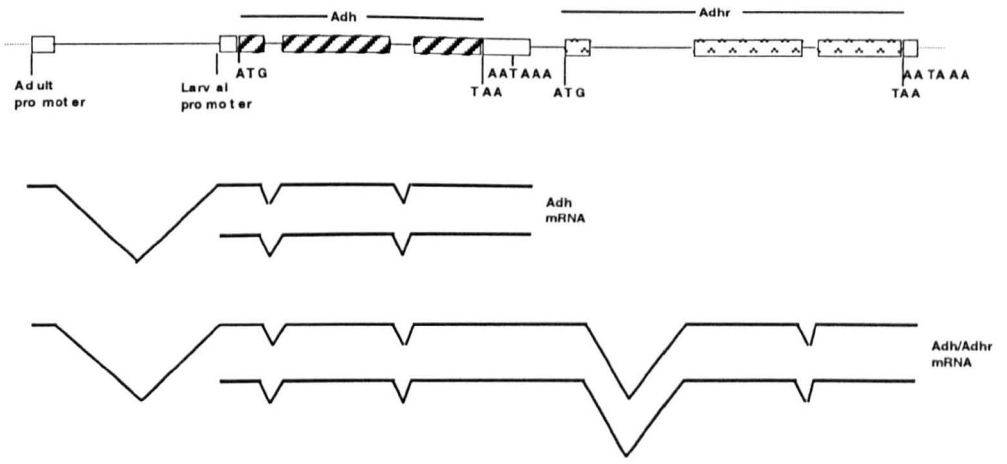
The *Adh* gene of Drosophila encodes for the alcohol dehydrogenase enzyme (ADH, NAD+ oxidoreductase, EC 1.1.1.1). In fruit flies ADH is thought to have evolved to detoxify the ethanol produced by the fermenting fruits on which most of these flies feed and breed (Chambers 1988; Ashburner 1998). In flies, the gene encoding for ADH was first cloned and characterised in *D. melanogaster* where there is a single *Adh* gene with two promoters (Benyajati et al., 1983; Goldberg, 1980). Since its first characterisation in *D. melanogaster*, the *Adh* gene has been cloned in more that 40 species of fruit flies; *Adh* probably represents, from a molecular evolutionary perspective, the most studied gene of Drosophila.

The *Adh* gene is found closely linked to *Adh-related* (*Adhr*) in several species of Drosophila. The two genes are separated by an average 300 bp and are arranged in tandem; most likely they originated from an ancient gene

duplication (Schaeffer and Aquadro 1987). The *Adh* gene in several species is transcribed into two mRNAs from two promoters that are developmentally regulated (Benyajati et al. 1983).

Recently, however it has been reported that *Adh* and *Adhr* constitute a single transcription unit, and from transcripts initiated from the either of the two *Adh* promoters are also generated a class of dicistronic mRNAs encoding for both *Adh* and *Adhr* (Brogna and Ashburner 1997), figure 2. ADHR appears to be exclusively expressed from these dicistronic mRNAs.

Figure 2



Schematic diagram of the *Adh-Adhr* locus. Rectangles represent exons, and lines represent introns. The coding regions of *Adh* and of *Adhr* are indicated by the line above. Below are shown diagrams of the monocistronic *Adh* mRNAs and of the dicistronic *Adh-Adhr* mRNAs generated from either the distal (adult) or proximal (larval) promoters.

The level of the dicistronic mRNA is approximately 100 times lower than the monocistronic *Adh* mRNA, and it may depend upon the choice of polyadenylation sites used, either that immediately downstream of *Adh* or that downstream of *Adhr*. It has been reported that the relative level of monocistronic (*Adh*) and dicistronic (*Adh/Adhr*) mRNA is affected by mutation in *suppressor of forked* [*su(f)*] (Brogna and Ashburner 1997), a gene

that is involved in pre-mRNA 3' end processing (Mitchelson et al. 1993; Takagaki and Manley 1994).

During the characterization of the *Adhr* gene (Brogna and Ashburner 1997), it was also noted that although a nonsense mutation in the *Adh* coding region caused a severe reduction in the level of the monocistronic mRNA , the same mutation seemed not to affect the level of the *Adh/Adhr* dicistronic transcript (figure 3, more information in (Brogna and Ashburner 1997). This was an unexpected result based of the current models of NMMR in yeast and in other systems. Because the current model is, in summary, that a cycle of translation initiation and termination upstream of an appropriate cis-acting element (a DSE in yeast and an RNP tag in mammalian cells) is necessary and sufficient to trigger NMMR in most transcripts, and, importantly an ORF downstream of the DSE elements can not prevent NMMR (Czaplinski et al. 1999).

In order to clarify this issue, I have studied the effect of nonsense mutations on both the *Adh* monocistronic transcript and the dicistronic *Adh/Adhr* transcript.

Chapter 2

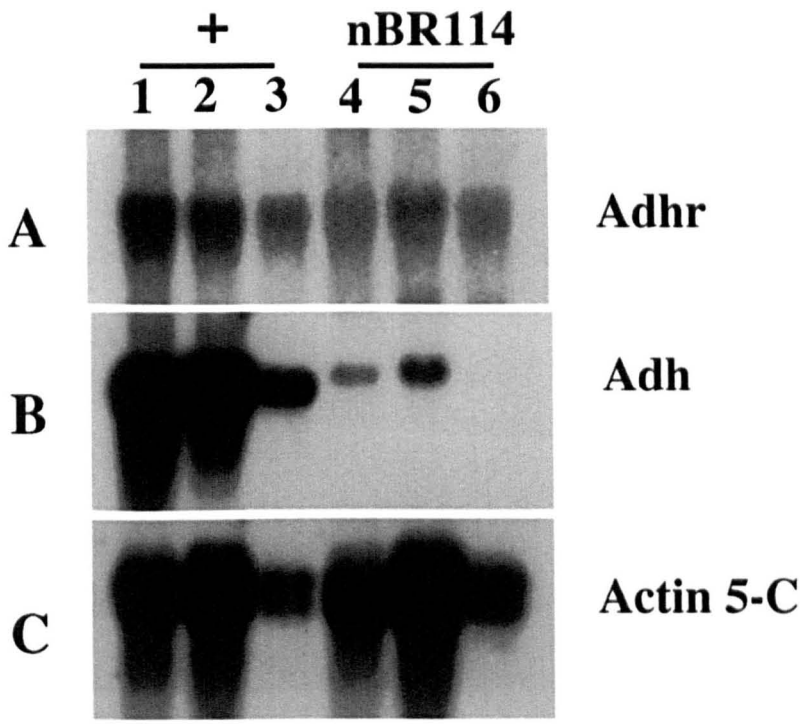
RESULTS - part one

Nonsense mutations affect mRNA metabolism.

The data of this thesis are presented into two separate chapters. The first one is a description of the data regarding the characterisation of the Adh *and* Adhr *mRNA and pre-mRNA in mutants and wild type flies, while the next one will be about the cloning and characterisation of the fly homolog of* Upf1. *The two chapters are divided into separate sections each presenting different steps of the study.*

A number of *Adh* null alleles have been molecularly characterised. These data indicate that in some nonsense alleles the level of *Adh* mRNA is drastically reduced (Chia et al. 1987). Since nonsense mediated mRNA reduction is a general phenomenon (see introduction), this property of *Adh* null mutations received no special attention. However in the course of previous work, we observed that the *Adh* mRNA of the nonsense allele *nBR114* not only is reduced in abundance but also has a lower electrophoretical mobility than that of the wild-type (Figure 3, compare lanes 2-3 to lanes 4-5 -

Figure 3



(**A-C**) Northern blot analysis of RNA. RNAs from wild-type flies (lanes 1-3) and RNAs from *Adh nBR114* flies (lanes 4-6). *Adh nBR114* carries a nonsese mutation at codon 64. Total RNA extracted from adult flies (lanes1 and 4), RNA extracted from the all polysomal fractions (before sucrose gradient sedimentation, lanes 2 and 5) and RNA extracted from polysomes associated with more than two ribosomes (see Brogna and Ashburner, 1997). Note that the amount of RNA in lane 3 is about 3 times less than in lanes 1 and 2, and the amount of RNA in lane 6 is about two-thirds of lane 4 (the relative amounts were estimated by comparing the relative intensity of the Actin signals with the NIH Image program, from autoradiograms. The Adhr and Adh probes and the Actin probe are described in Brogna and Ashburner, 1997. Two similar gels equally loaded were transferred, and one was probed with Adhr and the other first probed with Adh and then stripped and reprobed with Actin-5C . The filter in figure 3A was exposed 5 days at -70 ^{0}C and those in figure B and 3C approximately five hours at -70 ^{0}C.

more information in (Brogna and Ashburner 1997). This preliminary observation raised the suspicion that in the *nBR114* mutant and perhaps in others, the mRNA processing of the *Adh* transcript may be affected.

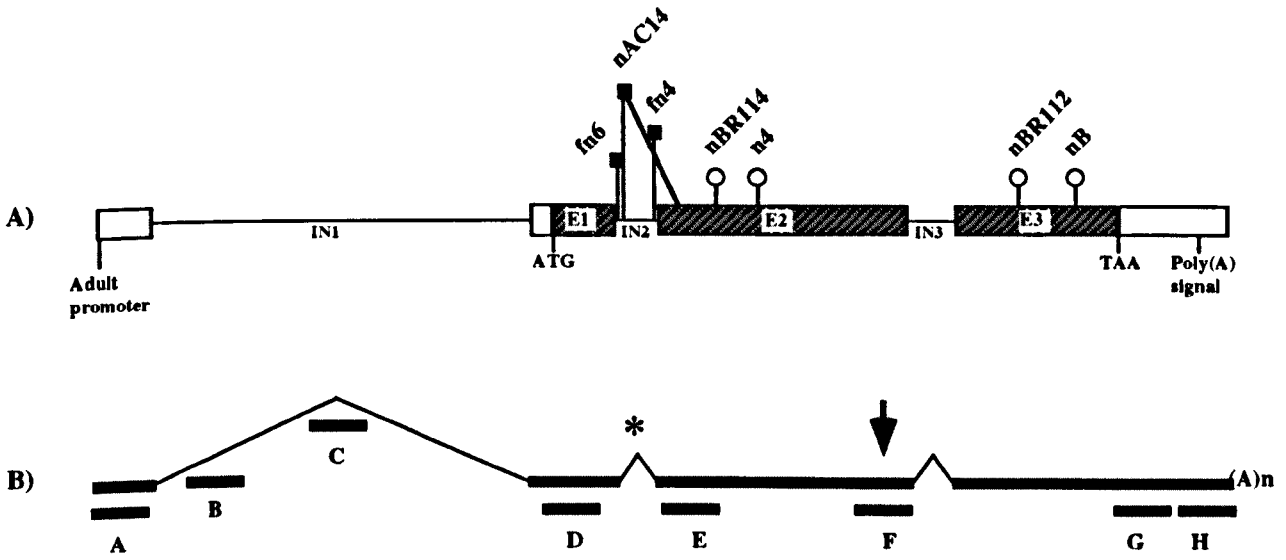
Because of this unexpected result, and since the phenomenon of nonsense mediated mRNA down regulation has not so far been studied in Drosophila to any significant extent, I decided to further investigate the effect that nonsense mutations have on the *Adh* transcript. Mutations that prevent splicing of one of the *Adh* introns are also known to lead to reduced levels of mRNA and have been included in this study.

The steady-state level of mRNA is reduced in *Adh* mutations that produce either a premature stop codon or prevent splicing.

In adult flies, *Adh* is transcribed from the distal promoter, producing a pre-mRNA with three introns (Figure 4). The first intron (IN1) is located in the 5' untranslated region and the other two (IN2 and IN3) are in the coding region. Intron 2 lies between codons 33 and 34 and intron 3 between codons 136 and 137, the complete coding sequence having 257 codons.

Here, I analysed the level of accumulation of the *Adh* mRNA in several *Adh* alleles that have previously been sequenced and partially characterised (figure 4A shows a diagram of the *Adh* gene and the location of the mutations; the name of the alleles refer to the different mutagenesis screens in which the

Figure 4



A schematic map of the Adh gene and a diagram of the structure of the distal transcript. (A) Genomic region. Rectangles represent exons. Hatched boxes represent coding regions, labelled E1 to E3, white boxes represent untranslated regions. The horizontal lines represent introns, IN1 to IN3. The positions of the nonsense mutations are indicated with vertical lines marked with an open circle, the mutations affecting splicing are marked with lines ending in filled squares. The accession number for the sequence of this region is X78384. (B) Pre-mRNA transcribed from the distal (adult) promoter. Thicker lines represent exons, thinner lines introns. Black arrows indicate the approximate positions of the oligodeoxynucleotides used. The vertical arrow indicates the position of cleavage of RNase H + oligo F. The asterisk indicates the intron whose splicing is prevented by the mutations indicated above. The sequence of the oligos, 5' to 3', are:

A - ctg ga(c/a) ctg (g/a)ca act (a/g)(t/c)a ctg aga ca,
B- ttg ggt gcc ctt ttg cta ct, C- cac ggt ttg ttt ttg ttt gg,
D- acg tt(c/a) ttg ttg gtc a(a/g)(a/t) gt(a/g) (g/a)(t/a)c gac at,
E - gat tga ttg cct tca gct cgg c, F - cca ggg agc tgg tga agt tga c,
G - cat cta aga agt gat aat cc, H - cac act cac att ctt ctc cta a.

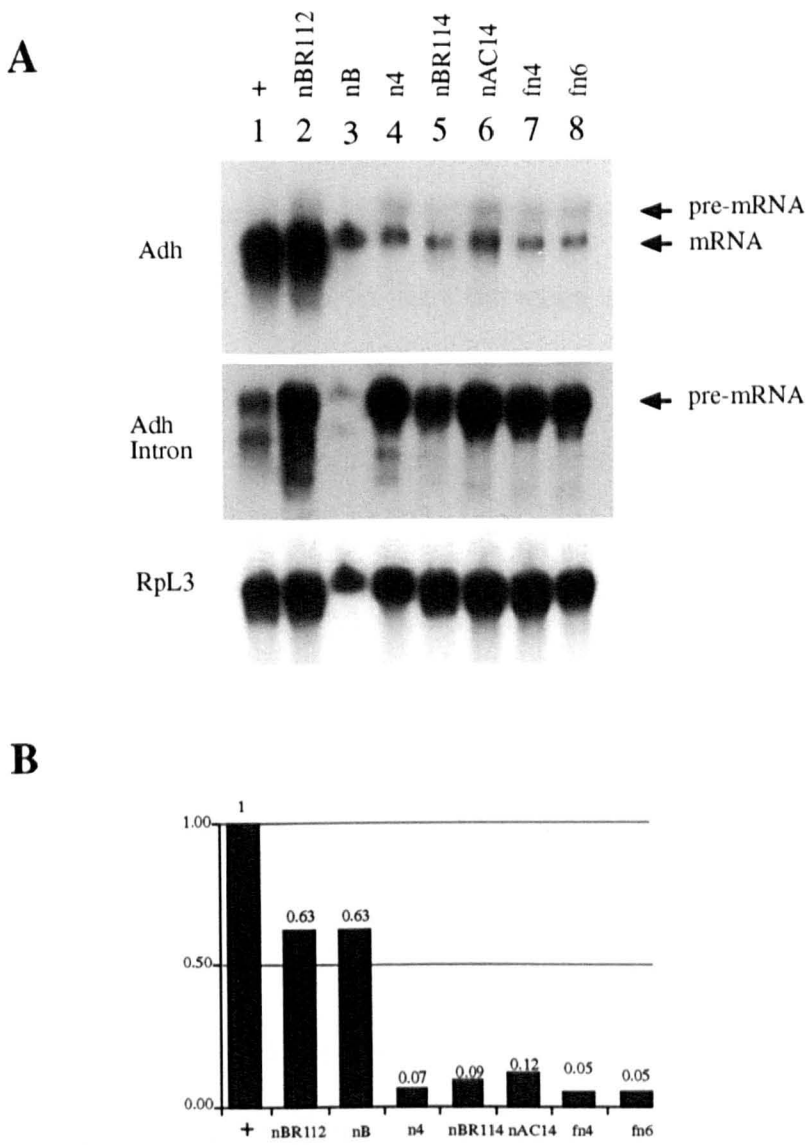
mutants were isolated, additional information and references are given in "Materials and Methods").

Four of these mutations are single substitutions that generate premature termination codons in the Adh coding region: *Adh* ^{nBR114} (Tyr 64, TAT --> TAA), *Adh* ^{n4} (Gln83, CAG-->TAG), *Adh* ^{nBR112} (Gln 203, CAG -->TAG) and *Adh* ^{nB} (Trp 236, TGG-->TGA). Two other mutants, *Adh* ^{fn6} and *Adh* ^{fn4}, are small deletions that destroy respectively the 5' and 3' splicing signals of intron 2. In *fn6* and *fn4*, splicing of this intron is therefore prevented. In both mutants, the retention of the intron leads to in-frame premature stop codons. *Adh* ^{nAC14} is a naturally occurring allele which has an insertion in the same intron, that inhibits splicing. This allele also has a second mutation that generates a stop codon in exon 3 (Glu 49, GAG--> TAG). The mutations described are the only known differences to the progenitor wild-type allele.

In order to quantify the relative levels of *Adh* RNA, total RNA was extracted from the mutants and from wild-type, and the *Adh* RNA was visualised by Northern Blot hybridisation (Figure 5A). The RNA levels in the mutants were quantified relative to the wild-type allele. Figure 5B shows the results of this quantification.

The data show that, relative to the wild-type allele, all mutants have more or less reduced levels of *Adh* mRNA. The most affected alleles are *n4*, *nBR114*, *nAC14*, *fn4* and *fn6*, where the levels of *Adh* RNA are between 5%

Figure 5



Adh mRNA levels in nonsense and splicing defective mutants. (A) Northern blot analysis of total RNAs isolated from wild-type (lane 1) and mutant flies (lanes 2-8). The filter was first hybridised with two Adh specific probes and then stripped and re-hybridised with a cDNA clone for the ribosomal protein L3 (RpL3, Chan et al., 1998). The first *Adh* probe was amplified from a genomic clone by PCR with primers at positions 2299 and 2926 relative to X78384 and is specific for the last two exons. The second *Adh* probe is specific for one of the introns (IN1), see figure 7. All alleles have been crossed to a stock carrying the *Df(2L)A63* deficiency, and the RNA was purified from hemizygous flies selected from the F1 offspring. Note that lane 3 is under loaded relative to the others, which contain 5-10 µg of total RNA. (B) Diagram of *Adh* mRNA levels relative to the wild-type control. The amount of radioactivity in each band was measured directly from the filter with an InstantImager (Packard Instruments Co.). The net total number of radioactive counts corresponding to each Adh band was standardised to the RpL3 signal and divided by the *Adh* counts in the wild-type lane. Similar values were obtained in an independent replica experiment in which Actin-5C was the loading control (not shown).

and 12% of the wild-type. The least affected are the two alleles *nBR112* and *nB*, which map closer to the normal stop codon, where the RNA levels are about 65% of the wild-type. Also note that *nBR112* and *nB* are located in the last exon (see discussion).

The level of *Adh* pre-mRNA, on the other hand, appears to be increased in the mutants (figure 5A, see also figure 11). However, the extent of this increase in *Adh* pre-mRNA seems to vary between experiments (see figure 6, for a comparison). The reasons for this variation are addressed below.

In conclusion, these data indicate that nonsense mutations in the coding region or the failure to excise intron 2 lead to a reduced level of the corresponding mRNA and usually to an increase in the level of pre-mRNA. The closer the premature stop codon is to the natural termination codon, the less severe is the mRNA reduction.

The level of nuclear mRNA is also reduced.

As mentioned in the introduction, the mRNA reduction associated with nonsense mutations in mammalian systems often happens while the mRNA is still associated with nuclei. I therefore checked the level of the nucleus-associated *Adh* RNA of the nonsense mutants.

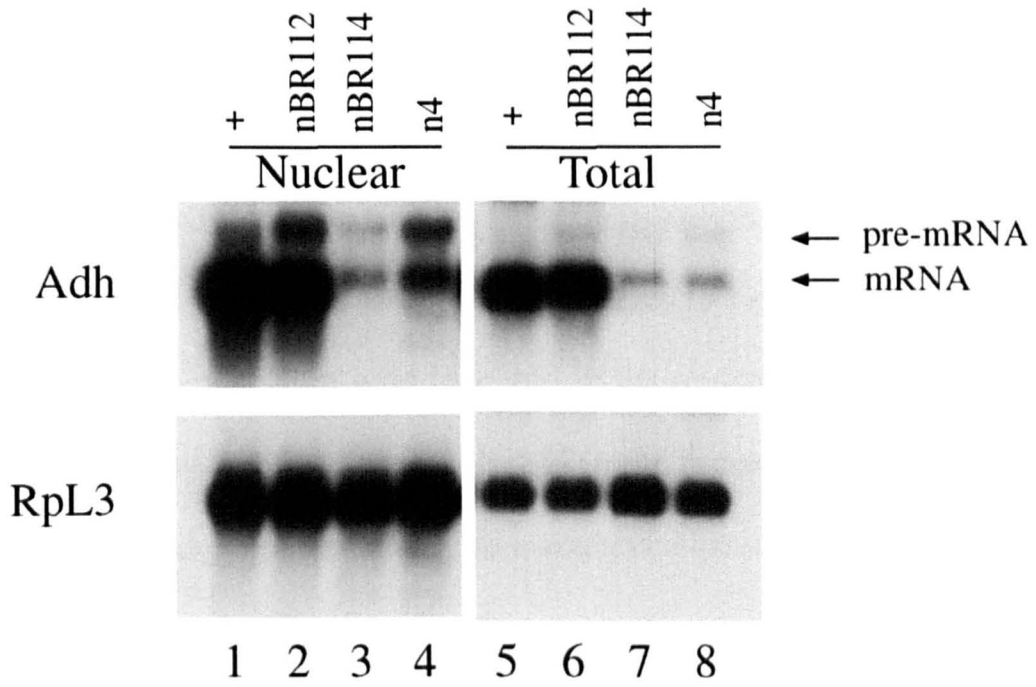
Nuclear cell fractions were prepared from adult flies (Materials and Methods). Adult flies were homogenised in two steps in an isotonic buffer,

and the ionic detergent deoxycholate was added in the second step. Several studies have shown that deoxycholate has the ability to strip off the outer layer of the nuclear envelope and surrounding cytoplasm (Belgrader et al. 1994). The nuclei were purified by centrifugation through a hypertonic sucrose cushion. Nuclei prepared with this method looked, by phase-contrast microscopy, to be clear of cytoplasmic contaminations or incompletely lysed cells.

