Studies of the Relationship between mRNA Stability and Gene Function in *Saccharomyces cerevisiae.*

A thesis submitted for the Degree of
Doctor of Philosphy at the University of Glasgow

by

Joyce Moore

Institute of Genetics
Church Street
Glasgow

Febuary 1991

Dedicated to Myra and Jimmy Moore
for all their love and support over the years.

The research reported in this thesis is my own original work, except where otherwise stated, and has not been submitted for any other degree.

Contents

Acknowledgements

Thanks to Howy Jacobs for his supervision and for critical reading of the manuscript, to Roger Sutcliffe for suggesting the project, and to Adrienne Jessop and Richard Wilson for their encouragement. Thanks also to T.C. Santiago and Alistair Brown for providing the cDNAs, to Derek Jamieson (Edinburgh) who provided the temperature sensitive splicing mutant and hints on how to handle it, to Priyal Die Zoysa (Reading) who provided the OFAGE blot, and to Iain Johnstone for help in identifying the protein encoded by cDNA13. A special thanks to the first, third and fifth floors of the Genetics department for providing essential equipment, even though they did not know about it most of the time.

This work would have been much more laborious without the help of the prep. room ladies, and I would like to thank them more specifically for nagging me towards the end of this work.

There is no longer a yeast group, but past members deserve thanks for their help, especially Paul for his encouragement and for keeping me in touch.

A special thanks to those people who made my stay in the Genetics department more bearable and who, thank goodness, do not want their virtues spelled out, in particular Jane, Fiona and Arvind .

My mother and father helped enormously, especially in regard to making living on a student grant somewhat more bearable, and by always being on my side.

Lastly thanks to Kim for his understanding, useful suggestions, very critical reading of the manuscript, and his friendship throughout.

This work was supported by a Science and Engineering Research Council Studentship.

Summary

Twelve *S. cerevisiae* cDNAs, characterised by Santiago (1986) on the basis of the half-lifes of the respective mRNAs, have been partially or completely sequenced. To four have been assigned a definite function. cDNA10, which generates a long half-life mRNA, encodes the glucose-inducible form of the glycolytic enzyme, enolase. cDNAs 90, 39 and 13, each of which generates a short half-life mRNA, encode ribosomal proteins L3, L29 and YL6 respectively. cDNA46, which generates a long half-life mRNA, is related to a mouse gene of unknown function, *MER5*. In addition, there is circumstantial evidence that among the unidentified cDNAs that generate short half-life mRNAs are several that encode ribosomal proteins.

cDNA13 has been the subject of the most study, since it represents a newly-discovered ribosomal protein gene of *S. cerevisiae*. This was first suggested by its homology at both the DNA and putative amino acid levels with the gene encoding ribosomal protein K37 of *S. pombe*. In addition, mRNA13 "behaves" like the mRNA for a ribosomal protein, in that its abundance increases sharply when glucose is introduced to a yeast culture grown on ethanol (nutrient-upshift), and declines when yeast are subjected to a mild heat-shock. Further, the gene(s) corresponding to cDNA13 contains an intron, as first indicated by the accumulation of unspliced mRNA13 when cells harbouring a temperature-sensitive splicing mutation are grown at the restrictive temperature.

Genomic southern blot analysis suggested the presence of two genes related to cDNA13 in the *S. cerevisiae* genome. This was confirmed by OFAGE analysis which, in addition, shows that the two copies reside on different chromosomes. An *S. cerevisiae* genomic DNA library was constructed in the bacteriophage lambda replacement vector, EMBL3. From it was isolated one of the genes related to cDNA13. Sequencing of this gene provided a complete amino acid sequence for the corresponding protein. The putative N-terminal sequence is identical to that of the previously characterised *S. cerevisiae* ribosomal protein, YL6. Further, there is clear evidence within the sequence for the predicted intron, as well as of elements characteristic of UAS_{rpg}s (upstream activation element that regulate the expression of *S. cerevisiae* ribosomal protein genes).

CHAPTER 1

INTRODUCTION

1. Introduction

Since this thesis revolves around the observation that *S. cerevisiae* ribosomal proteins are in general encoded by short half-life mRNAs, and more specifically concerns the identification and characterisation of a new ribosomal protein gene of *S. cerevisiae*, this introduction is split into two parts. The first concerns mRNA stability and the second ribosome structure, function and assembly.

1.1. Background to this study

In 1978 Johnston and Singer reported that 1,10-phenanthroline inhibits RNA synthesis in *S. cerevisiae*. RNA polymerase II is a zinc metalloenzyme; 1,10-phenanthroline inhibits its action by chelating zinc (Lattke and Weser, 1976). Based on these observations, Santiago (1986) developed a new method for the estimation of mRNA half-life in *S. cerevisiae*.

The method is as follows: 1,10-phenanthroline (final concentration 100 μg/ml) is added to a culture of *S. cerevisiae* growing exponentially in YPG medium. This is taken to be time point zero. Thereafter, aliquots of the culture are removed at 3 to 5 minute intervals and stored in 2 volumes of ethanol at -20°C until required. Sampling continues until approx. 45 minutes after the addition of the drug, normally 8 time points per mRNA half-life estimation. Longer incubation times have not been used, both because mRNAs with half-lifes longer than 100 minutes are not common in *S. cerevisiae* (Koch and Friesen, 1979), and because the long term effects of phenanthroline on transcription and other aspects of cellular physiology have not been investigated.

Total RNA is extracted from the samples, and the relative abundance of specific mRNAs at each time-point is determined by quantitation of dot-blot hybridisation with gene-specific probes. The data are normalised with respect to the abundance of 18S rRNA (small subunit rRNA). Results obtained by this method are therefore not absolute measurements but rather an indication of the stability of any one mRNA relative to that of the 18S control.

To select genes representing mRNAs with both short and long half-lifes, Santiago (1986) used a differential screening approach. The rationale for this approach was that ,while a subset of genes is turned on during heat shock (genes encoding the heat-shock proteins), transcription of the majority of genes is switched off. Thus, when duplicate filters are screened with both normal and heat-shock probes, colonies hybridising only

with the former might be expected to correspond to short half-life mRNAs, while those hybridising with both should correspond to long half-life mRNAs. A cDNA library was constructed from *S. cerevisiae* DBY746 by a method involving homopolymeric tailing of mRNA:cDNA hybrids (Wood and Lee, 1976): extensions of C-residues were added to the mRNA:cDNA hybrids and extensions of G-residues to the vector (pBR322 cleaved with *Pst*I). The library was then screened differentially with probes generated by reverse transcription of poly(A)$^+$ mRNA isolated from DBY746, either grown at 23°C or grown at 23°C and then heat-shocked for 20 minutes at 36°C. A 20 minute heat-shock was used as yeast growing fermentatively (though not by respiration; Section 4.1.4.) are able to resume normal growth and transcription after approximately 30 to 40 minutes at the elevated temperature (Linquist, 1984).

The results of this experiment were disappointing: after 3 rounds of screening only 3 clones behaved consistently as if they contained copies of short half-life mRNAs (cDNA9, cDNA13 and cDNA19), and only one appeared to be derived from a long half-life mRNA (cDNA11). In fact, while the assignment of cDNAs 9, 13 and 19 turned out to be correct, the assignment of cDNA11 was wrong: later experiments showed that cDNA11 was also a copy of a short half-life mRNA (Fig. 3.1.). These results may indicate that the original assumptions were not valid, and indeed Kim and Warner (1983a) have provided evidence that while mRNAs encoding ribosomal proteins are significantly depleted after 20 minutes at 36°C, synthesis of the bulk of mRNAs continues normally after such a mild temperature-shift. Other studies do not demonstrate quite such a clear-cut distinction. In a study of 500 individual proteins, some 300 were found to be reduced in abundance after a mild heat-shock, although to an extent that varies from protein to protein (Miller et al, 1982). However, it appears that ribosomal proteins are amongst the most severely affected. It is thus possible that Santiago may have enriched preferentially for copies of mRNAs that encode ribosomal proteins, rather than for copies of all short half-life mRNAs, as he had intended.