Nuclei were prepared from *nBR112*, *nBR114*, *n4* and from the wild-type strain. Figure 6 shows that relative to the wild-type (lane1), the level of *Adh* mRNA is drastically reduced in *nBR114* (lane3) and *n4* (lane 4) and to a much lesser extent in *nBR112* (lane2). The relative mRNA reduction observed in the nuclei is similar to that observed in total-RNA preparations (figure 6, compare lanes 1-4 with lanes 5-8). Note that in the nuclear RNA the ratio between the level of pre-mRNA and mRNA is increased, as expected from a nuclear RNA fraction. It should also be noted that the mRNA of *Adh nBR114* and *Adh n4* appears to have a lower electrophoretic mobility compared to *Adh nBR112* and wild-type. This is true for both the nuclear and the total RNA preparations.

In summary, the data indicate that *Adh* nonsense mutations reduce the nuclear level of the corresponding mRNA. The overall reduction of *Adh* mRNA is probably the direct consequence of down-regulation of the steady-state level of the nuclear *Adh* mRNA.

Figure 6



Northern blot analysis of nuclear RNA and total RNA in nonsense mutants. The RNA was extracted either from nuclear fractions purified from adults (lanes 1-4) or from whole adult flies (lanes 5-8). About 2 mg of RNA was loaded in lanes 1-4 and about 10 mg in lanes 5-6. The filters were first probed with an Adh probe, then stripped and hybridised with the RpL3 probe (see figure 2). The position of the mRNA and pre-mRNA bands are indicated. Note that in the nuclear RNA, the ratio between pre-mRNA and mRNA is increased. The filter was also hybridised with a probe specific for the pre-mRNA (not shown, refer to figure 11).

mRNAs carrying premature termination codons have longer poly(A) tails as assayed by Northern blot analysis.

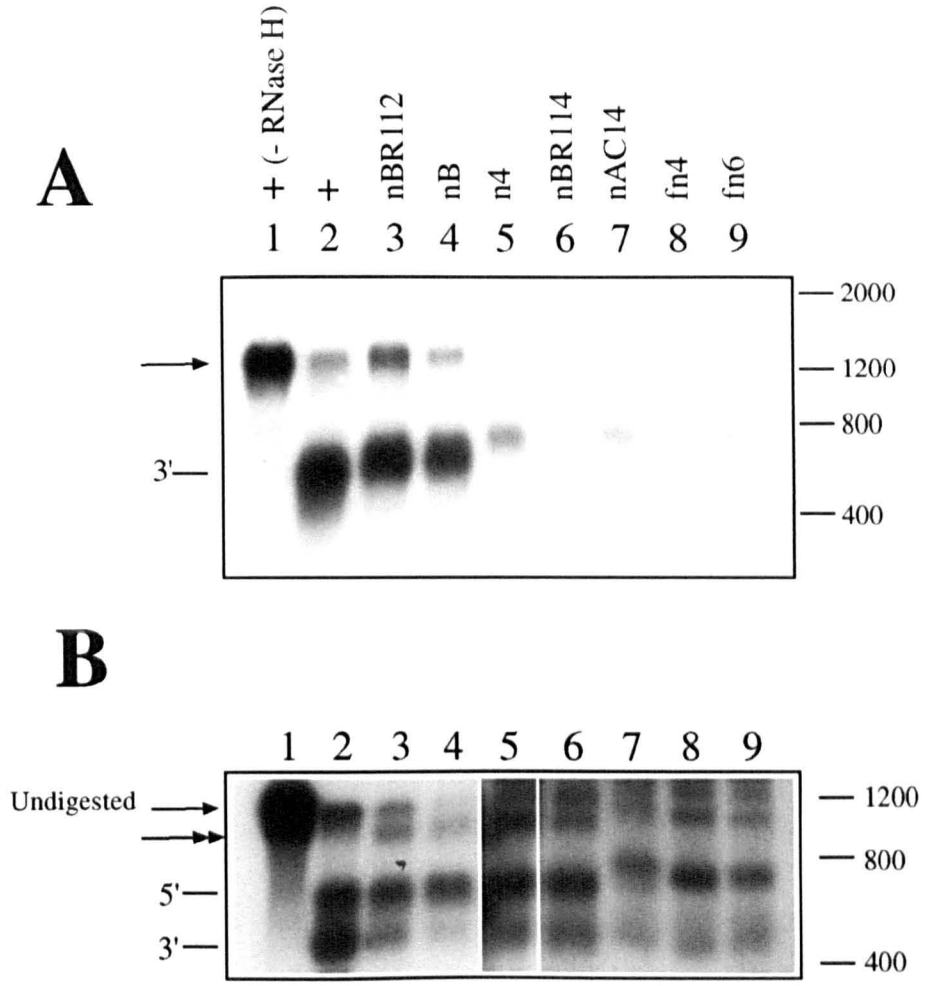
As mentioned before, Northern blot analysis indicates that in some *Adh* mutants which carry a nonsense mutation, the *Adh* mRNA may be longer than expected (Figure 3). At least for one of these alleles, *Adh n4*, S1 nuclease protection studies showed a normal splicing pattern (Chia, et al., 1987). Therefore, one possible reason for this increase in length could be the presence of a longer poly(A) tail. Here, I report experiments to better characterise the poly(A) tail of *Adh* mRNA in wild type and in the mutants described above.

Total RNA was cleaved with RNase H in the presence of an oligodeoxynucleotide complementary to a sequence located in the 3' half of the *Adh* mRNA (Figure 4B, oligo F). As a control, a parallel experiment was performed with the addition of oligo-$d(T)_{16}$ in order to remove the poly(A) tail. After cleavage, the total RNA was fractionated on a high-resolution agarose gel and the *Adh* mRNA was analysed by Northern blot (see Materials and Methods). This digestion splits the *Adh* mRNA in two moieties: a 5' fragment of 603 nucleotides (nt) and a smaller 3' fragment of 443 nt (excluding the poly(A) tail).

When the filter was hybridised with a probe specific for the 3' fragment (Figure 7A), it was apparent that relative to the wild-type *Adh* mRNA, the

mutants have increasingly longer 3' ends (compare lane 2 with lanes 3 to 9). I estimate that relative to the wild-type (lane 2), the 3' fragment in *Adh nBR112* (lane 3) and *Adh nB*(lane 4) is about 20 nucleotides longer, and over 100 nucleotides longer in the other mutants (lane 5-6). If the increase in size of the 3' end is indeed due to a longer poly(A) tail, deadenylating the RNA should lead to the production of 3' fragments of similar length by RNase H cleavage of all mRNAs.

Figure 7



Poly(A) length variation in different alleles: RNase H digestions and Northern blot analysis. (**A**) Northern blot analysis of approximately equal amounts of total RNA (~5 µg) that have been digested with RNase H and an oligo-deoxyribonucletide complementary to the Adh mRNA (oligo F, see figure 1B). Lane 1 contains RNA that has not been digested with RNase H. The filter was hybridised with an Adh probe specific for the 3' end moiety generated by RNase H digestion. The probe was PCR amplified from a genomic clone with primers at positions 2724 and 3123 relative to X78384 (see figure 4A). The arrow indicates the position of the undigested Adh mRNA. The RNA was fractionated on 3% NuSieve 3:1 agarose gels in the presence of formaldehyde. The marker is a DNA size standard (Gibco-BRL) that has been denatured in 98% formamide for 5 minutes at 95 ^{0}C.

(**B**) Northern blot of the Adh mRNA that has been digested with RNase H and two oligo-deoxyribonucleotides, one is the same as in fig. 4A, the other is a 16 long oligo d(T). The probe is as in figure 2A and detects both the 5' and 3' fragments. Here, the poly(A) tail of the mRNA is removed. Note that in lanes 7-9 the 5' end fragment migrates slightly slower that in the other lanes (lane labelling is as in A). All lanes are from the same filter that was exposed for different lengths of time in order to visualise all the bands. The single arrow indicates the position of the undigested Adh mRNA. The double arrow indicates the position of the fragment generated by oligo-d(T) mediated cleavage of a region in the Adh 3' UTR that has a stretch of 11 As. Note that the extent of this cleavage is variable between lanes. This cleavage in the 3' UTR causes the reduction of intensity of the 3' end fragments. It should also be noted that in this experiment the electrophoresis was slightly distorted, therefore a small smear is noticeable between lanes.

Figure 7B shows that cleavage of deadenylated *Adh* mRNAs with RNase H, in the presence of oligo F and oligo-d(T), produces two fragments of about 600 and 450 nucleotides. No significant size differences are noticeable between the 3' fragment of wild-type *Adh* mRNA (lane2) and that of the mutants (lanes 2-9). The significance of the slight variation in size between some of the 3' end fragments is not clear. Note that the 3' UTR contains a stretch of 11 As, and there is some oligo-d(T) mediated cleavage in this sequence (this causes reduction of the signal from the 3' fragment). Interestingly, the extent of this cleavage is higher in the mutants than in wild-type (see legend of figure 6B and Materials and Methods)

As expected, due to the retention of the second intron (65 nucleotides long), the 5' fragment of *Adh* ^{nAC14} (lane 7), *Adh* ^{fn6}(lane 8) and *Adh* ^{fn4} (lane 9) is accordingly longer. It should be noted that in the natural allele *Adh* ^{nAC14}, an additional 5' fragment is accumulated. Previous work carried out with this mutant suggests that this fragment is produced by aberrant splicing of the first intron (Freeth et al. 1990).

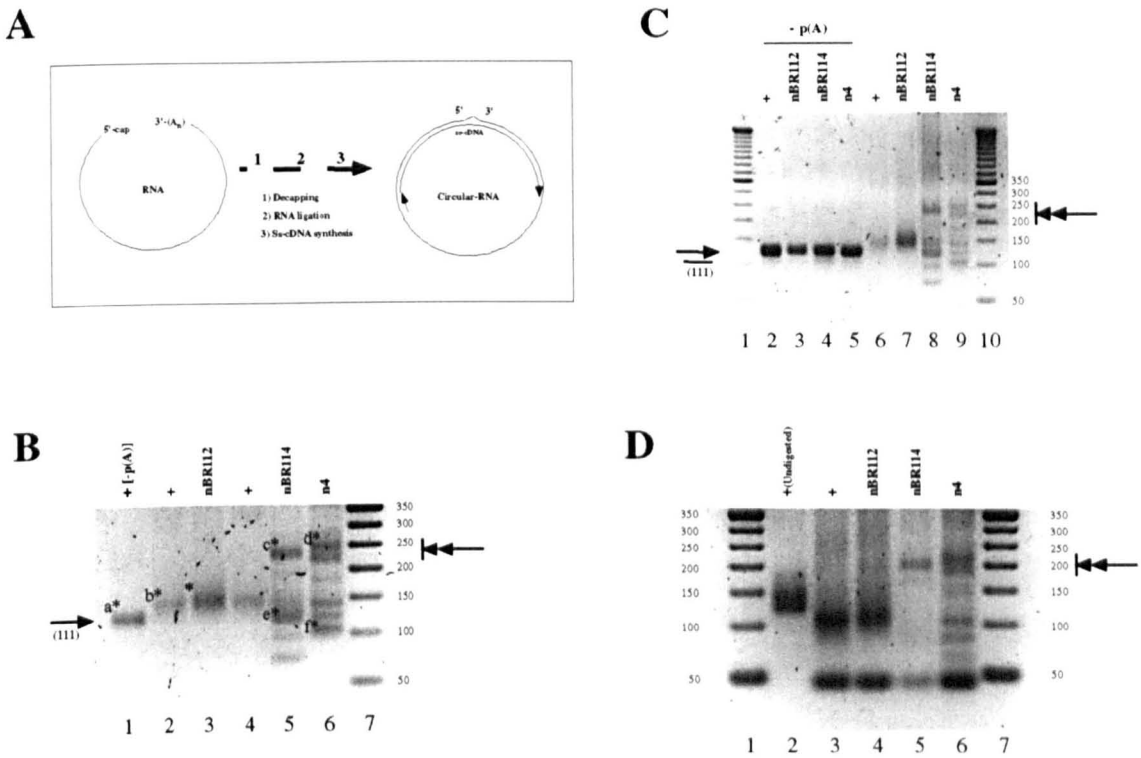
The conclusion from these experiments is that nonsense and splicing mutations in *Adh* lead to mRNAs with a longer poly(A) tail. The closer the premature stop codon is to the translation initiation codon, the longer is the poly(A) tail. Similar conclusions were also reached with a different experimental approach (see below).

Transcripts carrying premature termination codons have longer poly(A) tails as assayed by circular-RT-PCR.

In order to gain more detailed information on the effect of nonsense mutations on the length of the poly(A) tail, I followed an alternative experimental approach that makes use of the sensitivity of PCR. The basic principle of this technique is the circularisation of in vitro decapped RNA followed by cDNA synthesis across the junction and subsequent PCR with primers hybridising to the flanking sequences (Couttet et al. 1997); a schematic diagram of this technique is shown in figure 8A). This technique can be applied in order to precisely determine the length of the poly(A) tails and to identify other potential changes in 5' and 3' ends of the *Adh* transcripts (see Materials and Methods).

Circularised RNA was prepared and cDNA synthesised using a primer complementary to the beginning of exon 2 (Figure 4, primer E). Two rounds of PCR were performed, the first with primers G and D and the second with the two nested primers A and H (Figure 4B shows the position of the primers). Two parallel experiments were performed. In one of them, the RNA was first deadenylated using RNase H and oligo-d(T). Figure 8B shows that without deadenylation, longer fragments are produced from the cDNA of the mutants (lanes 3, 5 and 6 compared to 2 and 4). In particular, in *Adh* ^{nBR114} and *Adh* ^{n4}

Figure 8



Poly(A) length variation between different nonsense alleles: mRNA cRT-PCR.
(**A**) A schematic diagram of the circular RT-PCR technique. Decapping is done with Tobacco Acid Pyrophosphatase. The circularisation is achieved by directly ligating, with T4 RNA ligase, the 5' and 3' ends together. The ligation is done at low concentration of RNA, in order to prevent inter-molecular events (see Materials and Methods).
(**B**) Gel electrophoresis of cRT-PCR products. About 10 mg of purified total RNA were first decapped and recircularised. Then, the cDNA was synthesised across the junction. (Materials and Methods). The first round of PCR was with primers G and D and the second with primers H and A (Figure 4B). Lanes 1, 2 and 4 are wild-type and lanes 3, 5-6 are mutant RNAs. Note that in lane 1, the RNA was first deadenylated (see text). The asterisk indicates the fragments that have been sequenced (Materials and Methods).The single arrow indicates the fragment generated from deadenylated RNA, the expected size of a completely deadenylated template is indicated in brackets (see below). The double arrow and the vertical line indicates the position of the longer fragments produced in the mutants. The size marker (lane 7) is a 50 bp ladder (Gibco-BRL). The PCR fragments were separated in 3% Nusieve 3:1 agarose gel. Note that this image is a photographic negative.
(**C**) Similar experiment as in A. Lanes 2-5 are circular-RT-PCR of RNA from wild-type (lanes 2 and 6) and mutants (lanes 3-5 and 6-9). The RNAs used in lanes 2-5 were first deadenylated (see text). The line indicates the size (111 bp) of the fragment expected from a completely deadenylated mRNA. Note that lanes 6-9 show a replica of the experiment shown in A, the same cDNA was used as template.
(**D**) Restriction digestion analysis of c-RT-PCR products like those shown in B. The whole PCR reaction was purified with Qiaquick (Qiagen) columns and digested with Nla III. Gel electrophoresis conditions are as above. Lanes 1 and 7 are molecular weight markers (as above). Lane 2 is undigested c-RT-PCR products from wild-type, lanes 3-6 are c-RT-PCR products that have been digested.

it is apparent that products of up to 100 nucleotides longer than the wild type product are amplified (indicated by a double arrow). In the mutant *Adh* ^{BRII2}, the length of the fragments is also increased but to a much smaller extent (this is more obvious in figure 8C, compare lane 6 with lane 7). Here, the fragments are about 10-20 nucleotides longer.

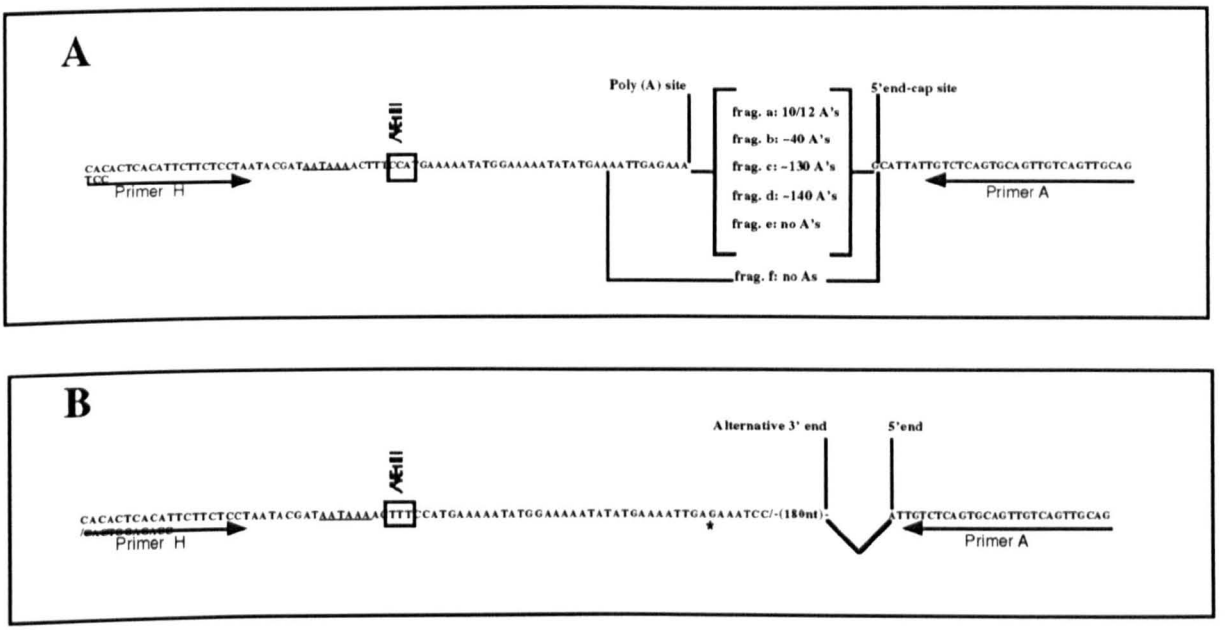
In the control c-RT-PCR experiment, the RNA was deadenylated before circularisation. Figure 8C (lanes 2-5) shows that here, fragments of equal size are produced in both wild-type and in the mutants. These data, therefore, support the hypothesis that the longer 3' end of the transcripts of the mutants is due to a longer poly(A) tail.

The nature of these c-RT-PCR fragments was checked by restriction analysis. The restriction endonuclease *Nla* III cuts 40 bp downstream of the annealing site of primer H, before the polyadenylation site, and it is therefore expected that digestion of the fragments in figure 8B should generate a diagnostic 40 bp fragment constant in all alleles and a variable longer fragment containing the poly(A) tail. Figure 8D shows that indeed in all alleles (lanes 3-6) digestion with *Nla* III generates a constant fragment of about 40 bp and longer fragments variable between alleles. The longer fragments migrate as bands about 40 bp shorter than those in figure 8B, as expected from bona-fide c-RT-PCR products.

In order to fully characterise these c-RT-PCR products, the fragments indicated with an asterisk in figure 8B were sequenced (figure 9A shows the sequences). As predicted, these fragments did indeed originate from circular transcripts produced by joining the 5' and 3' ends. In all cases, the 5' end coincides with the capping site (position +1) and in most cases (fragments a-e) the canonical cleavage/polyadenylation site was used. The smaller fragments produced in the mutants *Adh* ^{nBR114} and *Adh* ^{n4} (labeled e and f in figure 8B) derived from transcripts without any poly(A) tail. Note that the smallest fragment in *Adh* ^{n4} (lane 6, fragment f) also has a 10 nt shorter 3' end. The fact that these products are not visible by Northern blot analysis (Figure 7) indicates that they constitute a very small fraction of the total *Adh* RNA. One possibility is that they may represent degradation intermediates accumulating in the mutants. Interestingly, most of these smaller products, but not the longer ones, were readily amplified in c-RT-PCR experiments in which the decapping step was omitted (data not shown). It should also be noted that the pattern of this group of slow migrating bands varies between experiments (data not shown). Such smaller products are expected to be preferentially amplified by PCR due to their small size and probably more importantly, their lack of a poly(A) tail.

In summary, these data show that the poly(A) tail of the mRNA is about 40 nucleotides long in the wild-type, but up to more then 100 nt longer in

Figure 9



Sequence of the c-RT-PCR products. (A) The sequence of the six fragments labelled a through f in figure 5B is shown. Between the fragments a through e the only difference is the length of the A:T stretch, this is indicated in the square brackets. Fragment f has a 10 nuclotides shorter 3' end and no A:T stretch.
The sequence was obtained from sequencing two independent plasmid clones of each fragment (see Material and methods). In fragment a, the number of A:T in the two clones sequenced was 10 and 12. The length of the A:T stretch of the fragments b through d was estimated from the size of the cloned fragments. The small box highlights the recognition sequence of Nla III (B) Sequence of the larger fragment shown in figure 7A (lane 5, fragment 1). Two independent clones were sequenced. The asterisk indicates that 180 bp just downstream from the wild-type polyadenylation site are not shown, the alternative 3' end is located 191 nucleotides downstream from the normal one. Note that two other Nla III sites are located in the 180 bp not shown, 83bp and 140 bp from the one shown. For more information refer to the sequence of the genomic region (x78384 in GenEMBL).

nonsense mutants. The degree of extension of the poly(A) tail correlates inversely with the distance of the premature termination codon from the AUG.

Premature termination codons also affect the polyadenylation of the pre-mRNA.

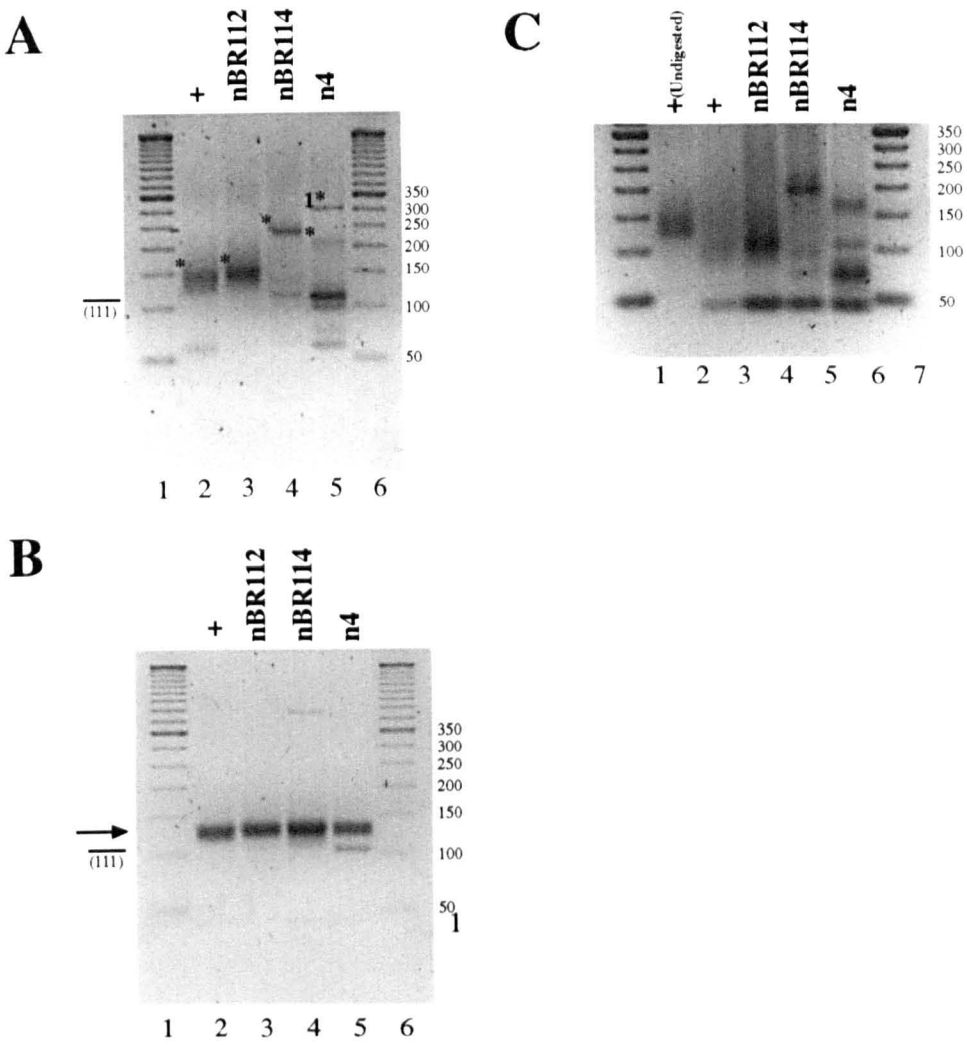
Next, I addressed the question of at which stage the mutant mRNAs acquire their longer poly(A) tails. In the vast majority of cases, the maximal length of the poly(A) tail is determined by the extent of nuclear polyadenylation. However, there are examples in which the poly(A) tail is elongated in the cytoplasm. This is often associated with the activation of translationally repressed transcripts (Wickens 1992). The fact that by Northern blot analysis I could not detect any difference between the mobility of nuclear and cytoplasmic mRNA (see above) of any alleles, strongly suggests that the longer poly(A) is added in the nucleus.

To test whether the pre-mRNA already has the longer poly(A) tail, the circular RT-PCR approach was used. In order to specifically amplify the pre-mRNA, one of the primers used was complementary to sequences located in the intron (IN1). In this experiment, the first round of PCR was with primer G and primer C (see Figure 4B), and the second with primer H and A. This first round of PCR with one intron-specific primer (C) generates a vast excess of unspliced specific template for the second round of PCR. This experiment

shows (figure 10A) that in the wild type, a band of 140-150 bp is amplified, while in the mutants longer fragments are seen (lanes 3-5). In particular, fragments up to 100-150 nt longer were apparent in the two more "5' terminal" mutants, *nBR114* and *n4* (lane 4-5). The maximal length of these longer products in *nBR114* and *n4*, varied slightly between similar experiments. *nBR112*, the "3' terminal" mutant, appears also to generate a slightly longer fragment (10-20 bp longer). The nature of these c-RT-PCR fragments was again checked by digestion with *Nla* III (figure 10C). In the control experiment (figure 10B), in which deadenylated RNA was used, a fragment of 130 bp was generated in all cases. These data are consistent with the hypothesis that the longer poly(A) tails of the mutant mRNAs are a direct consequence of the fact that the poly(A) tails of the corresponding pre-mRNAs are longer. These circular-RT-PCR data are also supported by Northern blot analysis which shows that the pre-mRNA of the mutants is apparently longer than wild-type (figure 11), this is also apparent in figure 6 (also see discussion).

It should be noted that the variation in the relative levels of pre-mRNA does not correlate with the length of the poly(A) tail. For example, although *nBR114* shows large variations in the level of pre-mRNA relative to the wild-type (compare the levels of pre-mRNA in lane 1 and 5 of figure 5 to that in

Figure 10



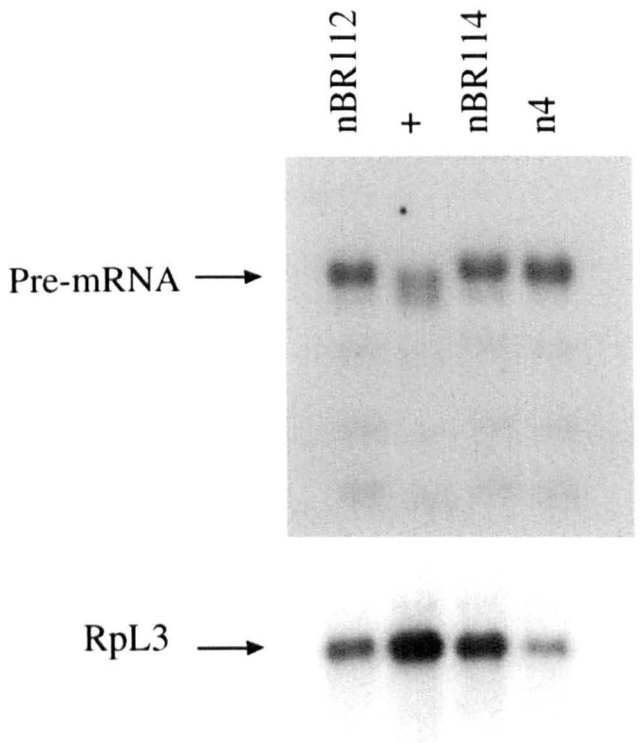
Nonsense mutants affect the 3' end processing of the Adh pre-mRNA: pre-mRNA specific cRT-PCR.

(**A**) Gel electrophoresis of cRT-PCR products. RNA decapping, ligation and cDNA synthesis are as described in figure 5. The first round of 30 cycles of PCR was with primers G and C and the second 30 cycles with primers H and A (Figure 4B). The size marker (lanes 1 and 6) is a 50 bp ladder (Gibco-BRL). The line indicates the position of the fragment that should correspond to a complete deadenylated product (111 bp). The asterisks indicates the fragments that have been sequenced. The fragment 1 in lane 5 has unusual ends. The 5' end is 5 nuclotides shorter and the 3' end is 191 nuclotides downstream from the canonical poly(A) site. The PCR fragments were separated on a 3% Nusieve 3:1 agarose gel. Note that the image is a photographic negative.

(**B**) Similar experiment to the one described in (A), but the RNA has been deadenylated with oligo-d(T) and RNase H. The single arrow indicates the fragments that have been cloned and sequenced. The sequences showed that they contain a residual poly(A) tail ranging from 15-20 nucleotides. The line indicates the position of a fragment that would be generated from a completely deadenylated product (111 bp).

(**C**) Restriction enzyme digestion analysis of c-RT-PCR products similar to those shown in A. Conditions are as in figure 5D. Lanes 1 and 7 are molecular weight markers, as above. Lane 2 contains undigested PCR products from wild-type, lanes 3-6 are PCR's products that have been digested. Note that the largest fragment shown in figure 7A (lane 5, fragment 1) has a longer 3' end that contains two additional Nla III sites (see legend figure 6B), and after digestion 4 fragments are expected: 103, 83, 68 and 40 bp long.

Figure 11



The Adh pre-mRNA is longer in nonsense mutants. Northern blot analysis of total-RNA (5-10 µg). The RNA was fractionated overnight on a 1.2 % agarose formaldehyde gel; formaldehyde was also added to the buffer. The filter was first hybridised with a probe specific for the first Adh intron, then stripped and probed with a probe specific for RpL3 (see fig. 5). The intron specific probe was PCR amplified from a genomic clone with two primers at positions 1392 and 1777 relative to the sequence X78384 (see fig. 1A). Note that the lane with wild-type RNA contains more RNA than the others.

lanes 1 and 3 in figure 6) the length of the poly(A) tail is about 100 nucleotides longer than the wild-type in all cases tested.

I completed this analysis by sequencing most of the PCR products generated in this last experiment (figure 10A, the bands sequenced are indicated with an asterisk). The sequencing essentially confirms that these are genuine fragments generated from circularised *Adh* transcripts and that in the mutants, the increase in RNA size is indeed due to a longer poly(A) tail. However, I found one interesting exception. In the *n4* reaction, the largest product (Figure 10A, band 1 in lane 5) appears to have been generated from a transcript with a longer 3' end, a truncated 5' end and without a poly(A) tail (sequence shown in figure 9B). At this stage, I suspected that in the mutants transcripts with longer 3' ends may be accumulating. In several similar experiments, it was indeed possible to amplify from all mutants transcripts with longer 3' ends. Since these unspliced (or partially spliced) transcripts have also truncated 5' ends, they might be degradation intermediates. However, at this stage it can not be excluded that these products may also be present in the wild-type, probably at a much reduced level.

In summary, these data show that nonsense mutations in the *Adh* gene in Drosophila are associated with an abnormal 3' end processing that results in mRNAs with longer poly(A) tails.

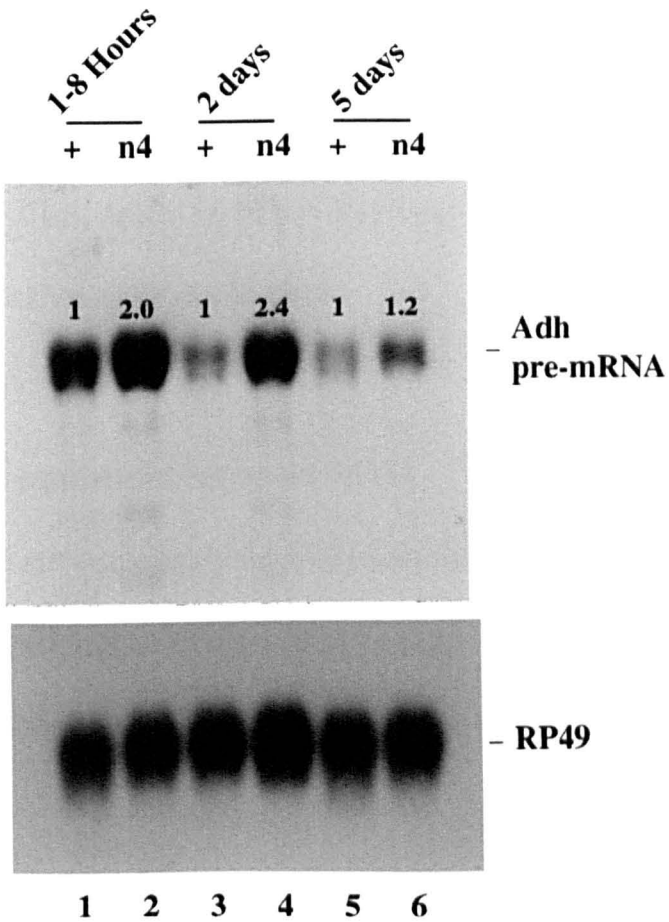
Nonsense mutations increase the level of pre-mRNA in an age dependent manner.

The data shown above suggest that nonsense mutations may also lead to an increase in the level of pre-mRNA (figure 5A). However, as mentioned above, this issue was complicated by the fact that the extent of this increase was apparent in some experiments but not in others. By exclusion, a possible explanation for these fluctuations could have been that the age of the flies used to extract the RNA may have been different between experiments.

In order to test this hypothesis, I extracted RNA from flies of given ages and analysed the level of the unspliced *Adh* transcript by Northern blot analysis. The experiment was done by using the wild-type and Adh^{n4} strains - *n4* being one of the nonsense mutations with a strong mRNA reduction phenotype.

The data indicate that the level of the unspliced *Adh* transcript (pre-mRNA) is increased in the nonsense mutant strain (figure 12). Interestingly these data also show that the ratio between the level of pre-mRNA in the wild-type flies and that in the mutant decreases with the age of the flies. For example in newly emerged flies (lanes 1-2) there is 2 times more pre-mRNA in the mutant strain that there is in the wild-type - the numbers above the pre-mRNA bands refer to the relative level of pre-mRNA in each age group (see figure legend). In contrast in 5 days old flies (lanes 5-6), the difference

Figure 12



Nonsense mutations increase the level of unspliced transcript. Northern blot analysis of total RNA from flies of three different age groups: 1-8 hours old (virgins), 2 days (+/- 8 hours) and 5 days (+/- 8 hours) adult flies. The flies were kept at 25^0 under standard condition, under these conditions wild-type flies start dying after about two weeks and earlier for the Adh mutant. The filter was first hybridised with a probe corresponding to the first Adh intron (panel above); the band corresponding to the unspliced transcript (pre-mRNA) is indicated. For more information on the probe refer to the legend of Figure 5). The same filter was stripped and reprobed with a cDNA for Ribosomal protein 49, as loading control. The number above the pre-mRNA bands are the relative level of radioactivity between wild type and the mutant, the level of radioactivity of the pre-mRNA bands was standardised for the variation in the loading. The level of radioactivity was measured directly from the filter using an Instant Imager (see figure 5).

between the level of pre-mRNA in the mutant and that in the wild-type was minimal (1.2). The highest increase was observed in 2 day old flies (lanes 2-3), where there was 2.4 times more pre-mRNA in the mutant that in the wild-type. It should also be noted that the level of pre-mRNA decreases with the age of the flies regardless of whether or not they carry a nonsense mutation.

In conclusion these data establish that the level of the unspliced *Adh* transcript is increased by a nonsense mutation. The extent of this increase is most apparent in young flies, while in 5 day old flies the pre-mRNA is minimally increased relative to the wild-type.

The effect of nonsense mutations happens in cis.

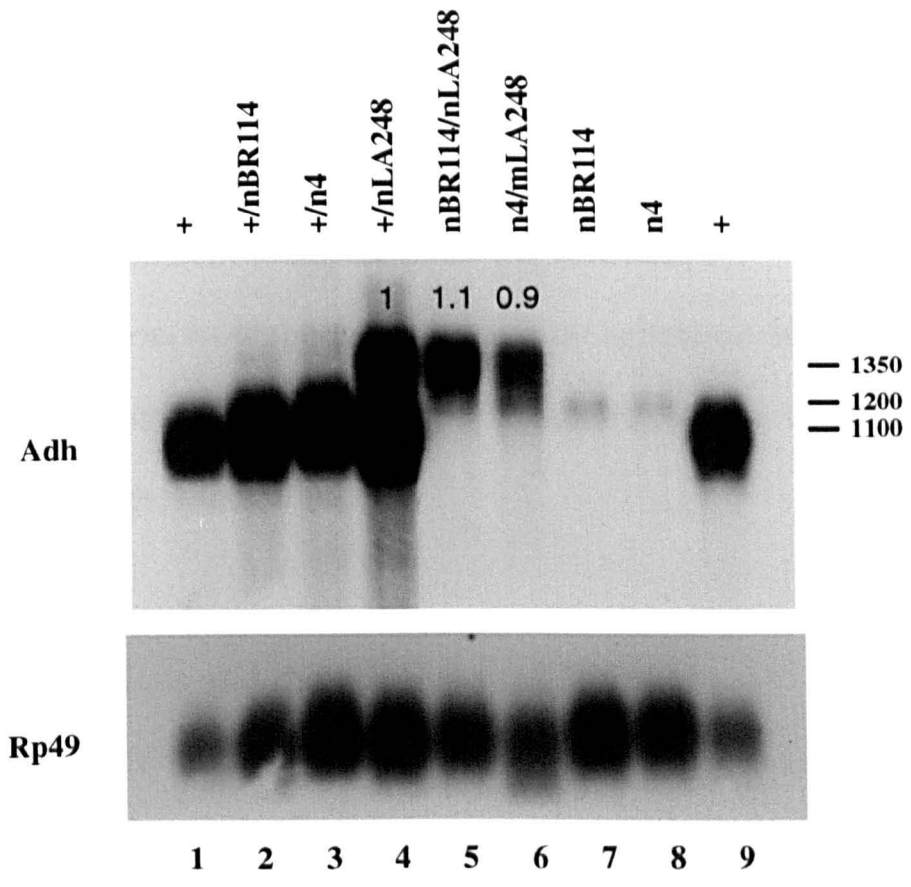
The data presented here strongly indicate that nonsense mutation can affect nuclear events, such as polyadenylation and splicing.

Studies in mammalian systems and yeast indicate that the effect of nonsense mutations on mRNA metabolism happens in cis, but since cytoplasmic translation is the only know means for detecting stop codons, it is conceivable that there may be a feed-back mechanism between translation in the cytoplasm and pre-mRNA processing in the nucleus. In addition ADH is a very abundant protein, it is estimated that it may represent as much as 1% of the total protein content and therefore translation inhibition of the *Adh* mRNA could potentially have some sort of pleiotropic effect on the metabolism of a

subset of mRNAs. More specifically, one possibility is that premature translation termination of *Adh* could have some effects on the abundance of the *Adh* transcript in trans. For example, an RNA binding protein could specifically bind to the *Adh* transcript and monitor its translation, and that this protein could shuttle between the nucleus and cytoplasm and in doing so affect the *Adh* pre-mRNA processing in the nucleus.

Here, contrary to the above hypothesis, it is shown that the effects of nonsense mutations on mRNA metabolism are strictly allele specific (figure 13). In heterozygous flies carrying a wild-type allele and either *nBR114* or *n4*, two of the most affected nonsense alleles, the level of the wild-type mRNA is similar, both in size and quantity to that of wild-type homozygous flies (lanes 2-3) . In addition, the same test was performed with flies heterozygous for *Adh* ^{nLA248} (lanes 4-6). This allele has an internal 244 bp duplication, spanning the last intron/exon junction of *Adh*, which produces a longer mRNA (Chia et al. 1985). Using this allele makes it easier to distinguish the transcripts from the two alleles by Northern blot analysis since the wild-type is about 1.1 kb while the *LA248* allele is about 1.3 kb (see lane 4). As for the other heteroallelic combinations, no trans-allelic effects are visible. For example, there is no evidence of the *nBR114* or *n4* alleles in trans significantly reducing the level of the *LA248* mRNA. The number just above the bands (lanes 4-6) indicates the relative level of the *LA248* mRNA in the presence of the wild-

Figure 13



Northern blot analysis of the Adh transcript in flies with different heteroallelic combinations of the Adh gene. Total RNA was purified from adult flies (heterozygous flies were selected from the F1progeny of a cross between the two given parental strains, using visible markers). The filter was first hybridised with an Adh specific probe, then stripped and reprobed with a cDNA from Ribosomal protein 49 (O'Connell and Rosbash, 1984). The numbers above the slow migrating Adh band (LA248) are ratios between the net number of counts for that band and the number of counts of the corresponding band in lane 4. The amount of radioactivity in each lane was measured directly from the filter using an InstantImager (Packard Instrumjents Co.). The number of counts was standarised for the variation in loading using the number of counts for the Rp49 band.

type allele (lane 4) or in the presence of the *nBR114* and *n4* alleles (lanes 5-6). The small variation between the relative level of the *LA248* transcript in the three lanes is within the experimental error detected in these kind of experiments (see figure 13 legend).

In addition, based on Northern blot analysis there is no indication of *nBR114* or *n4* increasing the mobility of the transcript of the other allele. On the other hand, the difference in mobility between the *Adh* transcript of *n4* and *nBR114* is very apparent (compare lanes 7-8 with lane 9).

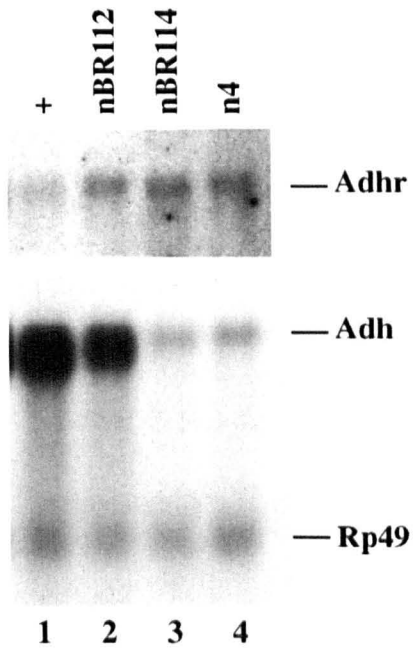
In conclusion, from these data there is no indication of premature translation termination of the mRNA of one allele affecting the metabolism of the mRNA from the other allele.

Nonsense mutations and the *Adh/Adh*r mRNA level.

As described in the introduction (see figure 2), the dicistronic mRNA that encodes for *Adhr* is essentially a longer version of the *Adh* transcript that has an extension at the 3' end. This extension contains the open reading frame encoding for ADHR while the first half of this dicistronic mRNA - up to the *Adh* polyadenylation site - is identical to the monocistronic *Adh* mRNA. Therefore, based on the current understanding of NMMR it was expected that nonsense mutations affecting the *Adh* monocistronic mRNA should also affect the dicistronic transcript (see introduction).

In order to test this hypothesis, I have first analysed the level of *Adh/Adhr* mRNA in nonsense mutants and in wild-type by Northern blot. These data, in figure 14, show that the level of the *Adh/Adhr* dicistronic transcript is clearly not reduced in the mutants (compare lane 1 (wild-type) with lanes 2-4 (mutants). Instead, it appears that the level of the dicistronic mRNA is slightly increased in the mutants relative to wild-type. This slight increase in the level of the *Adh/Adhr* transcript in nonsense mutants is apparent in several similar Northern blots, however its significance is not yet clear. One explanation, may be that the *Adh/Adhr* transcript is increased in the mutants because nonsense mutations in *Adh* by negatively affecting cleavage at the *Adh* poly(A) site would increase the usage of the downstream *Adhr* poly(A) site. This explanation is suggested by the observation that the level of the *Adh/Adhr* can be increased by mutations in *suppressor of forked* [*su(f)*] (Brogna and Ashburner 1997). *su(f)* is a gene involved in pre-mRNA 3' end processing (Mitchelson et al. 1993; Takagaki and Manley 1994). The increase in the level of the *Adh/Adhr* transcript in *su(f)* mutants is probably due to a reduced 3' processing at the first *Adh* poly(A) site, leading to an increase in the usage of the downstream *Adhr* poly(A) site.

Figure 14



Adhr and *Adh* mRNA levels in nonsense mutants.
Northern blot analysis of poly(A)+ RNA isolated from wild-type (lane 1) and mutant flies (lanes 2-4). Equal amouts of RNA were loaded in each lane. The filter was first hybridised with an ***Adhr*** specific probe (refer to figure 3) and then stripped and re-hybridised with both an *Adh* specif probe (refer to figure 5) and one specific for the ribosomal protein 49 (Rp 49, refer to figure12) as loading control..