Because of the problems he encountered during the differential screen, Santiago (1986) chose several additional cDNA clones at random from his cDNA library. The only prerequisite was that they contained *Pst*I sites at the boundaries of the cDNA insert. In total he studied 13 cDNAs, details of which are given in Fig. 3.1 and Table 3.1. It can be seen that the respective mRNAs fall into two broad classes, one stable and one unstable, and that within each class mRNA half-life is inversely proportional to mRNA length.

1.2 mRNA Stability in Yeast and Other Organisms

Although transcription is an obvious initial step at which expression of protein coding genes can be controlled, it is now clear that in both prokaryotes and eukaryotes post-transcriptional events play an important role in establishing the final level of a protein product. In the eukaryotic nucleus these events include maturation of the primary transcript (5'-end capping; 3'-end selection and polyadenylation; RNA splicing) and transport of mature mRNA from the nucleus to the cytoplasm. In the cytoplasm additional control mechanisms regulate the rate of translation and the stability of the mRNA.

The earliest evidence that mRNA stability plays a role in eukaryotic gene expression came from studies of the steroid hormone, oestrogen. Avian ovalbumin mRNA is stabilised by oestrogen treatment: withdrawal of the hormone results in a tenfold decrease in its half-life (Palmiter and Carey, 1974). Since that time, control of mRNA stability has come to be recognised as an important determinant of gene expression in a number of different systems (reviewed below). Further a number of different mechanisms can be seen to be at work.

Specific sequence elements affecting mRNA stability have been identified in 5'- and 3'-untranslated regions, as well as in coding regions. Thus stability is, at least in part, sequence- or structure-dependent. However mRNA length may also play a role, at least in some instances, as may on-going translation, and the structures at each end of the mRNA, i.e. the 5'-cap and the 3'-poly(A) tail. Examples illustrating the control of stability at various levels are discussed below. A comparison of various methods used for measuring mRNA half-life can be found in Appendix 4.

1.2.1. mRNA Stability in Prokaryotes

Prokaryotic genes that encode related functions are often organised into polycistronic operons. Operons facilitate the coordinate induction and repression of the genes they contain, but restrict the mechanisms by which differential expression of the individual genes can occur. Differential expression is on occasion achieved by differential termination of transcription, but is more commonly the result of differences in translation efficiency (reviewed by McCarthy and Gualerzi, 1990). However, there are also examples of differential expression within an operon that can only be accounted for by regional differences in stability within an mRNA. The *puf* (previously referred to as *rxcA*) operon of *Rhodobacter capsulatus* encodes both the light-harvesting and the reaction-centre proteins required for photosynthesis. Belasco et al (1985) demonstrated

that preferential accumulation of the light-harvesting proteins results predominantly from segmental differences in the stability of the polycistronic mRNA. The 5' domain that encodes the light-harvesting proteins is more stable than the rest of the mRNA, from which it is separated by a highly-stable stem and loop structure. Removal of the stem and loop prevents preferential accumulation. It also impairs growth, presumably due to altered stoicheiometry of the two complexes in the intracytoplasmic membrane (Klug et al, 1987). It has been hypothesised that the function of the stem and loop structure is that of an mRNA decay terminator that protects the upstream region from a 3'-exoribonuclease (Chen et al, 1988). However, replacing the stem and loop structure at the 3'-end of the *puf* operon fails to stabilise the labile 3'-domain, suggesting that the initial target for endonucleolytic attack is not the 3'-end (Chen et al, 1988).

REP (**r**epetitive **e**xtragenic **p**alindromic) sequences are present in both *Salmonella typhimurium* and *Escherichia coli*, in such amounts that they represent as much as 1% of the genome (Stern et al, 1984). The presence of two REP sequences in inverted orientation, and thus with the potential to form a stem and loop, within the *hisJ-hisQ* intergenic region of the *S. typhimurium* histidine-transport operon, and at the boundary of the *malE* and *malF* genes within the *malEFG* operon of *E. coli*, preferentially stabilises 5' sequences. It has been hypothesised that this is due to protection from a 3' to 5' exonucleolytic activity (Newbury et al, 1987a; 1987b).