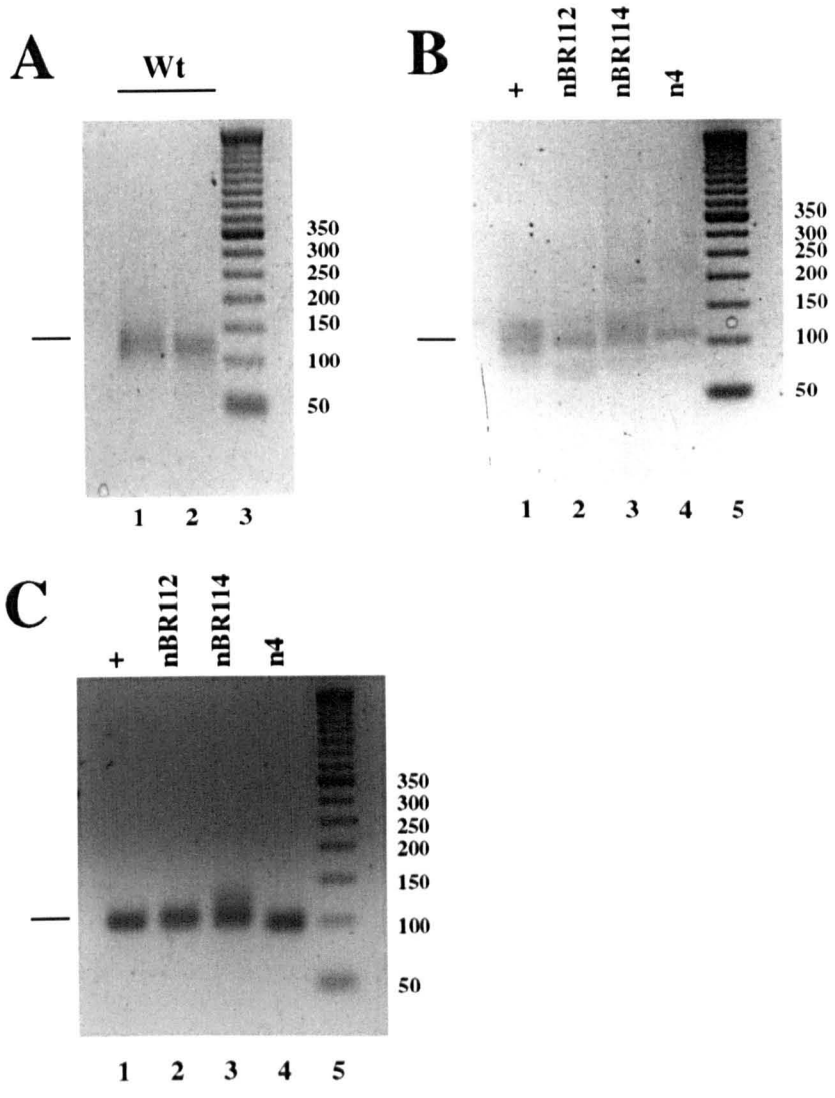
The poly(A) tail of *Adhr* is not affected by nonsense mutations.

Northern blot analysis does not show any indication of the *Adhr* transcript migrating differently in the nonsense mutants relative to the wild-type (see figure 14 for example). However, since nonsense mutations in the monocistronic *Adh* mRNA lead to a longer poly(A) tail of the transcript, I decided to further investigate this possibility by using the circular-RT-PCR technique previously used to look at the state of the 3' and 5' ends of the *Adh* transcript.

As I have described with the *Adh* transcript (figure 8) the mRNA was circularised with T4 RNA ligase and then single strand cDNA was synthesised across the junction using a primer in the *Adh* region. Then the region flanking the junction of the 5' and 3' ends was PCR amplified with one primer located very close to the 5' end and one close to the 3' end of *Adhr* (see legend figure 15).

Based on these c-RT-PCR experiments the length of the poly(A) tail of the wild-type *Adhr* mRNA should be about 40 nucleotides long - this is the difference between the size of the c-RT-PCR fragments produced from polyadenylated mRNA (about 140-150 bp, figure 15A, lanes 1-2) and that of the fragment generated from deadenylated mRNA (about 100 bp, Figure 15C, lane 1). This poly(A) tail is rather shorter compared to that found in

Figure 15



The length of the poly(A) tail of the Adhr mRNA does not vary between wild type and nonsense mutants - as assayed by cRT-PCR.

A: Gel electrophoresis of cRT-PCR products produced from wild type RNA. About 55 micrograms of total RNA was used for the experiment in lanes 1 and 10 micrograms for that in lane 2. Lane 3 is a 50 bp DNA marker. The line to the left of the picture indicates the position of the fragments. The cDNA was synthesised using primer E as in figure 8. The first round of PCR was with primer Adh2002.rev - 5'.TTAGCAGGCTCTTTCGAT.3') and primer Adhr.2962 - 5'.ACCCTCTTTACTTATTCCCAA.3'. The second round was with primers Adh1270.rev - 5'.AACTGACAACTGCACTGAGA.3' - and Adhr.3248 - 5'.CAGAGCAATGATGAAGAGGA.3'.

B: Similar cRT-PCR experiment as in A, from wild-type and nonsense mutants. 10 micrograms of RNA were used.

C: Experiment similar to that described in A but in which the RNA was treated with oligo(T) and RNase H to remove the poly(A) tail. Note that in this experiment the primers Adh2002.rev and Adh1270.rev were substituted with primers D and A as in figure 8 - primer A is located 7 bp upstream of primer Adh1270.rev. This fragment, based on the DNA sequence, is 118 bp long, but it runs slightly faster on the gel.

mammalian mRNAs, which are about 200 nt long, but is similar to that of the wild type *Adh* mRNA which is also about 40 nt long (see above).

In order to check whether the poly(A) is affected by nonsense mutations in the upstream *Adh* ORF, the same c-RT-PCR assay was used for the mutant transcripts. These experiments show a band of similar size both in wild-type and mutants (Figure 15B, compare the band indicated by the line between lane 1 with lanes 2-4), which indicate that the length of the poly(A) tail of the mutants is the same as in wild-type. In the mutants but not in wild-type are often visible also faint bands of higher molecular weight (for example, figure 15, lane 3). These faint bands are seen in some c-RT-PCR experiments but not others and their size is highly variable; at this stage is not clear what may originate them.

In summary, these experiment indicate that the length of the poly(A) tail of the dicistronic *Adhr* transcript does not appear to be affected by mutations in the upstream *Adh* ORF.

Chapter 3

RESULTS - part two

Drosophila Upf1.

As mentioned above, the *UPF1* gene was first identified in yeast in a screen for suppressors of nonsense mutations. The UPF1 protein is thought to be necessary for accurate translation termination and to be one of the key components of the *surveillance complex* (also known as the NMD system) that is assembled following a translation termination event (see introduction). This gene has also been cloned in human and *C. elegans* and it was therefore expected to be found also in flies.

Here I report the molecular characterisation of the *Drosophila melanogaster* homolog.

cDNA Cloning.

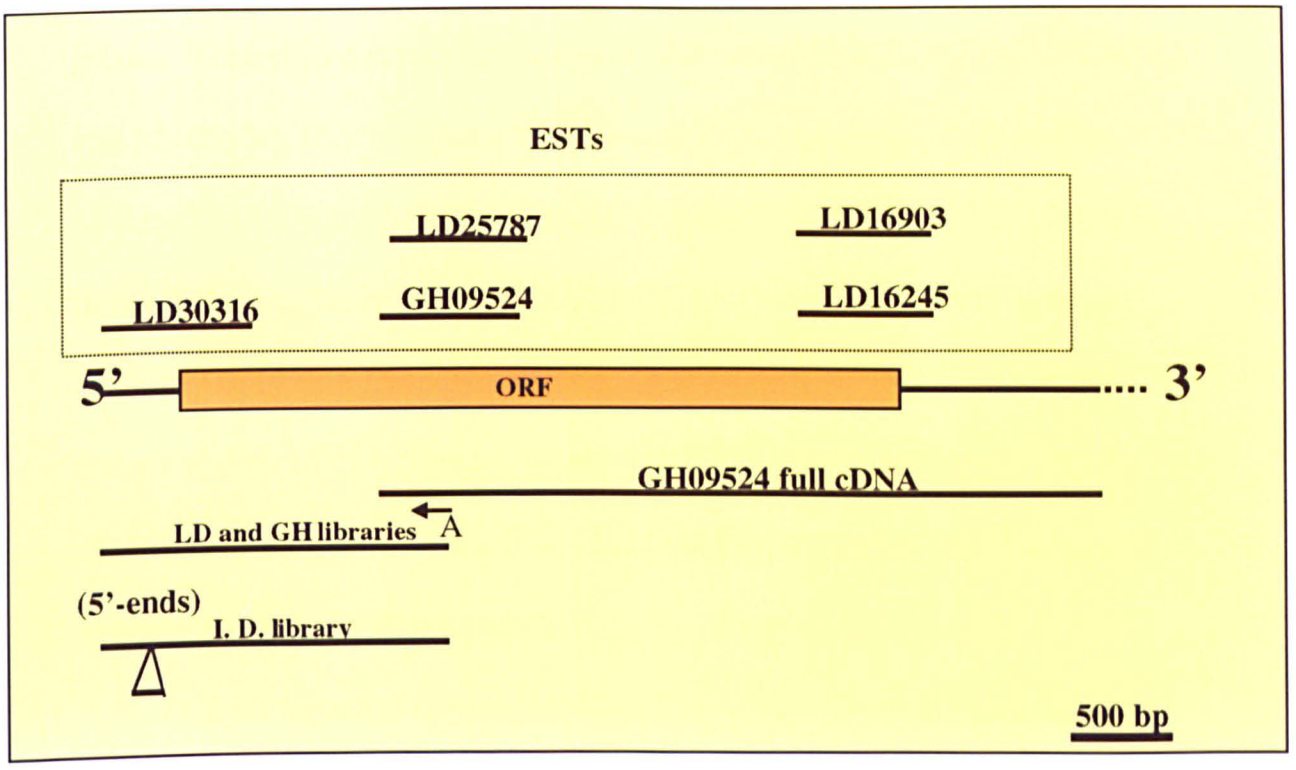
The characterisation of the *DmUpf1* gene (*D. melanogaster Upf1*) started with the identification of several sequences in the Drosophila EST database (Berkeley Drosophila Genome Project, BDGP-http://www.fruitfly.org/EST/) which, at the protein level, had strong similarity to the yeast and Human UPF1 proteins. I have restriction

mapped two of these clones, LD25787 and GH09524, and based on the restriction map they appear to derive from the same gene - the GH09524 had the longer insert and was completely sequenced.

The GH09524 clone is 3550 bp long and has a potential ORF of about 2553 bp. The translation of this ORF is predicted to encode for a protein of 851 aa that is very similar to the *S. ceravisiae* UPF1 sequence along most of its length. However, based on the alignment with the other UPF1 sequences it seemed that the *Drosophila* protein was about two hundred amino acids shorter at the N-terminus. It was therefore very possible that GH09524 was not a full-length cDNA.

In order to determine the correct 5' end of the gene, several cDNA libraries -including the GH and LD libraries from which the original EST clones were isolated - were screened by PCR for longer 5' ends, using a primer specific for the insert (GH09524) and a vector primer located in the region flanking the 5'-end of the insert (see Material and Methods). Fragments corresponding to longer cDNAs were amplified from most libraries. Most of these fragments, which were up to 1800 bp long, were sequenced and the sequence confirmed that they correspond to the missing 5' portion of the gene. The longer fragments (1800 bp) appears to include the part of the ORF missing in GH09524, plus 400 bp of 5' UTR (figures 16 and 17); in contrast,

Figure 16



A schematic diagram of the DmUpf1 cDNA and a scheme of its characterisation. In the dashed box, the lines correspond to the 5 hits in the Drosophila ESTs database. Below the diagram of the full-length cDNA, the lines correspond to the complete sequence of the GH09524 cDNA and that of the PCR fragments corresponding to the 5' end of the cDNA (note the two alternative 5' ends). The small line with the arrow shows the position of the most 3' located primer used to amplify the 5' end of the cDNA (information on the others primers are given in Materials and Methods. The dashed line at the 3' end of the ORF indicates that the 3' end of the GH09524 cDNA is probably upstream of the actual polyadenylation site (see text).

the shorter fragments (≤ 500) still lacked part of the ORF, they probably are partially extended cDNAs. The structure and sequence of this 5' terminal region was also checked by re-amplifying it by RT-PCR with one primer corresponding to the newly determined 5' end and the other primer located in a region included in the GH09524 cDNA (see Material and Methods); this fragment was completely sequenced. The sequencing of the cDNA 5' end fragments amplified from the GH and LD libraries plus all the available *DmUpf1* ESTs (figure 16) indicate that there is probably only one *DmUpf1* mRNA. One exception was found in the imaginal disk cDNA library, where the 5' UTR is 57 bp longer due to the retention of an intron in the 5' UTR, probably caused by an alternative splicing event (see also below).

Taken together the sequence of the GH09524 clone and the sequence of the PCR amplified cDNA fragments indicate that this putative Drosophila *Upf1* mRNA is about 5000 nuclotides long and contains a 3543 nt long ORF, a 400 bp 5' UTR (457 in the imaginal disk cDNA) and a 988 bp 3' UTR (figures 16 and 17).

Genomic organisation.

Very recently, the complete sequence of the genome of *D. melanogaster* was released (Adams et al. 2000) http://www.fruitfly.org/annot/, together with my cDNA sequence information this allowed the structure of the genomic region of *DmUpf1* to be determined.

The cDNA sequence was used to search the genome sequence database, and one hit was found in a region corresponding to the 10F7-10F8 segment of the X chromosome.

The alignment of the cDNA sequence with the genome sequence (figure 17) revealed that the gene contains 3 introns, one in the 5' UTR (57 bp) and two in the coding region - one near the beginning of the ORF, between codons 56 and 57 (61 nt long) and one at toward the end between codons 1110 and 1111 (68 nt long).

Interestingly, according to the cDNA sequencing data the 3' end of the transcript correspondences to a stretch of 11 A's on the coding strand of the DNA (position 6025 in the genomic sequence, figure 17). However, I suspect that the real 3' end cleavage site may be located further downstream; the 3' end of the cDNA is probably the result of the oligo(dT) primer annealing to this run of A's during the cDNA synthesis - this would therefore result in a cDNA deleted of the downstream region, between the stretch of A's and the real poly(A) site. This suspicion is further supported by the fact that there is no canonical poly(A) signal in the region preceding the cDNA 3' end - the poly(A) signal , AAUAAA, is one of the most conserved feature of the 3' end of mRNAs of higher eucaryotes (Proudfoot 1991). However, a

Figure 17

Sequence of the *DmUpf1* gene, in bold is the sequence of the cDNA. The three introns are labeled, intron1 through 3. Underlined are the two putative TATA boxes and the putative the poly(A) signal.

```
1   ---------+---------+---------+---------+---------+---------+   60
    GAGTCGATCTAGCTATGTCCCTCTGTCTGTCTGTCCGTATGAGGAACTATAAAAGCTAGA

61  ---------+---------+---------+---------+---------+---------+  120
    AGGTTGAGATTCAGCATACAGATTCTAGAGACAAAGACGCAGCGCAAGTTTGGTGACCCA

121 ---------+---------+---------+---------+---------+---------+  180
    TGTTGCCACGCCCACTCTAACGCCCACACTTCTGAACCATTTTTTGAAATTTTTCATAAT

181 ---------+---------+---------+---------+---------+---------+  240
    TTTTTTAGTTGCCAAACTGCGAAAAGACACGCCCACTCTAACGCCCTAGGGCAGCCGAAC

241 ---------+---------+---------+---------+---------+---------+  300
    CGGTCACGCCCACAGTTTGGAACATTTTTTTTTAATTTTTCTCATTTTATTCCCCAATATC

301 ---------+---------+---------+---------+---------+---------+  360
    TATCGATATCCCAGAAAAATGATAAAATTTCGCGTTCGCATTCACACTTGCTGAGTAACG

361 ---------+---------+---------+---------+---------+---------+  420
    GGTATCTTATAGTCGGGGAACTTGACTATAGCACTCTCTCTTGTTTTTTTTTATTTCCCT

421 ---------+---------+---------+---------+---------+---------+  480
    TAATGATTACAAGGAACCCTTGCTTCTGAAGGCAAATCTTGATGAGTCACATTATTGACA

481 ---------+---------+---------+---------+---------+---------+  540
    AAGAAAATATATATTTATATCATATATATTTATTGGGAACGAAAAAAGTTTTGATTATAA

541 ---------+---------+---------+---------+---------+---------+  600
    TGATTGCCTTCGCAACTAAAACTTTAACCTGTAGGAAATACTTGAAGGTTTTATTTTTTC

601 ---------+---------+---------+---------+---------+---------+  660
    TTCCATTTTATCCAACAAGTTATTTCAAGTTAAACCTTATGGTTAAATCCTGTTTGTAAT

661 ---------+---------+---------+---------+---------+---------+  720
    AGTTTATTATAAGGTCGACTATATAAGTACTTTTAAATAAATACGCAATACGGGAGTCCT

721 ---------+---------+---------+---------+---------+---------+  780
    ATGTTTGTTTATTCATTTCATAAATTTTGGCGATTTTGGCGTTTTCCAAAATAAACAGTA

781 ---------+---------+---------+---------+---------+---------+  840
    ATTTATGGTATAATTACTAAAAAGACAATACTTAAACGATATATTATTCACCCCATCACC

841 ---------+---------+---------+---------+---------+---------+  900
    GATAGTATCTATAGTCAACTTGCCCAACACTAGCGATTGATGTTGCCATACGCGATTATC

901 ---------+---------+---------+---------+---------+---------+  960
    GATAGACCCCTCGTGAGTCAAATGCAAACTGACGTATATTTATTACTGGCAAAACAGGCC
        ! cDNA 5' end.
```

961 ---------+---------+---------+---------+---------+--------- 1020
GGGCAAATTTGTTGAACAAGCAAGAAAGTTTTCGCTTTTTTTAATCCTACATTCCCCGCA

1021 ---------+---------+---------+---------+---------+--------- 1080
ATAGCGTGGGAATCTCGTACAGACTCAATTTAAAACACACAACAGCCACAGTTTCGAAAG

1081 ---------+---------+---------+---------+---------+--------- 1140
AGTGCCGAACGGCGGGCTCCCGAACACGGAAAGCGCGAAACGCACACACGACACCATCGC

1141 ---------+---------+---------+---------+---------+--------- 1200
CATACCATAATCATCGCGGCCACGCTCCAAGCTTAGGAATTTTTCAATTCCGGAAACACT

1201 ---------+---------+---------+---------+---------+--------- 1260
GACCAAGCGATCCGTGGTCTAGGAGCAGGTACTATAGGTGAGCATCAGTCCCATCCGAGT
Intron-1
1261 ---------+---------+---------+---------+---------+--------- 1320
TACGTACCTATCGGCCAATGCACATTGCTAACTAGATCTCATCGCGATCCCTGCGGAAAC

1321 ---------+---------+---------+---------+---------+--------- 1380
GCAGCATAATCACCTTGGGTGAGAAGAAACAACAGGCGCCATGAGCGTGGACACGTACGC
MetSerValAspThrTyrAl

1381 ---------+---------+---------+---------+---------+--------- 1440
GCCCAGCTCGGCGCTCTCGTTCCTGGACATGGACGACAACGAGCTGCTTCCGGGAGCGGA
aProSerSerAlaLeuSerPheLeuAspMetAspAspAsnGluLeuLeuProGlyAlaAs

1441 ---------+---------+---------+---------+---------+--------- 1500
TACTCAACCCACGCAGTACGATTACCGCGACTTTACCATGCCCTCCACCTCGCAGAGCCA
pThrGlnProThrGlnTyrAspTyrArgAspPheThrMetProSerThrSerGlnSerGl

1501 ---------+---------+---------+---------+---------+--------- 1560
GACCCAGAACGATCAGCTGGAGATTGCCGTGAGTTGTCCCCCAAATCATTCCGCGAAACC
nThrGlnAsnAspGlnLeuGluIleAla Intron 2
1561 ---------+---------+---------+---------+---------+--------- 1620
CAAATGGCATATACCCTTTTTAATTTACAGCAACGCTGCTCTGCCGGAGACTCGCATCCA
GlnArgCysSerAlaGlyAspSerHisPro

1621 ---------+---------+---------+---------+---------+--------- 1680
CGACTGGCCAGCATCACCAACGATCTGGCCGATCTGCAGTTCGAAGAGGAGGACGACGAG
ArgLeuAlaSerIleThrAsnAspLeuAlaAspLeuGlnPheGluGluGluAspAspGlu

1681 ---------+---------+---------+---------+---------+--------- 1740
CCTGGCAGCTCGTATGTGAAGGAGCTGCCGCCGCATGCGTGCAAGTATTGCGGCATCCAT
ProGlySerSerTyrValLysGluLeuProProHisAlaCysLysTyrCysGlyIleHis

1741 ---------+---------+---------+---------+---------+--------- 1800
GATCCAGCCACGGTGGTCATGTGCAACAACTGCCGCAAATGGTTTTGCAACGGTCGTGGA
AspProAlaThrValValMetCysAsnAsnCysArgLysTrpPheCysAsnGlyArgGly

1801 ---------+---------+---------+---------+---------+--------- 1860
AGCACTTCCGGTTCGCACATCATCAACCATCTGGTGAGGGCCAAGCATCGCGAGGTGACG
SerThrSerGlySerHisIleIleAsnHisLeuValArgAlaLysHisArgGluValThr

1861 ---------+---------+---------+---------+---------+--------- 1920
CTCCACGGGGAAGGTCCCCTGGGCGAGACAATCCTGGAGTGCTACTCCTGTGGTGTGCGC
LeuHisGlyGluGlyProLeuGlyGluThrIleLeuGluCysTyrSerCysGlyValArg

1921 ---------+---------+---------+---------+---------+--------- 1980
AACGTCTTTGTGCTGGGCTTCATTCCGGCCAAGGCCGATTCTGTGGTCGTGCTGCTCTGC
AsnValPheValLeuGlyPheIleProAlaLysAlaAspSerValValValLeuLeuCys

1981 ---------+---------+---------+---------+---------+--------- 2040
CGTCAGCCGTGTGCCGCCCAGAATTCGCTAAAGGATATGAACTGGGACCAGGAACAGTGG
ArgGlnProCysAlaAlaGlnAsnSerLeuLysAspMetAsnTrpAspGlnGluGlnTrp

2041 ---------+---------+---------+---------+---------+--------- 2100
AAGCCTCTAATTGCAGACCGCTGCTTTTTGGCCTGGCTGGTAAAGCAACCCAGCGAACAG
LysProLeuIleAlaAspArgCysPheLeuAlaTrpLeuValLysGlnProSerGluGln

```
2101 ---------+---------+---------+---------+---------+---------+ 2160
     GGACAGCTGCGAGCTCGCCAAATCTCAGCCGCTCAGATCAACAAGCTGGAGGAGCTATGG
     GlyGlnLeuArgAlaArgGlnIleSerAlaAlaGlnIleAsnLysLeuGluGluLeuTrp

2161 ---------+---------+---------+---------+---------+---------+ 2220
     AAGGAGAATATTGAGGCCACGTTTCAGGATCTGGAGAAGCCAGGCATTGACTCGGAGCCA
     LysGluAsnIleGluAlaThrPheGlnAspLeuGluLysProGlyIleAspSerGluPro

2221 ---------+---------+---------+---------+---------+---------+ 2280
     GCACATGTGCTACTCCGCTACGAGGATGGCTATCAGTACGAGAAGACCTTTGGGCCGCTG
     AlaHisValLeuLeuArgTyrGluAspGlyTyrGlnTyrGluLysThrPheGlyProLeu

2281 ---------+---------+---------+---------+---------+---------+ 2340
     GTCCGCCTTGAGGCCGAATACGACCAAAAACTGAAGGAGTCTGCCACGCAGGAGAACATC
     ValArgLeuGluAlaGluTyrAspGlnLysLeuLysGluSerAlaThrGlnGluAsnIle

2341 ---------+---------+---------+---------+---------+---------+ 2400
     GAAGTACGCTGGGACGTCGGCCTCAACAAAAAGACCATTGCCTACTTTACGCTGGCGAAG
     GluValArgTrpAspValGlyLeuAsnLysLysThrIleAlaTyrPheThrLeuAlaLys

2401 ---------+---------+---------+---------+---------+---------+ 2460
     ACCGATTCGGACATGAAGCTCATGCATGGCGACGAGCTGCGCCTGCATTATGTGGGCGAG
     ThrAspSerAspMetLysLeuMetHisGlyAspGluLeuArgLeuHisTyrValGlyGlu

2461 ---------+---------+---------+---------+---------+---------+ 2520
     CTGTACAATCCGTGGAGCGAGATCGGCCACGTTATCAAGGTGCCGGACAATTTCGGCGAT
     LeuTyrAsnProTrpSerGluIleGlyHisValIleLysValProAspAsnPheGlyAsp

2521 ---------+---------+---------+---------+---------+---------+ 2580
     GACGTCGGCCTGGAGCTGAAATCCTCAACGAATGCCCCGGTTAAGTGCACCAGTAACTTT
     AspValGlyLeuGluLeuLysSerSerThrAsnAlaProValLysCysThrSerAsnPhe

2581 ---------+---------+---------+---------+---------+---------+ 2640
     ACGGTGGACTTCATCTGGAAGTGCACGTCATTTGATCGCATGACACGTGCTCTGTGCAAA
     ThrValAspPheIleTrpLysCysThrSerPheAspArgMetThrArgAlaLeuCysLys

2641 ---------+---------+---------+---------+---------+---------+ 2700
     TTCGCCATCGATCGCAATTCAGTGTCGAACTTCATCTACTCGCGCCTGTTGGGCCACGGT
     PheAlaIleAspArgAsnSerValSerAsnPheIleTyrSerArgLeuLeuGlyHisGly

2701 ---------+---------+---------+---------+---------+---------+ 2760
     CGTGCGGATTCCAACGACGAGGTGCTGTTCCGCGGCCCACAACCCAAGCTCTTCAGTGCC
     ArgAlaAspSerAsnAspGluValLeuPheArgGlyProGlnProLysLeuPheSerAla

2761 ---------+---------+---------+---------+---------+---------+ 2820
     CCGCATCTGCCGGATTTGAATCGCAGCCAGGTGTATGCCGTGAAACACGCGCTTCAGCGT
     ProHisLeuProAspLeuAsnArgSerGlnValTyrAlaValLysHisAlaLeuGlnArg

2821 ---------+---------+---------+---------+---------+---------+ 2880
     CCGCTCTCGCTAATCCAAGGGCCGCCTGGCACGGGCAAAACCGTGACCTCGGCGACCATC
     ProLeuSerLeuIleGlnGlyProProGlyThrGlyLysThrValThrSerAlaThrIle

2881 ---------+---------+---------+---------+---------+---------+ 2940
     GTTTACCAGCTGGTCAAGCTCCATGGTGGCACAGTGCTGGTGTGCGCTCCCAGCAACACG
     ValTyrGlnLeuValLysLeuHisGlyGlyThrValLeuValCysAlaProSerAsnThr

2941 ---------+---------+---------+---------+---------+---------+ 3000
     GCCGTGGATCAGCTAACTGAGAAGATCCACCGAACAAACCTTAAAGTGGTGCGTGTTTGC
     AlaValAspGlnLeuThrGluLysIleHisArgThrAsnLeuLysValValArgValCys

3001 ---------+---------+---------+---------+---------+---------+ 3060
     GCCAAGAGCCGTGAGGCCATCGATAGCCCGGTAAGCTTCCTGGCGCTGCACAACCAAATC
     AlaLysSerArgGluAlaIleAspSerProValSerPheLeuAlaLeuHisAsnGlnIle

3061 ---------+---------+---------+---------+---------+---------+ 3120
     CGTAACATGGAGACCAACTCGGAGCTAAAGAAGCTGCAGCAGCTGAAAGACGAGACCGGC
     ArgAsnMetGluThrAsnSerGluLeuLysLysLeuGlnGlnLeuLysAspGluThrGly
```

```
3121  ---------+---------+---------+---------+---------+---------+  3180
      GAGCTGAGCTCAGCAGACGAAAAGCGATACCGCAACCTGAAACGTGCCGCCGAGAACCAA
      GluLeuSerSerAlaAspGluLysArgTyrArgAsnLeuLysArgAlaAlaGluAsnGln

3181  ---------+---------+---------+---------+---------+---------+  3240
      CTGCTGGAGGCTGCCGACGTTATCTGCTGCACATGCGTAGGCGCCGGCGATGGTCGTCTA
      LeuLeuGluAlaAlaAspValIleCysCysThrCysValGlyAlaGlyAspGlyArgLeu

3241  ---------+---------+---------+---------+---------+---------+  3300
      TCGCGAGTCAAGTTCACCTCGATCCTGATCGATGAGTCTATGCAGTCGACGGAGCCGGAG
      SerArgValLysPheThrSerIleLeuIleAspGluSerMetGlnSerThrGluProGlu

3301  ---------+---------+---------+---------+---------+---------+  3360
      TGCATGGTGCCAGTGGTGCTGGGCGCTAAGCAGCTGATCCTCGTGGGCGATCACTGCCAG
      CysMetValProValValLeuGlyAlaLysGlnLeuIleLeuValGlyAspHisCysGln

3361  ---------+---------+---------+---------+---------+---------+  3420
      CTGGGACCGGTTGTTATGTGCAAGAAAGCAGCTCGTGCCGGCCTCTCGCAAAGTTTGTTC
      LeuGlyProValValMetCysLysLysAlaAlaArgAlaGlyLeuSerGlnSerLeuPhe

3421  ---------+---------+---------+---------+---------+---------+  3480
      GAGCGCCTGGTGGTTCTGGGCATCCGTCCGTTCCGGCTGGAGGTGCAATATCGCATGCAC
      GluArgLeuValValLeuGlyIleArgProPheArgLeuGluValGlnTyrArgMetHis

3481  ---------+---------+---------+---------+---------+---------+  3540
      CCCGAGCTGTCCCAGTTCCCGTCCAACTTCTTCTACGAGGGATCGCTGCAAAACGGCGTC
      ProGluLeuSerGlnPheProSerAsnPhePheTyrGluGlySerLeuGlnAsnGlyVal

3541  ---------+---------+---------+---------+---------+---------+  3600
      TGCGCGGAGGATCGTCGCCTTAAGCTTGATTTCCCCTGGCCACAGCCGGAGAGACCGATG
      CysAlaGluAspArgArgLeuLysLeuAspPheProTrpProGlnProGluArgProMet

3601  ---------+---------+---------+---------+---------+---------+  3660
      TTCTTTCTGGTAACACAGGGACAGGAGGAGATTGCCGGCTCCGGCACCTCTTTCCTCAAC
      PhePheLeuValThrGlnGlyGlnGluGluIleAlaGlySerGlyThrSerPheLeuAsn

3661  ---------+---------+---------+---------+---------+---------+  3720
      CGCACAGAGGCGGCCAACGTGGAAAAGATTACGACGCGATTCCTTAAGGCAGGCATCAAG
      ArgThrGluAlaAlaAsnValGluLysIleThrThrArgPheLeuLysAlaGlyIleLys

3721  ---------+---------+---------+---------+---------+---------+  3780
      CCGGAACAAATTGGAATCATCACGCCTTACGAAGGTCAGCGCGCTTACCTGGTGCAGTAC
      ProGluGlnIleGlyIleIleThrProTyrGluGlyGlnArgAlaTyrLeuValGlnTyr

3781  ---------+---------+---------+---------+---------+---------+  3840
      ATGCAATACCAGGGCAGCCTGCACTCTCGTCTATACCAGGAGATCGAGATCGCCAGTGTG
      MetGlnTyrGlnGlySerLeuHisSerArgLeuTyrGlnGluIleGluIleAlaSerVal

3841  ---------+---------+---------+---------+---------+---------+  3900
      GACGCGTTCCAGGGACGTGAGAAGGACATCATTATCATGTCTTGCGTGCGATCTAACGAA
      AspAlaPheGlnGlyArgGluLysAspIleIleIleMetSerCysValArgSerAsnGlu

3901  ---------+---------+---------+---------+---------+---------+  3960
      CGTCAAGGCATCGGCTTTTTGAACGACCCACGTCGCCTTAATGTTGCCCTTACACGGGCC
      ArgGlnGlyIleGlyPheLeuAsnAspProArgArgLeuAsnValAlaLeuThrArgAla

3961  ---------+---------+---------+---------+---------+---------+  4020
      AAGTTCGGGATCATCATTGTGGGCAATCCCAAAGTGCTCGCCAAGCAGCAGCTGTGGAAC
      LysPheGlyIleIleIleValGlyAsnProLysValLeuAlaLysGlnGlnLeuTrpAsn

4021  ---------+---------+---------+---------+---------+---------+  4080
      CATCTGCTCAACTTCTACAAGGACCGTAAGGTGCTCGTCGAGGGATCTCTGAATAACCTT
      HisLeuLeuAsnPheTyrLysAspArgLysValLeuValGluGlySerLeuAsnAsnLeu

4081  ---------+---------+---------+---------+---------+---------+  4140
      AAGGAGTCGCTAATCCACTTCCAGAAGCCCAAAAAGCTTGTCAACAGCATGAACATTGGG
      LysGluSerLeuIleHisPheGlnLysProLysLysLeuValAsnSerMetAsnIleGly
```

```
4141 ---------+---------+---------+---------+---------+---------+ 4200
     GCACACTTTATGTCCACCATTATTGCCGATGCCAAGGAAGTGATGGTGCCAGGCTCCATT
     AlaHisPheMetSerThrIleIleAlaAspAlaLysGluValMetValProGlySerIle

4201 ---------+---------+---------+---------+---------+---------+ 4260
     TACGACCGCAGTGGCGGTTACGGCCAAGGTCGCCAAATGGTGGGACAGTCAATGAATGGC
     TyrAspArgSerGlyGlyTyrGlyGlnGlyArgGlnMetValGlyGlnSerMetAsnGly

4261 ---------+---------+---------+---------+---------+---------+ 4320
     GGACAGTACGGTGGCAGTGGAGGTGGTCCCTACGGAAACTCACCCCTCGGCTACGGTACT
     GlyGlnTyrGlyGlySerGlyGlyGlyProTyrGlyAsnSerProLeuGlyTyrGlyThr

4321 ---------+---------+---------+---------+---------+---------+ 4380
     CCCAGCTCCAATTCCATGGTGGGCTTTGGCCTGGGCAACGGAGGCAATGGCGCGGCCGGT
     ProSerSerAsnSerMetValGlyPheGlyLeuGlyAsnGlyGlyAsnGlyAlaAlaGly

4381 ---------+---------+---------+---------+---------+---------+ 4440
     GGCAACAACAACTTCGGAGGGGCTGGACCCAGTTGGGCGGCTGCCCACCTCCACCACGAC
     GlyAsnAsnAsnPheGlyGlyAlaGlyProSerTrpAlaAlaAlaHisLeuHisHisAsp

4441 ---------+---------+---------+---------+---------+---------+ 4500
     TCCATTGGCTATATATCCAACGAGCATGGAGCAGCAGCACTGGGCAACATGCCAGTTCCG
     SerIleGlyTyrIleSerAsnGluHisGlyAlaAlaAlaLeuGlyAsnMetProValPro

4501 ---------+---------+---------+---------+---------+---------+ 4560
     GTTGGCATGTTCATGAACATGAGCAATATTCCGCCGCGTTTCTACAACCAGCACCAGCAG
     ValGlyMetPheMetAsnMetSerAsnIleProProArgPheTyrAsnGlnHisGlnGln

4561 ---------+---------+---------+---------+---------+---------+ 4620
     GCGATCATGGCGGTCAAGCAGAATCGCGCCATTCAACAACAGACGGGTAATTTCTCTCCC
     AlaIleMetAlaValLysGlnAsnArgAlaIleGlnGlnGlnThrGlyAsnPheSerPro

4621 ---------+---------+---------+---------+---------+---------+ 4680
     GGTAACTCGGGTCCTGGAGTCACTGGAGTCGGAGTCGGACGAAGCGCCACCCCAGGCGGC
     GlyAsnSerGlyProGlyValThrGlyValGlyValGlyArgSerAlaThrProGlyGly

4681 ---------+---------+---------+---------+---------+---------+ 4740
     AATAAGAAGACCAACAAGCTGGGAAAATCGCGCGTAACGGGCGGTGGAACTGGCGGGGCA
     AsnLysLysThrAsnLysLeuGlyLysSerArgValThrGlyGlyGlyThrGlyGlyAla

4741 ---------+---------+---------+---------+---------+---------+ 4800
     CCGCTAACACAAGGAAGCTCGGTATGCAATGCTGCTCCATACAGTCAGCACCCGATGCCT
     ProLeuThrGlnGlySerSerValCysAsnAlaAlaProTyrSerGlnHisProMetPro

4801 ---------+---------+---------+---------+---------+---------+ 4860
     TTGTCGCTGCAGATGACCCAGCCCAGCGGATTTGCTCTGTCCCAGCAGCCGGAACTTTCA
     LeuSerLeuGlnMetThrGlnProSerGlyPheAlaLeuSerGlnGlnProGluLeuSer

4861 ---------+---------+---------+---------+---------+---------+ 4920
     CAGGACTTTGGGCAAATATCGCAGATGGACGGTTTGCTATCCCAGGATGTTGCCTTTAAC
     GlnAspPheGlyGlnIleSerGlnMetAspGlyLeuLeuSerGlnAspValAlaPheAsn

4921 ---------+---------+---------+---------+---------+---------+ 4980
     GCGGTAAGTGTGGAAGTGATATTTTTGAAACCCAAAGAGTGTTAGGTAATCTGAATTATT
     Ala
          ____________________________Intron 3____________________

4981 ---------+---------+---------+---------+---------+---------+ 5040
     TGTGTTTCGCAGTCGGGCGAGCGGAGCTTGAATCAGTTCTCACAGCCTTATTGAGAATCC
     ___________SerGlyGluArgSerLeuAsnGlnPheSerGlnProTyr

5041 ---------+---------+---------+---------+---------+---------+ 5100
     GGTTGGAGGTGCAACAACTACAAACGATAACAAGCAACACAAACGAAACAACAACAAACT

5101 ---------+---------+---------+---------+---------+---------+ 5160
     CAACTTTGAAGTCATCATCAAGGAGTAGCATCAGTAGGCAGCAGGAAAAAGCAGAGAGGA

5161 ---------+---------+---------+---------+---------+---------+ 5220
     CCATCACGGGAACAACGACAGATGCGGGAAACAATGCGACCTAAGAACATTAGGACAAAG
```

```
5221  ---------+---------+---------+---------+---------+---------+  5280
      CGGGACATTAAAAGATGGCAAATCGAGACGAGGCAGCGTGATAGATGGAGACGGATATG
5281  ---------+---------+---------+---------+---------+---------+  5340
      CAAGATGTGGAGTTACAAAGTAATAAAGAAGACTGAGACGGGAGAGAGATGTCGAAAGAT

5341  ---------+---------+---------+---------+---------+---------+  5400
      CGAAATATCGAGAGCAAATCACTGTAGGCTGCCAGGATGCACAGCTGCAGTTGCTGTTTC

5401  ---------+---------+---------+---------+---------+---------+  5460
      CGTAGCGCCGCGGGCCATGGAGTCATAGCCAGGCGACATGCCATTGGCCACCGGCCATTG

5461  ---------+---------+---------+---------+---------+---------+  5520
      GCCATTGGTCACGAGTGCGCTCCCGAGACGGAGTTCGCCCACGGATGAGCCAATCTGCCA

5521  ---------+---------+---------+---------+---------+---------+  5580
      GCAAGGCGCCACACAAGATGGACAACTTCTCGACACGTTTCTTTTCTCTTTTGCCAGCCC

5581  ---------+---------+---------+---------+---------+---------+  5640
      AGGATTCGCAGCCGGAGCCTAACAAAGACGCAATCACTCCCGCCCAGGACATGTATTCAC

5641  ---------+---------+---------+---------+---------+---------+  5700
      CAAGTGGCTTATCTTCTAACCTTCTTCTGGCCTCTAGTGCAGCAGCCAATCTTTAGCTCC

5701  ---------+---------+---------+---------+---------+---------+  5760
      TGTTTTAACATTTTCCCACCCAAAGCGGTTTATACCGATCCCCATATATGTATGCCCACT

5761  ---------+---------+---------+---------+---------+---------+  5820
      GTTGTACATTTTTGGACTCTTTTTTGATGCATTTGTTATCCTAACAAACACGAAAACAAA

5821  ---------+---------+---------+---------+---------+---------+  5880
      ATCAAACAAAAACAAGGAAAATCACGCAAACAAAGAATCAAGACGAAACAAAATCAAATT

5881  ---------+---------+---------+---------+---------+---------+  5940
      TAAAAAACATGGATACACCATTTATAATCGCTCGGGAACGCTGGCGGATCCCCCCAAGCG

5941  ---------+---------+---------+---------+---------+---------+  6000
      GATGCAGGCTTCCAGCGCAGTTCGGATATCGAATGGTTGATAGGAAAACGCATTGCGATT

6001  ---------+---------+---------+---------+---------+---------+  6060
      TACCGATCTACATAGGTACACTACAAAAAAAAAAAACACCACAAAAAAACAGGCGAACATT
                                  ! 3' end GH09524 cDNA clone

6061  ---------+---------+---------+---------+---------+---------+  6120
      AACAAACTGAACCAACGGAGTAAATGATCAGGATCAGAAAAGGTTAGTTATTAATAATAT

6121  ---------+---------+---------+---------+---------+---------+  6180
      TAATGCAGTTACAGATGCCTGATCATTGTTTTTAAAACGACATGAAAATCGTATATATTT

6181  ---------+---------+---------+---------+---------+---------+  6240
      TCTATATATTGTCAGTTAATTCTTCCTCCATGCGACTCTTTGTATCAAGAAAATGGAACA

6241  ---------+---------+---------+---------+---------+---------+  6300
      TTTTGTTTGAAATTCAAAATAAAGAATTCCCTTGCGAAAAAAAAAAAAACCAATATTCGCG
                          ! putative polyadenylation signal

6301  ---------+---------+---------+---------+---------+---------+  6360
      CCAAGAATTTTCTCTTTGGGGCACGTGAAAATTGAATTTATATTCGAGTTTCCAGACACT

6361  ---------+---------+---------+---------+---------+---------+  6420
      TGGCTGTTTTGTTGCTTTTTAATTTACAGAACTTGCAATTTTACATTTTCGGATGACATG

6421  ---------+---------+---------+---------+---------+---------+  6480
      TTGTTTTTCAAATAGATACATACACACCACTGGAGGAACGAACATCGCACACCTCCGTGG

6481  ---------+---------+-  6501
      CTGTTTACAATAATTAATGGA
```

canonical AATAAA sequence is found in the 223 bp further downstream in the DNA coding strand.

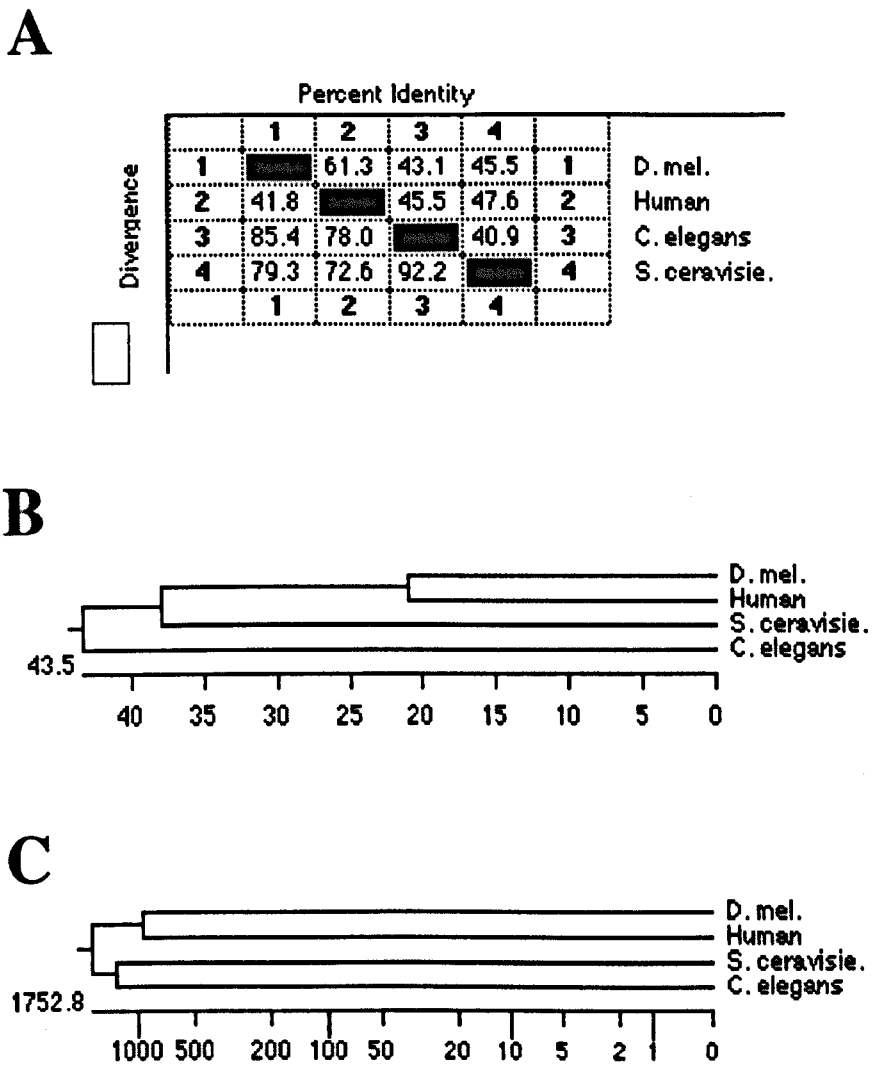
UPF1 proteins sequence analysis.