Two exonucleases, PNPase (**p**oly**n**ucleotide **p**hosphorylase, the *pnp* gene product) and RNaseII (the *rnb* gene product) are the enzymes primarily responsible for degrading mRNA to mononucleotides in bacteria (Donovan and Kushner, 1986). Both enzymes proceed in a 3' to 5' direction and no exoribonucleases with 5' to 3' activity have been identified (Belasco and Higgins, 1988). This is consistent with both the above examples.

1.2.2. mRNA Stability in Higher Eukaryotes

Oestrogen stabilises the already stable vitellogenin mRNA of *Xenopus laevis* liver cells, leading to an increase in half-life from 16 to approx. 500 hours (Brock and Shapiro, 1983). On-going protein synthesis seems not to be required, since inhibitors of translation elongation have no effect. Blocking translation-initiation on the other hand, which reduces the density of ribosomes on the mRNA, prevents stabilisation. Thus the maintenance of a high density of ribosomes would seem to be necessary for stabilisation. It cannot be sufficient, however, since removal of oestrogen from culture medium does not alter the density of ribosomes on polysomal vitellogenin mRNA (Blume and Shapiro,

1989) even though this mRNA becomes labile. The effects of oestrogen and other hormones on mRNA stability have been reviewed by Shapiro et al (1987).

DNA replication-dependent histone mRNAs of mammalian cells have a stem and loop structure at their 3'-ends instead of a poly(A) tail. The concentration of these mRNAs varies by a factor of 30 to 50 fold during the cell cycle, rapidly increasing at the onset of DNA synthesis and equally rapidly decreasing at the end of DNA replication. These fluctuations results from moderate changes in transcription rates, together with much greater changes in mRNA stability (Sittman et al, 1983). The stem and loop structure at the 3'-end of the histone mRNA is necessary and sufficient for the observed destabilisation of histone mRNAs after DNA synthesis (Pandey and Marzluff, 1987), though translation of each mRNA must first have proceeded to within 300 nucleotides of the 3'-end (Graves et al, 1987).

Both *in vivo* and *in vitro* experiments indicate that histone mRNA degradation proceeds in a 3' to 5' direction (Ross and Kobs, 1986; Ross et al, 1986; 1987). Further, an exonuclease that degrades histone mRNA *in vitro* (Peltz et al, 1987) is very inefficient in degrading poly(A)$^+$ mRNA, suggesting that the poly(A) tail plus its associated protein complex (Blobel, 1973; reviewed by Bernstein and Ross, 1989) protects polyadenylated mRNAs (Peltz et al, 1987). Peltz and Ross (1987), demonstrated, by use of the *in vitro* system, that histone mRNAs are degraded 4-6 fold faster in reactions containing free histones, and that histones form complexes with histone mRNAs. However, the exact nature of this autoregulation remains to be elucidated.

Although not a higher eukaryote, it is pertinent to discuss one aspect of *S. cerevisiae* mRNA stability in this section. In this organism, mRNAs encoding the DNA replication-dependent histones do have a poly(A) tract at their 3'-end (Fahrner et al, 1980), unlike their counterparts in higher eukaryotes. The abundance of these mRNAs increases at the onset of S-phase and decreases once DNA replication ceases, either naturally or through the use of temperature-sensitive mutants, or drugs (Hereford et al, 1981; Lycan et al, 1987). Unlike the situation in higher eukaryotes, inhibition of DNA chain elongation seems predominantly to repress transcription of histone genes. However, when transcription regulation is uncoupled, by placing a histone gene under the control of a constitutive promoter, from post-transcriptional events, a 3-fold difference is observed in the stability of histone mRNAs between S-phase and G2. Further, this difference appears to be mediated via the 3'-end of the message (Lycan et al, 1987). Since the rate of turnover of histone mRNAs in yeast, independent of any specific control mechanism, appears to be relatively fast (Osley and Hereford, 1981), post-transcriptional mRNA destabilisation may not be as critical as it is in higher eukaryotes for coupling mRNA levels to DNA replication (Lycan et al, 1987).