Based on the cDNA and genomic sequences, the predicted Drosophila UPF1 protein is 1181 amino acids long. Overall, this polypeptide is 61.3 % identical to the human homolog and 43.1 % and 45.5 % identical to the *C. elegans* and *S. cerevisiae* proteins, respectively (Figure 18). The Drosophila sequence is, as expected, more closely related to the human sequence; however the fly sequence is slightly more similar to the yeast sequence than it is to the *C. elegans* sequence. Phylogenetic trees constructed using two different methods indicate that the Drosophila sequence is equidistant from the worm and yeast sequences (Figure 18 B-C). This is a bit surprising, given that usually *C. elegans* genes are significantly more similar to genes from higher organisms than to yeast.

The multiple sequence alignment (Figure 19) of the fly sequence with the human, worm and yeast sequences shows that the central portion of the protein (approximately from position 100 to 900 of the fly sequence) is the most conserved: 83% similar and 78 % identical to the human sequence and approximately 67 % similar and 57 % identical to the yeast and worm proteins. The N-terminal part of this central region (position 100-190) is

Figure 18

UPF1 phylogenies



Phylogenetic relationships between *D. melanogaster* UPF1 and the corresponding Human, *C. elegans* and *S. ceravisie* proteins - sequence alignments are shown in figure 19.

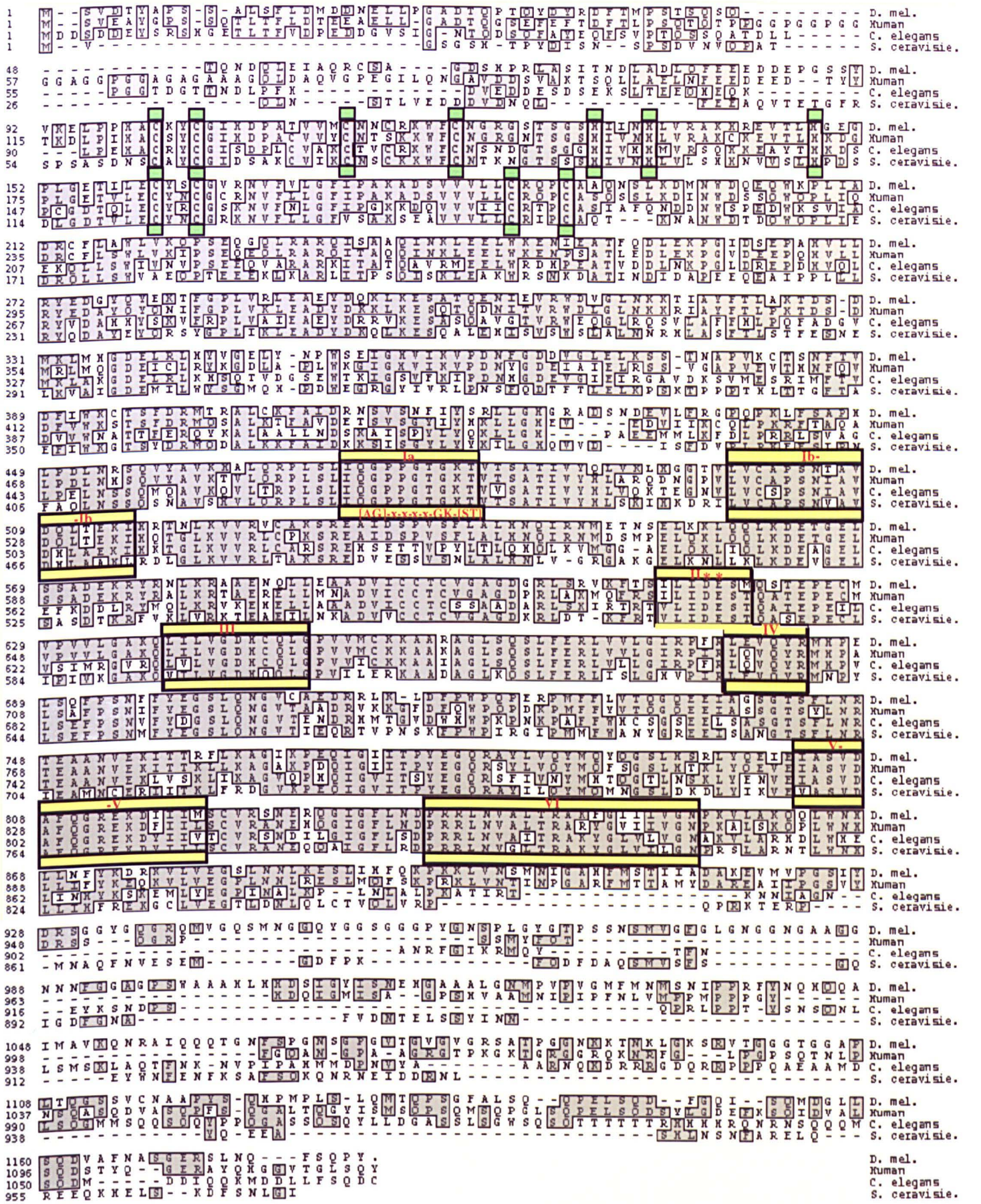
A: A matrix of simple (uncorrected for multiple hits) distances, above the diagonal are percentage of identities and below are percentage of differences.

B: This tree is produced by the MEGALIGN program with the Clustal method. Note that the different grouping of the Drosophila and human branches relative to the other two Sequences.

C: Phylogenetic tree produced using the MEGALIGN program of the DNA-STAR software package, this tree is according to the Jotun Hein method (parsimony).

Figure 19

UPF1 sequence alignment



Legend in the following page

Legend figure 19

Sequence comparison between UPF1 proteins.: *Drosophila melanogaster*, Human (AAC50771), *C. elegans* (AAC26789) and *S. cerevisiae* (CAA89226).

The sequences were aligned with the MEGALIGN program from the DNA-STAR package, using the Clustal algorithm. The numbers refer to the amino acids positions. In shaded grey are the residues conserved relatively to Drosophila or between any two sequences. In green are highlighted the cysteine and histidine residues thought to form a Zn finger-like structure. In yellow are indicated the putative seven (Ia-Ib and II-VI) putative motifs characteristic of group I RNA helicases. Motif I has the characteristic sequence of a ATP/GTP binding domain. In motif II, asterisks indicate the two residues mutated to alanine in the DmUpf1DE617AA construct.

cysteine and histidine rich with 8 cysteine and 3 histidine residues conserved in all four sequences; these residues are thought to form a Zn finger-like structure (see introduction). The C-terminal part of the highly conserved region (position 472- 856) contains the seven sequence motifs typical of group I helicases, which are found in the other UPF1 proteins and in other group I helicases (Koonin 1992). In addition the Drosophila protein contains all those residues that are thought to be essential for UPF1, based on genetical and biochemical studies in yeast (Weng et al. 1996b; Weng et al. 1996a).

Outside the central region the four sequences are more divergent. However there is still a significant level of conservation between the fly and human sequences with short regions of almost complete conservation. The C-terminus is the most diverged part of the protein, although several positions are conserved between the fly and human sequences; particular is noticeable a stretch of 7 completely conserved amino acids - QPELSQD.

Compared to the yeast UPF1, the Drosophila sequence is 210 residues longer, with a 32 amino acid longer N-terminus and a 188 residue longer C- terminus. The Drosophila sequence is also longer than the worm and human sequence by 112 and 63 residues, respectively.

In conclusion, based on the above sequence analysis, I conclude that this fly protein is a putative group I helicase and most probably the functional homolog of yeast, human and worm UPF1.

Chapter 4

DISCUSSION

Nonsense and splicing mutations reduce the level of the *Adh* mRNA in a position-dependent manner.

In this report, I show that in the *Adh* gene of *D. melanogaster*, nonsense mutations and mutations that prevent splicing of one of the introns lead to a reduction in the level of the corresponding mRNA. The level of this reduction correlates with the distance of the mutation from the beginning of the open reading frame: mutations at the beginning of the ORF cause severe reductions of the mRNA level while mutations in the second half of the ORF cause only minor reductions.

The issue of what determines whether or not a nonsense mutation reduces the mRNA level has not directly been investigated in this study, but it is interesting to note that those mutations that do not significantly affect the mRNA level are both located in the last exon.

In view that nonsense mutations may also directly affect pre-mRNA 3'-end formation (discussed in detail below), it is interesting to note that whether a nonsense mutation can or can not reduce the mRNA level may simply reflect the distance of the mutation from the poly(A) signal. The cleavage/polyadenylation machinery, which assembles on the poly(A)

signal and its flanking regions, is a complex RNP structure that interconnects several aspect of mRNA biogenesis. Therefore, as I will further discuss below, it is possible that the distance of the nonsense mutations from the cleavage/polyadenylation machinery is the key aspect to distinguish between those mutations that do affect mRNA abundance and those that do not.

As I describe in the introduction, another important determinant in distinguishing mutations that cause mRNA reduction and mutations that do not could be the presence of a downstream intron. However, not all genes have introns; furthermore in yeast downstream introns are not required for NMMR (very few yeast genes have introns). Instead a DSE is required downstream of the mutation for it to cause NMMR in yeast (see introduction). Therefore, the distance from the 3' end may be a more general feature to distinguish the two kinds of mutations. Interestingly, it was recently suggested that the effect of the DSE in yeast may be to simply change the spatial relationship of the nonsense mutation relative to the 3'-terminal mRNP domain (Hilleren and Parker 1999).

Nonsense mutations in *Adh* reduce the level of the nuclear mRNA. The results presented here, similarly to those observed in several mammalian systems, indicate that the overall reduction of the mRNA in the cell may simply be the consequence of a reduced pool of mRNA in the nucleus. This conclusion is based on the observation that the mRNA

reduction is apparent also in nuclear fractions, and the extent of the nuclear reduction is similar to that seen in the whole cell.

However, the distinction between what is nuclear and what is cytoplasmic relies only on the cell-fractionation protocol used. In the protocol I used, the nuclear fractions were treated with a ionic detergent which was previously shown to strip the outer layer of the nuclear envelope and the ribosome and other cytoplasm material associated with it (Belgrader et al. 1994). But it should be pointed out that although the nuclear associated mRNA could be located within the nucleoplasm (the nucleus proper) it could also be associated with the nuclear pore or be almost fully exported but not yet released from the nucleus. Hence the definition of nuclear is rather vague and depends on the fractionation protocol used (Hentze and Kulozik 1999). Therefore, taking the reduction of the level of mRNA in nuclear fractions alone as an indication that nonsense mutations can affect mRNA levels independently of cytoplasmic translation is a weak argument. The demonstration that nonsense mutations can affect pre-mRNA processing makes a case for a nuclear effect of nonsense mutations.

Nonsense mutations affect 3'-end pre-mRNA processing .

An unexpected finding of this work is that NMRR in the *Adh* gene is associated with an increase in the length of the poly(A) tail of the transcript. The longer poly(A) tail of the mRNA is probably the direct consequence of a longer poly(A) tail being added in the nucleus to the pre-mRNA. This conclusion is based on the observation that the poly(A) tails of both nuclear mRNA and pre-mRNA are as long as that of the cytoplasmic mRNA.

The poly(A) tail is added in the nucleus during the process of pre-mRNA 3' end formation, which consists of endonucleolytic cleavage followed by polyadenylation of the upstream fragment (Proudfoot 1996; Colgan and Manley 1997). The length of the poly(A) tail appears to be controlled by regulating the activity of the poly(A) polymerase. The poly(A) binding proteins, Pab I (in yeast) and Pab II (in mammals), are thought to be involved in this process. In yeast, cellular extracts from strains mutated or deleted for the PAB1 gene generate abnormally long poly(A) tails (Amrani et al., 1997; MinvielleSebastia et al., 1997). However, in mammalian cells, Pab1 seems to accumulate only in the cytoplasm. Here, it is Pab II which is thought to be the main factor in regulating the length of the poly(A) tail, and it does this by stimulating the elongation of a shorter poly(A) tail (Wahle 1995).

In comparison to the phenotype of PAB1 mutations in yeast, one could interpret the data in this thesis as an indication that in *Drosophila*

(at least for the *Adh* gene), the relatively short length of the poly(A) tail of the wild-type *Adh* transcript may be achieved by preventing its elongation to a 'default' longer tail. However, the *Drosophila* homolog of PAB2, in a reconstituted mammalian polyadenylation system, has the same functions as its bovine homolog: it stimulates poly(A) polymerase and leads to a rapid poly(A) extension of about 250 nt (Benoit et al. 1999). Therefore, since the poly(A) tails of several *Drosophila* mRNAs is, including *Adh* and *Adhr*, are much shorter than those of the typical mammalian mRNA, clearly there must be a further regulation of the activity of PAB2 to prevent potentially much longer poly(A) tails. The factors controlling in vivo the activity of PAB2 are unknown. One possibility is that in vivo PAB2 activity is regulated by PAB1. The prediction from this hypothesis would be that mutations in the *Drosophila* homolog of *PAB1* should lead to mRNAs with longer poly(A) tails.

Whatever the mechanism that controls the length of the poly(A) tail, from the results in this work it is apparent that the presence of a premature termination codon appears to affect this process and leads to longer poly(A) tails. Therefore, it seems to be a reasonable suggestion that there may be a link between nuclear polyadenylation and scanning of the coding region.

As discussed in detail below, more work is required to establish if these are direct effects on the machinery that generates the 3' end of the

mRNA or rather if the poly(A) extension is a secondary consequence of interference with other steps during pre-mRNA processing. For example, at this stage it cannot be excluded that the elongation of the poly(A) tail could just be a secondary consequence of a delay in splicing and/or trafficking.

What's the link between longer poly(A) tails and mRNA reduction?

The increase in length of the poly(A) tail is not expected to trigger mRNA decay. This is suggested by several studies in yeast and mammals indicating that poly(A) shortening actually precedes mRNA degradation (Decker and Parker 1994; Beelman and Parker 1995; Couttet et al. 1997). In addition, in mammalian cells it has been experimentally demonstrated that most often there is no difference between the stability of mRNAs carrying premature termination codons and wild-type transcripts (see introduction). Obviously it would be interesting to check the stability of these nonsense mRNAs. In Drosophila, unfortunately, it has been difficult so far to establish a reliable method of measuring mRNA half-life in the whole fly.

Here, however, the preliminary observation that unspliced transcripts with abnormal 3' ends accumulate in the mutants may suggest that the reduction of mRNA levels, rather than reflecting a change in the stability of the mature transcript, might actually just be a consequence of an increase in non productive pre-mRNA processing events. This

hypothesis is also in agreement with the observation that in the mutants the 5' end is truncated, which may indicate that these abnormal transcripts are degradation intermediates.

However, it cannot be ruled out that the mRNA reduction, the elongation of the poly(A) tail and the increase of unspliced transcripts may be independent events. For example, the mRNA reduction may be due to a peculiar decay mechanism that acts on the mRNA while it is still associated with the nucleus. This second model is equivalent to the translation-transport-decay model proposed to explain the 'nuclear' reduction of nonsense mRNAs that are stable once in the cytoplasm (Maquat, 1995). However, this latter model would still not explain the elongation of the poly(A) tail and the increase of the level of "pre-mRNA" in the mutants (see below).

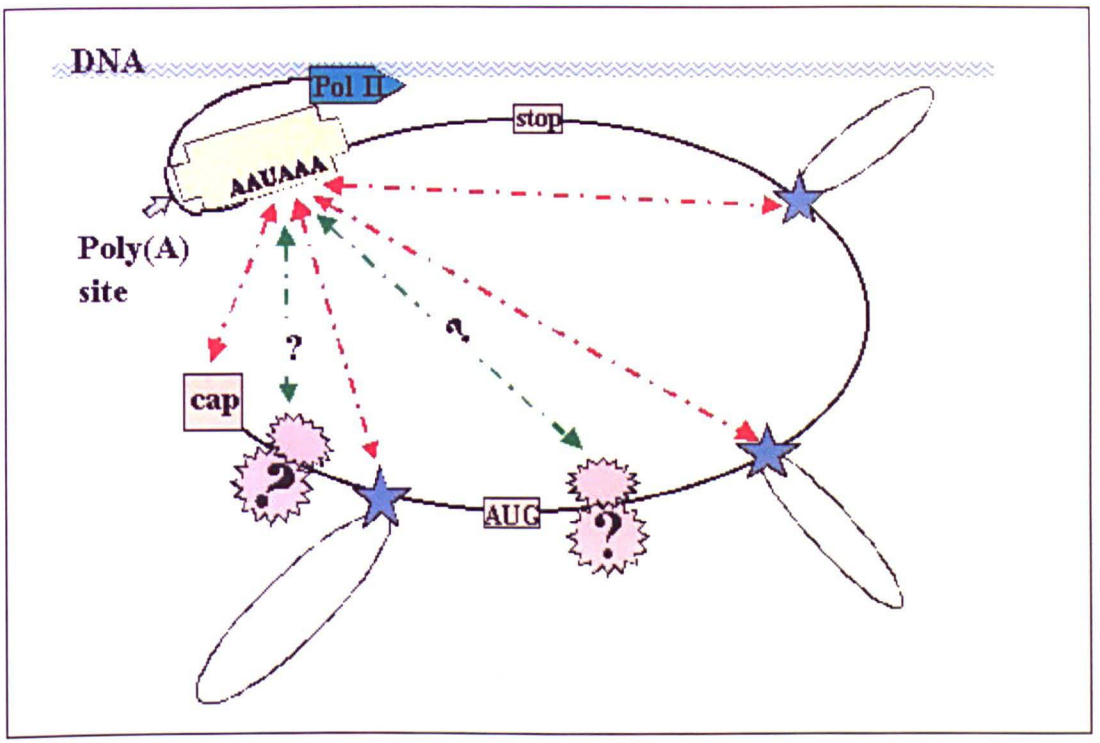
A nuclear scanning machinery?

Translation in the cytoplasmic compartment is the only known mechanism which recognises termination codons. In order to explain the observation of nonsense codons having an effect on the nuclear RNA, it has been assumed that the recognition of the premature termination codon happens on the cytoplasmic side of the nuclear envelope, during mRNA export (Maquat 1995; Hentze and Kulozik 1999). However, it seems improbable that the pre-mRNA can be scanned by the cytoplasmic ribosomes before 3' processing, especially considering the emerging view

that RNA 3' end processing is probably a co-transcriptional event (Dantonel et al. 1997; McCracken et al. 1997a; McCracken et al. 1997b; Birse et al. 1998; Hirose and Manley 1998). In addition, the *Adh* pre-mRNA contains a 654 nucleotide long intron in the 5' untranslated region. Unless one assumes that the pre-mRNA is first partially exported to the cytoplasm, then checked for eventual nonsense mutations and finally re-imported into the nucleus for pre-mRNA processing, the "direct" involvement of cytoplasmic translation does not seem to be the strongest possibility. Therefore, the data presented appear to give credit to the hypothesis that there is a nuclear scanning mechanism. What this would be in molecular terms has been the subject of several speculations, but there is still no direct evidence for any of them (Dietz and Kendzior 1994; Aoufouchi et al. 1996; Li et al. 1997b; Li et al. 1997a).

Based on the observation that nonsense mutations affect 3' end formation one could speculate that, despite the question remaining of how this nuclear scanning is achieved, it may be required , in vivo, for correct pre-mRNA 3' end formation. As mentioned above, 3'-end formation is a very early step in pre-mRNA processing intrinsically linked to transcription termination, even in vitro (Yonaha and Proudfoot 1999), and therefore, this putative scanning is likely to occur on the nascent transcript (a graphic description of this speculative model is shown in figure 20). The fact that nonsense mutations cause an accumulation of unspliced transcript may be a

Figure 20



A cartoon of the interactions between the 3'-end pre-mRNA machinery, transcription, capping, splicing and the putative "nuclear scanner". Note that the transcript is still attached to the chromosome. The yellow box represents the 3'-end pre-mRNA machinery, the blue stars the splicing apparatuses (the loops represent introns). The red arrows indicate known interactions with capping and splicing, the green lines indicate the interaction with the proposed nuclear scanning machinery (indicated by a two subunit ribosome like structure). The question mark indicates that there is still no experimental evidence for these interactions.

secondary consequence of inhibiting the termination/cleavage/polyadenylation machinery. This later aspect of the model is supported by several indications that splicing and polyadenylation, as well as pre-mRNA processing events, can influence each other. For example, recent data indicate that the polyadenylation machinery stimulates splicing of the upstream intron through a direct interaction of the poly(A) polymerase with U2AF 65, which in turn enhance the binding of U2AF 65 to the pyrimidine tract of the adjacent 3' splice site (Vagner et al. 2000).

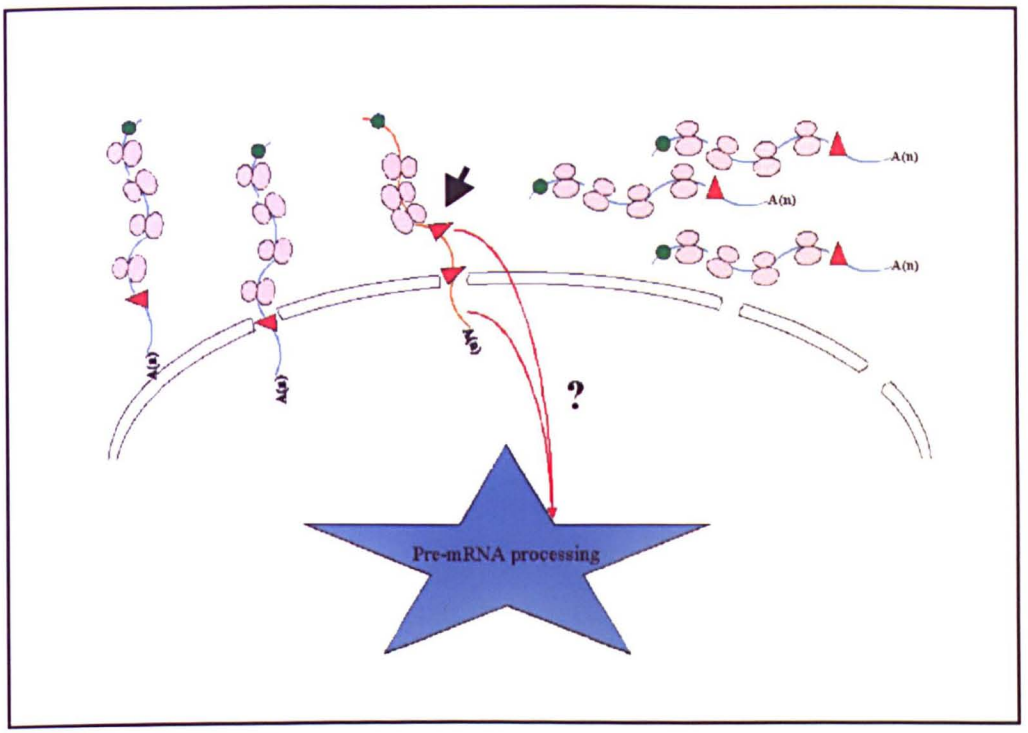
One of the major complications of scanning the ORF of an unspliced transcript, is of course represented by the fact that introns in the coding region would not allow a continuous scanning, like that happening during translation for example. Theoretically, one possibility would be to first define the intron and then juxtapose the two flanking exons; the intron would still be covalently bound to the rest of the transcript but kept in a distinct structural domain (a cartoon of such a speculation is shown in figure 20).