Stability of mRNA encoding the human transferrin receptor (hTR), a protein required for transporting iron into cells, is regulated by iron. hTR mRNA is relatively stable in cells starved of iron and *vice versa*. Sequence elements involved in this control have been shown to be located within the 3'-untranslated region (Owen and Kuhn, 1987), and more precisely within a 678-nucleotide domain with the capacity to form five stem and loop structures (Casey et al, 1988). Strikingly, elements of this region have significant homology with the sequence of an Iron Responsive Element (IRE) located in the 5'-untranslated region of human ferritin mRNA, and which mediates iron-dependent translation (Hentze et al, 1987). Further, two of the hTR stem and loop structures can function as IREs, mediating ferritin-like translation regulation when individually inserted into the 5'-untranslated region of a human growth hormone gene (Casey et al, 1988). A cytosolic protein (IRF; **i**ron **r**egulatory **f**actor) that binds to the IRE involved in translational regulation of ferritin has been isolated from rodent and human cells (Leibold and Munro, 1988; Rouault et al, 1988). IRF also binds to two of the five IRE-like elements in the 3'-untranslated region of hTR mRNA (Koeller et al, 1989). In the presence of an iron chelator, the RNA binding activity of IRF is increased 25-fold; conversely the addition of iron salts causes rapid inactivation of IRF followed by hTR degradation (Mullner et al, 1989). It is envisaged that in cells starved of iron, IRF protects hTR mRNA from endonucleolytic cleavage, the rate limiting step in its turnover, and inhibits translation of ferritin mRNA. In the presence of iron the half-life of hTR mRNA is prolonged by treatment of cells with either actinomycin D or cycloheximide (inhibitors of RNA synthesis and protein synthesis respectively; Mullner and Kuhn, 1988). However, IRF inactivation is not inhibited by either of these compounds indicating that a second component that is regulated independently of IRF may be involved, possibly the ribonuclease itself (Mullner et al, 1989).

AU-rich elements (AREs) are found in the 3'-untranslated regions of a variety of transiently expressed mRNAs, for example proto-oncogene mRNAs (e.g. *c-fos* and *c-myc*), as well as mRNAs encoding several lymphokines, cytokines and inflammatory mediators. The high degree of conservation seemed to imply functionality, and indeed AREs appear to play a direct role in determining the stability of an mRNA (see Shaw and Kamen, 1986). The best documented example is the 51nt ARE found in the 3'-untranslated region of the mRNA encoding lymphokine GM-CSF (**g**ranulocyte **m**onocyte-**c**olony **s**timulating **f**actor). Transplantation of this ARE from GM-CSF into the 3'-untranslated region of a human β-globin gene causes the half-life of the β-globin mRNA to be reduced from >17 hours to less than 30 minutes, similar to the half-life of the GM-CSF transcript itself (Shaw and Kamen, 1986). β-globin mRNA containing a transplanted ARE is stable in cells treated with cycloheximide (Shaw and Kamen, 1986), suggesting that active translation or association with ribosomes is required for

destabilisation, or alternatively that some component required for degradation is a short lived protein.

Although AREs contribute towards the observed short half-lifes of *c-myc* and *c-fos* mRNAs, they are insufficient to explain the entire effect. Activation of the oncogenic potential of *c-myc* often involve gene translocations that truncate exon 1 (a non-coding exon); some truncated *c-myc* mRNAs have increased stabilities (Rabbitts et al, 1985), even though they retain their 3' ARE. However, exon 1 does not significantly stabilise other messages, suggesting that the 3'-untranslated region plays the predominant role (Jones and Cole, 1987). Evidence for this role is provided by the human T-cell leukemia line HUT-78, in which the 3'-untranslated region of *myc* is truncated as a result of translocation. This alteration removes a 61 nt sequence from the ARE, and indeed the half-life of the mRNA is increased by at least five-fold (Aghib et al, 1990). Further, the translocation has fused *myc* to a region of chromosome 2 that appears to have no coding potential, and that possibly has no other genetic function. These experiments clearly demonstrate the importance of mRNA stability in the regulation of *c-myc*, and sustain the view that increased expression of an otherwise normal allele of *c-myc* can play a role in oncogenesis.