One alternative model is based on the emerging view that splicing may more often occur before the release of the transcript from the chromosome. This is supported for example by the report that mutations in both splicing and polyadenylation signals prevent the release of the transcript from the locus (Custodio et al. 1999). This second model proposes that splicing occurs while the transcript is still associated with

polyadenylation machinery and that scanning of the ORF may be required to complete 3'-end formation and to release the transcript from the chromosome. In this model the apparent increase in the level of unspliced transcript could be due to in trans splicing inhibition of the neighboring nascent transcripts by the "stalled" mRNP not yet released from the chromosome. The first prediction from this model would be that nonsense mutations cause an accumulation of nascent transcripts at the site of transcription. Interestingly, in agreement with this prediction, it has been observed that nonsense mutations in both the IgM and TCR genes cause an accumulation of pre-mRNA at the site of transcription (Oliver Muhlemann and Melissa Moore, Brandeis University, personal communication).

Of course all three of these models are highly speculative and, more importantly involve a mysterious nuclear scanning machinery. Therefore, alternatively and more conventionally, one could assume that the abortion of translation of the *Adh* mRNA has an effect in trans on nuclear pre-mRNA 3' end processing, implying that there is an *Adh* specific RNA binding protein that shuttles between the nucleus and cytoplasm, monitoring the translation of the *Adh* mRNA (a cartoon of this model is in figure 21). However, this later model it is also highly improbable because, since I have shown that the effects of nonsense mutations appear to be allele specific, the mechanism must be able to distinguish between two mRNAs differing in a single nucleotide.

Figure 21



A cartoon of the proposed feed-back mechanism between translation in the cytoplasm and pre-mRNA processing in the nucleus. According to this model the 5' end of the mRNA associates with ribosomes (the pink circles) as soon as it reaches the cytoplasm. The mRNA is exported as an RNP complex, which is remodelled during the first round of translation. If translation is prevented by a premature stop codon (indicated by the red triangle with the arrow) then the mRNP is not properly remodelled and a protein which is probably bound to the RNP, signals back to the pre-mRNA processing apparatus that the mRNA can not be translated. The question mark indicates that there is still no experimental evidence for the interaction.

In conclusion, from this work it is apparent that nonsense mutations can affect pre-mRNA processing; however future work and possibly new experimental approaches will be needed to distinguish between these three equally speculative hypotheses.

Does the mRNA surveillance system exists ?

Regardless of whether the mRNA reduction occurs in the cytoplasm or in the nucleus and regardless of whether the target is the mRNA or the pre-mRNA, the phenomenon of nonsense-mediated mRNA reduction is commonly viewed as an indication of the existence of a specialized cellular mechanism. This putative mechanism is regarded as a kind of "mRNA surveillance" system, that has evolved to rid the cell of aberrant mRNAs that could potentially produce deleterious truncated peptides.

This system is thought to require several evolutionary conserved proteins. Here I have reported that the sequence of the *Drosophila* homolog of UPF1 is very similar to that of the corresponding proteins found in other organisms; these data further support the view that UPF1 constitutes a family of highly conserved proteins. Since Upf1p is considered to be a major component of the mRNA surveillance complex in yeast, the fact that this protein is highly conserved in other organisms might suggest that the mRNA surveillance system is evolutionary

conserved, and that in higher organisms it should be similar to that of yeast.

However, despite its presupposed ancient evolution, this putative "mRNA surveillance system" is still not very effective in what is thought to be its main function: Preventing the synthesis of truncated polypeptides with a dominant negative function.

For instance, this putative mRNA surveillance system would appear to be effective in reducing the level of mRNAs carrying nonsense mutations in the first half of the coding region, but would not be effective in reducing the level of mRNAs with mutations located in the second half of the coding region. In other words, this putative "mRNA surveillance system" would be effective only in preventing the synthesis of polypeptides significantly shorter than the wild type protein.

However, it is expected that the longer is the truncated polypeptide the higher is the probability of it having a dominant negative function. Short polypeptides are probably intrinsically unstable and may have anyway a low probability of interacting with another factor - for example forming a defective heterodimer with the wild type allele.

Furthermore, it is curious to note that in *Adh* none of the nonsense mutations show any indication of phenotypic dominance, not even in the case of the most "C-terminal" mutations like Adh^{nB} and Adh^{nBR114}, which having only a minor reduction in the level of mRNA should be able to produce polypeptides only slightly shorter than the wild type (Schwartz

and Sofer 1976; Martin et al. 1985). The ADH enzyme functions as a dimer; homodimers and heterodimers can be distinguished by nondenaturing electrophoresis and previous work indicates that some missense mutations or small in-frame deletions can form defective heterodimers with the wild type subunit - that is to say that some mutant alleles of *Adh* are partially dominant (Jiang et al. 1992). Therefore it is a bit surprising that none of the "C-terminal" mutations, which have high levels of mRNA, show any indication of heterodimer formation. Interestingly, the reason for this appears to be that the mutant polypeptide, although efficiently synthesized is very unstable and degraded soon after being synthesized (Martin et al. 1985).

Therefore, based on the *Adh* system, it seems reasonable to suggest that what may determine whether or not a nonsense mutation has a dominant effect may, in many instances, depend of the stability of the truncated polypeptide rather than the mRNA level or the level of its synthesis. Prematurely terminated polypeptides are most probably incorrectly folded and rapidly degraded by the ubiquitin-proteasome pathway. Interestingly, very recent studies in HeLa cells indicate that more than 30% of newly synthesized proteins are defective, probably because of errors in translation or post-translational processes which prevent proper folding; these defective proteins are rapidly destroyed by the proteasomes (Schubert et al. 2000). The inactivation of the proteasome machinery with specific drugs leads to a rapid increase in the

cytosolic level of defective proteins, therefore it is conceivable that the proteasome machinery, by preventing the accumulation of defective polypeptides, would also prevent major abnormalities in the physiology of the cell (Johnston et al. 1998). One of the functions of the proteasome machinery could be to prevent the accumulation of partially folded proteins that could act in a dominant negative manner.

In conclusion, at least in *Drosophila*, it is premature to take the phenomenon of nonsense mediated mRNA reduction, NMMR, as proof of the existence of an mRNA surveillance mechanism. The data produced in this work instead suggest that nonsense mediated mRNA reduction may be the passive consequence of abnormal pre-mRNA processing. However, the mechanism by which this is achieved remains a mystery.

Chapter 5

MATERIAL AND METHODS

Alleles

The mutations *Adh* ^{nBR112} and *Adh* ^{nBR114} have been induced by N-ethyl-N-nitrosourea (ENU) mutagenesis and are molecularly characterised (Fossett et al. 1990). The allele *Adh* ^{nB} was induced by ethylmethane sulfonate (EMS) mutagenesis of a wild-type allele and has been molecularly characterised (Martin et al. 1985). *Adh* ^{n4} is EMS induced, on an *Adh* ^{D} chromosome. An *Adh* ^{n4} 2369 bp genomic fragment including flanking regions has been (Chia et al. 1987).

The alleles *Adh* ^{fn4} and *Adh* ^{fn6} were induced by formaldehyde on the functional allele *Adh* ^{D}, both affecting intron 2 (Benyajati et al. 1982). *Adh* ^{fn4} has a 17bp deletion and one substitution. The site of the deletion is between the conserved A and G at the 3' end of the intron. This creates a GG 3' end. *Adh* ^{fn6} has a 6 bp deletion and a 5 bp substitution, one break is between the 5' GT of the intron and creates a GC instead.

Adh ^{nAC14} is a naturally occurring null allele that has eight extra nucleotides near the 5' splice site and a nonsense mutation in exon 2 (Freeth et al. 1990) . The wild-type strain used is Canton-S.

Nuclei isolation.

The following method is a modification of that described by (Shaffer et al. 1994). About 300 flies were left in a clean empty tube for 2-3 hours to reduce contaminations. They were then homogenised in a Dounce glass homogenizer in 7 ml of buffer A supplemented with 1mM DTT and 10 mM Vanidyl Ribonucleoside Complex (Gibco-BRL), using a loose pestle. Subsequently, 0.35 ml of a solution of 10% Triton and 5% Deoxicolate were added and homogenisation continued with a tight pestle. The homogenate was spun for 1 min at 1000 rpm in a swing-out rotor, the supernatant then filtered through a layer of Miracloth (Calbiochem). The supernatant was then over-layed on a 7 ml 1.6M sucrose cushion (buffer A-1.6) and centrifuged in a SW28 rotor at 12K rpm for 20 min. The purity of the nuclear fraction was checked by phase-contrast microscopy.

RNA extraction and Northern blots

Total and nuclear RNA preparations were done according to the protocol described in (Ashburner 1989). Alternatively small preps were purified with the either RNeasy kit (Qiagen) or SV Total RNA isolation System (Promega. When required the DNA was removed by digestion with RNase free DNase (Promega).

The RNA was fractionated on either ordinary agarose gels or on high resolution 3-4% Nusieve 3:1 agarose (FMC) in the presence of formaldehyde. For overnight runs, formaldehyde was also added to the running buffer.

The RNA was transferred by over-night capillary transfer onto a nylon membrane (Hybond-N, Amersham) and hybridised as described (Yang et al., 1993). All other manipulations are essentially as described in Sambrook et al. (1989).

RNase H digestion

The hybridisation and RNase H digestion was essentially as described (Brogna 1999). 5-10 µg of total RNA was hybridised to the oligodeoxynucleotide(s) in 40 µl hybridisation solution supplemented with 40 units of RNasin (Promega) for 30 min at RT. Then 40 µl of digestion buffer containing 1 unit of RNase H (Promega) was added and the mixture was incubated for 30 min at 37 0C. Then, the samples were extracted with 50:50 phenol/chloroform, ethanol precipitated and resuspended in H_2O. The RNA was then analysed by Northern-blot.
In the control experiment with added oligo-d(T)16, we found digestion in the *Adh* 3' UTR, which has a stretch of 11 As. We could partially prevent the cleavage of this stretch of A' by not denaturing the RNA before hybridisation with the oligodeoxynucleotides.

Circular-RT-PCR

This procedure is a modification of a previously published procedure (Couttet et al. 1997). Total-RNA was purified with one of the methods described above. Between 1 and 50 micrograms of total RNA were de-capped using 2.5 units of Tobacco Acid Pyrophosphatase (TAP, Epicentre Technologies) in 20 microliters of reaction buffer [50mM sodium acetate (pH 6) 1mM EDTA, 0.1 % 2-mercaptoethanol, 0.01 % Triton X-100 and 40 units of RNasin (Promega). The reactions were incubated for 2 hours at 37°, diluted to 200 microliters with TE (pH 8) and then the RNA was purified by phenol/chloroform and chloroform extractions. The RNA was precipitated with ethanol - Pellet Paint (Novagen) was used as a carrier, which was added before the organic extraction.

The circularisation of the mRNA was achieved by incubating the de-capped RNA from above with 40 units of T4 RNA ligase (Promega or New England Biolabs) in 200 microliters of ligation buffer [50mM Tris-HCl (pH 7,8), 10mM MgCl2, 10mM 2-mercaptoethanol, 1 mM ATP and 80 units of RNasin (Promega)]. The ligations were incubated over night at 18 C°. The RNA was purified as above, precipitate and resuspended in 5 microliters of water. Denaturing the RNA for 10 minutes by heating at 70 C° prior the ligation was omitted because did not improve the efficiency of the reaction (Couttet et al. 1997).

The cDNA was synthesised with SuperScript II (GIBCO-BRL) following the instructions from the manufacturer. Standard PCR conditions were used and are described in (Brogna 1999). The sequence of the primers is given in the figure legends. The amplified products were analysed on 3-4% Nusieve 3:1 agarose gels. The sequence of these products was determined by either directly sequencing re-amplified gel purified fragments or after sub-cloning in pGEM-T (Promega). Sequencing was performed with the DyeDeoxy Terminator Cycle kit (ABI).

Characterisation of *DmUpf1*.

ESTs (Expressed Sequence Tags) for the *DmUpf1* gene were identified by querying the Drosophila EST database with the yeast (P30771) and human (Q92900) UPF1 sequences, using the TBLASTX program http://www.fruitfly.org/EST/. Two of these EST clones (GH09524 and LD25787, Research Genetics) were sequenced from the 5' and 3' ends using the PM001 and T7 primers respectively. The GH and LD cDNA libraries derive from adult heads and embryos mRNA respectively, the cloning vector is pOT2 (http://www.fruitfly.org/EST/pOT2vector.html).

The 5' end of the cDNA was amplified from the GH, LD and an imaginal disk library (provided by Nick Brown; Wellcome-CRC, Cambridge). In the GH and LD libraries the 5' and 3' ends of the cDNA

were amplified by nested PCR with the T7 primer and Upf1.Rev3 - 5'.CGATTCAAATCCGGCAGAT.3' - in the first round, and Upf1.Rev2, primer A in figure 14 - 5'.CACTGAAGAGCTTGGGTTGT.3' - in the second round. For the imaginal disk library two primers flanking the 5' and 3' ends were used : NB5.fow - 5'.CACTATAGAATACAAGCTTGCT.3' or NB852 - 5'.GAATACAAGCTTGCTTGTTC.3' for the 5' end and NB3.3198 - 5'.TTAATGCAGCTGGCTTATCG.3' for the 3' end.

A full length cDNA, referred as pDUpf1.wt, was constructed by ligating to the GH09524 cDNA the missing 5' end fragment, PCR amplified from the imaginal disk library with Pfu polymerases and specific primers. The primers used were: Upf1.5endFow - GGGAGATCTCGTGAGTCAAATGCAAACT -plus Upf1.Rev3 (see above).The Upf1.5endFow primer corresponds to the 5' end of the cDNA, from position 8 to 28. The fragment was digested with *Bgl* II (which was introduced in Upf1.5endFow) and *Pvu* I (which cut in the insert up stream of Upf1.Rev3) and cloned into the same sites in pGH09524. The fragment was completely sequenced.

The DE617AA mutant cDNA, called pDUpf1.mut, was created by first PCR amplifying the correspondent region from with Upf1.Mut2Rev - 5'.CGTCGACTGCATAGACgCAgCGATCAGGATCG.3' - which contains two substitutions relative to the wild-type, the codons mutated

are underlined, and Upf1.1660Fow - CGCCATCGATCGCAATTC -. The fragment was checked by sequencing. The mutated fragment was then substituted for the wild-type fragment by restriction enzymes digestion with *Pvu* I and *Sal* I and ligation into the same sites in pUpf1.wt.

Transformation constructs and transgenes

The wild-type (pDUpf1.wt) and the DE617AA (pDUpf1.mut) cDNAs were cut out with *Xho* I and *Bgl* II and ligated into the same sites in pP(UAST) - (Brand and Perrimon 1993). The constructs were integrated into the genome by transforming a *yellow white* strain by standard techniques using a helper plasmid as a source of transposase (Ashburner 1989).

Insertion sites were mapped to the chromosomes by crossing with marked balancer chromosomes.

BIBLIOGRAPHY

Adams, M.D., S.E. Celniker, and e. al. 2000. The genome sequence of Drosophila melanogaster. *Science* **287**: 2185-2195.

Altamura, N., O. Groudinsky, G. Dujardin, and P.P. Slonimski. 1992. Nam7 nuclear gene encodes a novel member of a family of helicases with a Zn-ligand motif and is involved in mitochondrial functions in *Saccharomyces cerevisiae*. *Journal of Molecular Biology* **224**: 575-587.

Aoufouchi, S., J. Yelamos, and C. Milstein. 1996. Nonsense mutations inhibit RNA splicing in a cell-free system - recognition of mutant codon is independent of protein synthesis. *Cell* **85**: 415-422.

Applequist, S.E., M. Selg, C. Raman, and H.M. Jack. 1997. Cloning and characterization of HUPF1, a human homolog of the Saccharomyces cerevisiae nonsense mRNA-reducing UPF1 protein. *Nucleic Acids Research* **25**: 814-821.

Ashburner, M. 1989. Drosophila: a laboratory handbook. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York.

Ashburner, M. 1998. Speculations on the subject of alcohol dehydrogenase and its properties in Drosophila and other flies. *Bioessays* **20**: 949-954.

Atkinson, J. and R. Martin. 1994. Mutations to nonsense codons in human genetic disease implications for gene therapy by nonsense suppressor transfer-RNAs. *Nucleic Acids Research* **22**: 1327-1334.

Baserga, S.J. and E.J. Benz. 1992. Beta-globin nonsense mutation deficient accumulation of messenger RNA occurs despite normal cytoplasmic stability. *Proc. Natl. Acad. Sci. USA* **89**: 2935-2939.

Beelman, C.A. and R. Parker. 1995. Degradation of messenger RNA in eukaryotes. *Cell* **81**: 179-183.

Belgrader, P., J. Cheng, and L.E. Maquat. 1993. Evidence to implicate translation by ribosomes in the mechanism by which nonsense codons reduce the nuclear level of human triosephosphate isomerase messenger RNA. *Proc. Natl. Acad. Sci. USA* **90**: 482-486.

Belgrader, P., J. Cheng, X.B. Zhou, L.S. Stephenson, and L.E. Maquat. 1994. Mammalian nonsense codons can be cis effectors of nuclear messenger RNA half life. *Molecular and Cellular Biology* **14**: 8219-8228.

Belgrader, P. and L.E. Maquat. 1994. Nonsense but not missense mutations can decrease the abundance of nuclear messenger RNA for the mouse major urinary protein, while both types of mutations can facilitate exon skipping. *Molecular and Cellular Biology* **14**: 6326-6336.

Benoit, B., A. Nemeth, N. Aulner, U. Kuhn, M. Simonelig, E. Wahle, and H. Bourbon. 1999. The Drosophila poly(A)-binding protein II is ubiquitous throughout Drosophila development and has the same function in mRNA polyadenylation as its bovine homolog in vitro. *Nucleic Acids Res 1999* **27**: 3771-3778.

Benyajati, C., A.R. Place, N. Wang, E. Pentz, and W. Sofer. 1982. Deletions at intervening sequence splice sites in the alcohol dehydrogenase gene of Drosophila. *Nucleic Acids Res* **10**: 7261-7272.

Benyajati, C., N. Spoerel, H. Haymerle, and M. Ashburner. 1983. The messenger RNA for alcohol dehydrogenase in Drosophila melanogaster differs in its 5' end in different developmental stages. *Cell* **33**: 125-133.

Benzer, S. and S.P. Champe. 1962. A change from nonsense to sense in the genetic code. *Proc. Natl. Acad. Sci. USA* **48**: 1114-1121.

Birse, C.E., L. Minvielle-Sebastia, B.A. Lee, W. Keller, and N.J. Proudfoot. 1998. Coupling termination of transcription to messenger RNA maturation in yeast. *Science* **280**: 298-301.

Bouvet, P. and J.G. Belasco. 1992. Control of Rnase E mediated RNA degradation by 5'-Terminal base pairing in *Escherichia coli. Nature* **360**: 488-491.

Brand, A.H. and N. Perrimon. 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**: 401-415.

Braun, F., J. LeDerout, and P. Regnier. 1998. Ribosomes inhibit an RNase E cleavage which induces the decay of the rpsO mRNA of Escherichia coli. *EMBO Journal* **17**: 4790-4797.

Brenner, S., L. Barnett, E.R. Katz, and F.H. Crick. 1967. UGA: a third nonsense triplet in the genetic code. *Nature* **213**: 449-450.

Brogna, S. 1999. Nonsense mutations in the alcohol dehydrogenase gene of Drosophila melanogaster correlate with an abnormal 3' end processing of the corresponding pre-mRNA. *RNA* **5**: 562-573.

Brogna, S. and M. Ashburner. 1997. The Adh-related gene of Drosophila melanogaster is expressed as a functional dicistronic messenger RNA: Multigenic transcription in higher organisms. *EMBO Journal* **16**: 2023-2031.

Caceres, J.F., G.R. Screaton, and A.R. Krainer. 1998. A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. *Genes & Development* **12**: 55-66.

Cali, B.M. and P. Anderson. 1998. mRNA surveillance mitigates genetic dominance in Caenorhabditis elegans. *Molecular and General Genetics* **260**: 176-184.

Cali, B.M., S.L. Kuchma, J. Latham, and P. Anderson. 1999. smg-7 is required for mRNA surveillance in Caenorhabditis elegans. *Genetics* **151**: 605-616.

Carter, M.S., S.L. Li, and M.F. Wilkinson. 1996. A splicing dependent regulatory mechanism that detects translation signals. *EMBO Journal* **15**: 5965-5975.

Castilho-Valavicius, B., H. Yoon, and T.F. Donahue. 1990. Genetic characterization of the *Saccharomyces Cerevisiae* translational initiation suppressors Sui1, Sui2 and Sui3 and their effects on His4 expression. *Genetics* **124**: 483-495.

Chambers, G.K. 1988. The Drosophila alcohol dehydrogenase gene-enzyme system. In *Advance in Genetics* (ed. E.W. Caspari and S. J.G.), pp. 39-107. Academic Press, Inc., London.

Chan, H.Y.E., Y. Zhang, J. Hoheisel, and C.J. O'Kane. 1998. Identification and characterization of the gene for Drosophila L3 ribosomal protein. *Gene* **212**: 119-125.

Chang, J.C. and Y.W. Kan. 1979. β^0 thalassemia, a nonsense mutation in man. *Proc Natl Acad Sci U S A* **76**: 2886-2889.

Cheng, J. and L.E. Maquat. 1993. Nonsense condons can reduce the abundance of nuclear messenger RNA without affecting the abundance of premessenger RNA or the half-life of cytoplasmic messenger RNA. *Molecular and Cellular Biology* **13**: 1892-1902.

Chia, W., C. Savakis, R. Karp, and M. Ashburner. 1987. Adh-n4 of *Drosophila melanogaster* is a nonsense mutation. *Nucleic Acids Res* **15**: 3931.

Chia, W., C. Savakis, R. Karp, H. Pelham, and M. Ashburner. 1985. Mutation of the Adh gene of Drosophila melanogaster containing an internal tandem duplication. *J Mol Biol* **186**: 679-688.

Colgan, D.F. and J.L. Manley. 1997. Mechanism and regulation of mRNA polyadenylation. *Genes & Development* **11**: 2755-2766.

Couttet, P., M. FromontRacine, D. Steel, R. Pictet, and T. Grange. 1997. Messenger RNA deadenylylation precedes decapping in mammalian cells. *Proc. Natl. Acad. Sci. USA* **94**: 5628-5633.

Cui, Y., C. I. Gonzalez, T. G. Kinzy, J. D. Dinman and S. W. Peltz, 1999. Mutations in the MOF2/SUI1 gene affect both translation and nonsense-mediated mRNA decay. RNA **5:** 794-804.

Cui, Y., J.D. Dinman, and S.W. Peltz. 1996. mof4-1 is an allele of the UPF1/IFS2 gene which affects both mRNA turnover and -1 ribosomal frameshifting efficiency. *EMBO Journal* **15**: 5726-5736.

Cui, Y., K.W. Hagan, S.A. Zhang, and S.W. Peltz. 1995. Identification and characterization of genes that are required for the accelerated degradation of messenger RNAs containing a premature translational termination codon. *Genes & Development* **9**: 423-436.

Culbertson, M.R., K.M. Underbrink, and G.R. Fink. 1980. Frameshift suppression in Saccharomyces Cerevisiae II. Genetic properties of group II suppressors. *Genetics* **95**: 833-853.

Custodio, N., M. Carmo-Fonseca, F. Geraghty, H.S. Pereira, F. Grosveld, and M. Antoniou. 1999. Inefficient processing impairs release of RNA from the site of transcription. *EMBO J* **18**: 2855-2866.

Czaplinski, K., M.J. RuisEchevarria, C.I. Gonzalez, and S.M. Peltz. 1999. Should we kill the messenger? The role of the surveillance complex in translation termination and mRNA turnover. *Bioessays* **21**: 685-696.

Czaplinski, K., M.J. Ruiz-Echevarria, S.V. Paushkin, X. Han, Y.M. Weng, H.A. Perlick, H.C. Dietz, M.D. TerAvanesyan, and S.W. Peltz. 1998. The surveillance complex interacts with the translation release factors to enhance termination and degrade aberrant mRNAs. *Genes & Development* **12**: 1665-1677.

Czaplinski, K., Y. Weng, K.W. Hagan, and S.W. Peltz. 1995. Purification and characterization of the Upf1 protein - a factor involved in translation and messenger RNA degradation. *RNA* **1**: 610-623.

Dantonel, J.C., K.G. Murthy, J.L. Manley, and L. Tora. 1997. Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA. *Nature* **389)**: 399-402.

Das, A., D. Court, and S. Adhya. 1976. Isolation and characterization of conditional lethal mutants of Escherichia coli defective in transcription termination factor rho. *Proc Natl Acad Sci U S A* **73**: 1959-1963.

Das, B., Z. Guo, P. Russo, P. Chartrand, and F. Sherman. 2000. The role of nuclear cap binding protein cbc1p of yeast in mRNA termination and degradation. *Mol Cell Biol* **20**: 2827-2838.

Decker, C.J. and R. Parker. 1993. A turnover pathway for both stable and unstable messenger RNAs in yeast - evidence for a requirement for deadenylation. *Genes & Development* **7**: 1632-1643.

Decker, C.J. and R. Parker. 1994. Mechanisms of messenger RNA degradation in eukaryotes. *Trends In Biochemical Sciences* **19**: 336-340.

Dietz, H.C. and R.J. Kendzior. 1994. Maintenance of an open reading frame as an additional level of scrutiny during splice-site selection. *Nature Genetics* **8**: 183-188.

Dietz, H.C., D. Valle, C.A. Francomano, R.J. Kendzior, R.E. Pyeritz, and G.R. Cutting. 1993. The skipping of constitutive exons in vivo induced by nonsense mutations. *Science* **259**: 680-683.

Dinman, J.D. and R.B. Wickner. 1994. Translational maintenance of frame-mutants of Saccharomyces cerevisiae with altered -1 ribosomal frameshifting efficiencies. *Genetics* **136**: 75-86.

Fossett, N.G., P. Arbourreily, G. Kilroy, M. McDaniel, J. Mahmoud, A.B. Tucker, S.H. Chang, and W.R. Lee. 1990. Analysis of ENU-Induced mutations at the Adh locus in *Drosophila Melanogaster. Mutation Research* **231**: 73-85.

Franklin, N.C. and S.E. Luria. 1961. Transduction by the bacteriophage P1 and the properties of the lac genetic region of E. coli and S. dysenteriae. *Virology* **15**: 299-311.

Freeth, A.L., J.B. Gibson, and A.V. Wilks. 1990. Aberrant splicing of a naturally occurring alcohol dehydrogenase null activity allele in Drosophila melanogaster. *Genome* **33**: 873-877.

Garen, A. and O. Siddiqi. 1962. Suppression of mutations in the alkaline phosphatase structural cistron of E. coli. *Proc. Natl. Acad. Sci.-USA* **48**.

Gersappe, A. and D.J. Pintel. 1999. A premature termination codon interferes with the nuclear function of an exon splicing enhancer in an open reading frame-dependent manner. *Molecular and Cellular Biology* **19**: 1640-1650.

Gonzalez, C.I., M.J. Ruiz-Echevarria, S. Vasudevan, M.F. Henry, and S.W. Peltz. 2000. The yeast hnRNP-like protein Hrp1/Nab4 marks a transcript for nonsense-mediated mRNA decay. *Mol Cell* **5**: 489-499.

Hagan, K.W., M.J. Ruiz-Echevarria, Y. Quan, and S.W. Peltz. 1995. Characterization of cis-acting sequences and decay intermediates involved in nonsense-mediated messenger RNA turnover. *Molecular and Cellular Biology* **15**: 809-823.

Hampsey, M., J.G. Na, I. Pinto, D.E. Ware, and R.W. Berroteran. 1991. Extragenic suppressors of a translation initiation defect in the Cyc1 gene of *Saccharomyces Cerevisiae. Biochimie* **73**: 1445-1455.

Hawkins, J.D. 1988. A survey on intron and exon lengths. *Nucleic Acids Research* **16**: 9893-9908.

He, F., A.H. Brown, and A. Jacobson. 1997. Upf1p, Nmd2p, and Upf3p are interacting components of the yeast nonsense-mediated mRNA decay pathway. *Molecular and Cellular Biology* **17**: 1580-1594.

He, F. and A. Jacobson. 1995. Identification of a novel component of the nonsense-mediated messenger RNA decay pathway by use of an interacting protein screen. *Genes & Development* **9**: 437-454.

Hentze, M.W. and A.E. Kulozik. 1999. A perfect message: RNA surveillance and nonsense-mediated decay. *Cell* **96**: 307-310.

Hilleren, P. and R. Parker. 1999. mRNA surveillance in eukaryotes: kinetic proofreading of proper translation termination as assessed by mRNP domain organization? *RNA* **5**: 711-719.

Hirose, Y. and J.L. Manley. 1998. RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* **395**: 93-96.

Hodgkin, J., A. Papp, R. Pulak, V. Ambros, and P. Anderson. 1989. A new kind of informational suppression in the nematode Caenorhabditis elegans. *Genetics* **123**: 301-313.

Hodgman, T.C. 1988. A New superfamily of replicative proteins. *Nature* **333**: 22-23.

Iost, I. and M. Dreyfus. 1995. The stability of *Escherichia coli* lacZ messenger RNA depends upon the simultaneity of its synthesis and translation. *EMBO Journal* **14**: 3252-3261.

Jacob, F. and J. Monod. 1961. On the regulation of gene activity. *Cold Spring Harbor Symp. Quant. Biol.* **26**: 193-211.

Jacobson, A. and S.W. Peltz. 1996. Interrelationships of the pathways of mRNA decay and translation in eukaryotic cells. *Annual Review of Biochemistry* **65**: 693-739.

Jiang, J.C., W.R. Lee, S.H. Chang, and H. Silverman. 1992. Mechanisms for dominance: Adh heterodimer formation in heterozygotes between ENU or X-ray induced null alleles and normal alleles in Drosophila melanogaster. *Environ Mol Mutagen* **20**: 260-270.

Johnston, J.A., C.L. Ward, and R.R. Kopito. 1998. Aggresomes: a cellular response to misfolded proteins. *J Cell Biol* **143**: 1883-1898.

Kessler, M.M., M.F. Henry, E. Shen, J. Zhao, S. Gross, P.A. Silver, and C.L. Moore. 1997. Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3'-end formation in yeast. *Genes Dev* **11**: 2545-2556.

Koonin, E.V. 1992. A new group of putative RNA helicases. *Trends in Biochemical Sciences* **17**: 495-497.

Kugler, W., J. Enssle, M.W. Hentze, and A.E. Kulozik. 1995. Nuclear degradation of nonsense mutated beta-globin messenger RNA - a posttranscriptional mechanism to protect heterozygotes from severe clinical manifestations of beta-thalassemia. *Nucleic Acids Research* **23**: 413-418.

Lee, B.S. and M.R. Culbertson. 1995. Identification of an additional gene required for eukaryotic nonsense messenger RNA turnover. *Proc. Natl. Acad. Sci. USA* **92**: 10354-10358.

Lee, S.I., J.G. Umen, and H.E. Varmus. 1995. A genetic screen identifies cellular factors involved in retroviral - 1 frameshifting. *Proc. Natl. Acad. Sci. USA* **92**: 6587-6591.

Leeds, P., S.W. Peltz, A. Jacobson, and M.R. Culbertson. 1991. The product of the yeast Upf1 gene is required for rapid turnover of messenger RNAs containing a premature translational termination codon. *Genes & Development* **5**: 2303-2314.

Leeds, P., J.M. Wood, B.S. Lee, and M.R. Culbertson. 1992. Gene products that promote mRNA turnover in Saccharomyces cerevisiae. *Mol Cell Biol* **12**: 2165-2177.

Le Hir, H., M.J. Moore, and L.E. Maquat. 2000. Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exox-exon junctions. *Genes & Development* **14**;1098-1108.

Li, S.L., D. Leonard, and M.F. Wilkinson. 1997a. T cell receptor (TCR) mini-gene mRNA expression regulated by nonsense codons: A nuclear-associated translation-like mechanism. *Journal Of Experimental Medicine* **185**: 985-992.

Li, S.L., D. Leonard, and M.F. Wilkinson. 1997b. T-cell receptor (TCR) mini-gene mRNA expression regulated by nonsense codons: A nuclear-associated translation-like mechanism (vol 185, pg 985, 1997). *Journal Of Experimental Medicine* **185**: 1883-1883.

Lopez, P.J., I. Marchand, O. Yarchuk, and M. Dreyfus. 1998. Translation inhibitors stabilize Escherichia coli mRNAs independently of ribosome protection. *Proc. Natl. Acad. Sci. USA* **95**: 6067-6072.

Losson, R. and F. Lacroute. 1979. Interference of nonsense mutations with eukaryotic messanger RNA stability. *Proc. Natl. Acad. Sci. USA* **76**: 5134-5137.

Lozano, F., B. Maertzdorf, R. Pannell, and C. Milstein. 1994. Low cytoplasmic messenger RNA levels of immunoglobulin-kappa light chain genes containing nonsense codons correlate with inefficient splicing. *EMBO Journal* **13**: 4617-4622.

Mackie, G.A. 1998. Ribonuclease E is a 5'-end-dependent endonuclease. *Nature* **395**: 720-723.

Maquat, L.E. 1995. When cells stop making sense - Effects of nonsense codons on RNA metabolism in vertebrate cells. *RNA* **1**: 453-465.

Martin, P.F., A.R. Place, E. Pentz, and W. Sofer. 1985. UGA nonsense mutation in the alcohol dehydrogenase gene of Drosophila melanogaster. *J Mol Biol* **184**: 221-229.

McCormick, J.R., J.M. Zengel, and L. Lindahl. 1994. Correlation of translation efficiency with the decay of lacZ messenger RNA in *Escherichia coli*. *Journal of Molecular Biology* **239**: 608-622.

McCracken, S., N. Fong, E. Rosonina, K. Yankulov, G. Brothers, D. Siderovski, A. Hessel, S. Poster, S. Shuman, and D.L. Bentley. 1997a. 5'-capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes & Development* **11**: 3306-3318.

McCracken, S., N. Fong, K. Yankulov, S. Ballantyne, G.H. Pan, J. Greenblatt, S.D. Patterson, M. Wickens, and D.L. Bentley. 1997b. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**: 357-361.

Mitchelson, A., M. Simonelig, C. Williams, and K. O'Hare. 1993. Homology with Saccharomyces cerevisiae RNA14 suggests that phenotypic suppression in Drosophila melanogaster by suppressor of forked occurs at the level of RNA stability. *Genes Dev* **7**: 241-249.

Morrison, M., K.S. Harris, and M.B. Roth. 1997. smg mutants affect the expression of alternatively spliced SR protein mRNAs in Caenorhabditis elegans. *Proc. Natl. Acad. Sci. USA* **94**: 9782-9785.

Morse, D.E. and C. Yanofsky. 1969. Polarity and the degradation of mRNA. *Nature* **224**: 329-331.

Muhlrad, D., C.J. Decker, and R. Parker. 1995. Turnover mechanisms of the stable yeast Pgk1 messenger RNA. *Molecular and Cellular Biology* **15**: 2145-2156.

Muhlrad, D. and R. Parker. 1994. Premature translational termination triggers messenger RNA decapping. *Nature* **370**: 578-581.

Muhlrad, D. and R. Parker. 1999. Aberrant mRNAs with extended 3' UTRs are substrates for rapid degradation by mRNA surveillance. *RNA* **5**: 1299-1307.

Nagy, E. and L.E. Maquat. 1998. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends in Biochemical Sciences* **23**: 198-199.

Nakielny, S. and G. Dreyfuss. 1997. Nuclear export of proteins and RNAs. *Curr. Opin. Cell Biol.* **9**:420-429.

Newton, W.A., J.R. Beckwith, D. Zipser, and S. Brenner. 1965. Nonsense mutants and polarity in the lac operon of Escherichia coli. *J Mol Biol* **14**: 290-296.

Nilsson, G., J.G. Belasco, S.N. Cohen, and A. Vongabain. 1987. Effect of premature termination of translation on messenger RNA stability depends on the site of ribosome release. *Proc. Natl. Acad. Sci. USA* **84**: 4890-4894.

O'Connell, P.O. and M. Rosbash. 1984. Sequence, structure, and codon preference of the Drosophila ribosomal protein 49 gene. *Nucleic Acids Res* **12**: 5495-5513.

Oliveira, C.C. and J.E.G. McCarthy. 1995. The relationship between eukaryotic translation and messenger RNA stability - a short upstream open reading frame strongly inhibits translational initiation and greatly accelerates messenger RNA degradation in the yeast *Saccharomyces cerevisiae*. *Journal of Biological Chemistry* **270**: 8936-8943.

Page, M.F., B. Carr, K.R. Anders, A. Grimson, and P. Anderson. 1999. SMG-2 is a phosphorylated protein required for mRNA surveillance in Caenorhabditis elegans and related to Upf1p of yeast. *Molecular and Cellular Biology* **19**: 5943-5951.

Peltz, S.W., A.H. Brown, and A. Jacobson. 1993. Messenger RNA destabilization triggered by premature translational termination depends on at least 3 cis-acting sequence elements and one trans-acting factor. *Genes & Development* **7**: 1737-1754.

Perlick, H.A., S.M. Medghalchi, F.A. Spencer, R.J. Kendzior, and H.C. Dietz. 1996. Mammalian orthologues of a yeast regulator of nonsense transcript stability. *Proc. Natl. Acad. Sci. USA* **93**: 10928-10932.

Pestova, T.V., S.I. Borukhov, and C.U.T. Hellen. 1998. Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons. *Nature* **394**: 854-859.

Pinto, I., J.G. Na, F. Sherman, and M. Hampsey. 1992. Cis-acting and trans-acting suppressors of a translation initiation defect at the Cyc1 locus of Saccharomyces Cerevisiae. *Genetics* **132**: 97-112.

Platt, T. and D.G. Bear. 1983. Role of RNA polymerase, rho factor, and ribosomes in transcription termination. In *Gene function in procaryotes.* (ed. J. Beckwith, J. Davies, and J. Gallant), pp. 123-161. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Proudfoot, N. 1991. Poly(A) signals. *Cell* **64**: 671-674.

Proudfoot, N. 1996. Ending the message is not so simple. *Cell* **87**: 779-781.

Pulak, R. and P. Anderson. 1993. Messenger RNA surveillance by the *Caenorhabditis elegans* smg genes. *Genes & Development* **7**: 1885-1897.

Ross, J. 1997. A hypothesis to explain why translation inhibitors stabilise mRNAs in mammalian cells: mRNA stability and mitosis. *BioEssays* **19**: 527-529.

Ruiz-Echevarria, M.J., K. Czaplinski, and S.W. Peltz. 1996. Making sense of nonsense in yeast. *Trends In Biochemical Sciences* **21**: 433-438.

Ruiz-Echevarria, M.J., C.I. Gonzalez, and S.W. Peltz. 1998. Identifying the right stop: determining how the surveillance complex recognizes and degrades an aberrant mRNA. *EMBO Journal* **17**: 575-589.

Ruiz-Echevarria, M.J. and S.W. Peltz. 1996. Utilizing the GCN4 leader region to investigate the role of the sequence determinants in nonsense-mediated messenger RNA decay. *EMBO Journal* **15**: 2810-2819.

Sarabhai, A.S., O.W. Stretton, and S. Brenner. 1964. Co-linearity of the gene with the polypeptide chain. *Nature* **201**: 13-17.

Schaeffer, S.W. and C.F. Aquadro. 1987. Nucleotide sequence of the Adh gene region of Drosophila pseudoobscura: evolutionary change and evidence for an ancient gene duplication. *Genetics* **117**: 61-73.

Schubert, U., L.C. Anton, J. Gibbs, C.C. Norbury, J.W. Yewdell, and J.R. Bennink. 2000. Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. *Nature* **404**: 770-774.

Schwartz, M. and W. Sofer. 1976. Alcohol dehydrogenase-negative mutants in Drosophila: Defects at the structural level. *Genetics* **83**: 125-136.

Semrad, K. and R. Schroeder. 1998. Ribosomal function is necessary for efficient splicing of the T4 phage thymidylate synthase intron in vivo. *Genes & Development* **12**: 1327-1337.

Shaffer, C.D., J.M. Wuller, and S.C.R. Elgin. 1994. Preparation of Drosophila nuclei. In *Drosophila melanogaster: practical uses in cell and molecular biology* (ed. L.S.B. Goldstein and E.A. Fyrberg), pp. 185-189. Academic Press, Inc., San Diego, California.

Stretton, A.O., S. Kaplan, and S. Brenner. 1966. Nonsense codons. *Cold Spring Harb Symp Quant Biol* **31**: 173-179.

Sun, X., H.A. Perlick, H.C. Dietz, and L.E. Maquat. 1998. A mutated human homologue to yeast Upf1 protein has a dominant-negative effect on the decay of nonsense-containing mRNAs in mammalian cells. *Proc Natl Acad Sci U S A* **95**: 10009-10014.

Takagaki, Y. and J.L. Manley. 1994. A polyadenylation factor subunit is the human homologue of the Drosophila suppressor of forked protein. *Nature* **372**: 471-474.

Thermann, R., G. NeuYilik, A. Deters, U. Frede, K. Wehr, C. Hagemeier, M.W. Hentze, and A.E. Kulozik. 1998. Binary specification of nonsense codons by splicing and cytoplasmic translation. *EMBO Journal* **17**: 3484-3494.

Urlaub, G., P.J. Mitchell, C.J. Ciudad, and L.A. Chasin. 1989. Nonsense mutations in the dihydrofolate reductase gene affect RNA processing. *Molecular and Cellular Biology* **9**: 2868-2880.

Vagner, S., C. Vagner, and I.W. Mattaj. 2000. The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing. *Genes Dev* **14**: 403-413.

Wahle, E. 1995. Poly(A) tail length control is caused by termination of processive synthesis. *Journal Of Biological Chemistry* **270**: 2800-2808.

Waldsich, C., K. Semrad, and R. Schroeder. 1998. Neomycin B inhibits splicing of the td intron indirectly by interfering with translation and enhances missplicing in vivo. *RNA* **4**: 1653-1663.

Welch, E.M. and A. Jacobson. 1999. An internal open reading frame triggers nonsense-mediated decay of the yeast SPT10 mRNA. *EMBO Journal* **18**: 6134-6145.

Weng, Y.M., K. Czaplinski, and S.W. Peltz. 1996a. Genetic and biochemical characterization of mutations in the ATPase and helicase regions of the Upf1 protein. *Molecular and Cellular Biology* **16**: 5477-5490.

Weng, Y.M., K. Czaplinski, and S.W. Peltz. 1996b. Identification and characterization of mutations in the UPF1 gene that affect nonsense suppression and the formation of the Upf protein complex but not mRNA turnover. *Molecular and Cellular Biology* **16**: 5491-5506.

Weng, Y.M., K. Czaplinski, and S.W. Peltz. 1998. ATP is a cofactor of the Upf1 protein that modulates its translation termination and RNA binding activities. *RNA* **4**: 205-214.

Wickens, M. 1992. Forward, backward, how much and when: mechanismes of poly(A) addition and removal and their role in early development. *Sem. Dev. Biol.* **3**: 399-412.

Wickner, S., M.R. Maurizi, and S. Gottesman. 1999. Posttranslational quality control: Folding, refolding, and degrading proteins. *Science* **286**: 1888-1893.

Yonaha, M. and N.J. Proudfoot. 1999. Specific transcriptional pausing activates polyadenylation in a coupled in vitro system. *Mol Cell* **3**: 593-600.

Yun, D.F. and F. Sherman. 1996. Degradation of CYC1 mRNA in the yeast Saccharomyces cerevisiae does not require translation. *Proc. Natl. Acad. Sci. USA* **93**: 8895-8900.

Zhang, J., X.L. Sun, Y.M. Qian, J.P. LaDuca, and L.E. Maquat. 1998a. At least one intron is required for the nonsense-mediated decay of triosephosphate isomerase mRNA: a possible link between nuclear splicing and cytoplasmic translation. *Molecular and Cellular Biology* **18**: 5272-5283.

Zhang, J., X.L. Sun, Y.M. Qian, and L.E. Maquat. 1998b. Intron function in the nonsense-mediated decay of beta-globin mRNA: Indications that pre-mRNA splicing in the nucleus can influence mRNA translation in the cytoplasm. *RNA* **4**: 801-815.

Zhang, S.A., M.J. Ruiz-Echevarria, Y. Quan, and S.W. Peltz. 1995. Identification and characterization of a sequence motif involved in nonsense-mediated messenger RNA decay. *Molecular and Cellular Biology* **15**: 2231-2244.

Zhang, S.A., E.M. Welch, K. Hagan, A.H. Brown, S.W. Peltz, and A. Jacobson. 1997. Polysome-associated mRNAs are substrates for the nonsense-mediated mRNA decay pathway in Saccharomyces cerevisiae. *RNA* **3**: 234-244.