There exist two distinct destabilising sequences within human *c-fos* mRNA, one a 75 nt ARE in the 3'-UTR, and the other located within the coding region. Placing the ARE within the 3'-UTR of the normally stable β-globin mRNA causes the chimeric transcript to degrade with a half-life approaching that of wild-type *c-fos* mRNA. However, little effect on the stability of *c-fos* mRNA itself is caused by its removal (Shyu et al, 1989). Replacement of the β-globin coding region with that of *c-fos* has a two-fold greater destabilising effect than transplanting the ARE. Most significantly the two elements appears to be targetted by different degradation pathways (Shyu et al, 1989). Two different transcription inhibitors (functioning by different mechanisms) interfere with the destabilising effect exerted by the ARE (in chimeric genes), while having only a small effect on decay mediated by the coding region. Neither inhibitor significantly affects wild-type *c-fos* mRNA degradation, again demonstrating that the coding region carries the major determinant of stability. From the above it appears that ARE-dependent mRNA degradation requires synthesis of a labile gene product that is not required for degradation mediated by the coding region. The existence of two independent degradation pathways for *c-fos* mRNA could serve an important biological function. Over-expression of *c-fos* can lead to oncogenesis (Meijlink et al, 1985); the presence of two destabilising sequence may act as some form of safeguard.

β-tubulin, a principal component of microtubules in higher eukaryotes, is encoded by a small multigene family (reviewed by Cleveland, 1987). As in the case of histone

mRNAs, human β-tubulin mRNAs are destabilised by high levels of their own unpolymerised translation products. Autoregulation is not at the level of transcription (Cleveland and Havercroft, 1983), rather it occurs as a consequence of cytoplasmic events (Pittenger and Cleveland, 1985). Sequences required for autoregulation of β-tubulin mRNA stability lie within the first 13 translated nucleotides of the mRNA (Gay et al, 1987). They are able to confer tubulin mediated instability on heterologous mRNAs, but again only if present as the first translated nucleotides. Further, by using various inhibitors of protein synthesis, or β-tubulin genes with premature translation termination codons (Pachter et al, 1987; Yen et al, 1988), it was shown that β-tubulin mRNA must be associated with polysomes in order to be a substrate for autoregulation, and that translation must proceed past the first 41 codons.

Yen et al (1988) systematically introduced 25 different nucleotide substitutions, both conservative and non-conservative, into the 13 nucleotide regulatory element of a human β-tubulin gene. Only mRNAs with conservative changes were regulated correctly. Thus it is not the mRNA, but the amino-terminal tetrapeptide (Met-Arg-Glu-Ile), that responds to tubulin levels. Translation past codon 41 is presumably required so as to allow this tetrapeptide to emerge from the ribosome. In bacterial ribosomes, for example, at least 30 to 40 amino acids are contained in a tunnel in the large subunit after translation (Yonath et al, 1987). The N-terminal tetrapeptide is absolutely conserved between β-tubulin proteins and is unique to them at this position (Yen et al, 1988). Thus it is hypothesised that regulated degradation of β-tubulin mRNAs, following elevation of the unpolymerised tubulin subunit pool, is caused by an interaction between an autoregulatory factor(s) (possibly tubulin itself) and the nascent amino terminal tubulin tetrapeptide as it emerges from the ribosome (Yen et al, 1988). This interaction might activate an RNase (which may be a ribosome component, or may be recruited to the ribosome); alternatively binding may induce transient stalling, leaving the mRNA in an exposed confirmation for cleavage by a non-specific endonuclease.

1.2.3. mRNA Stability in Yeast

Progress in the field of yeast mRNA stability has been much slower. It was stated earlier that, at least for 13 cDNAs of unknown function, stability is inversely proportional to mRNA length (Santiago, 1986). This conclusion was strengthened by measurement, also by the 1,10-phenanthroline method, of the half-lifes of actin, *PYK*, and *URA3* mRNAs as well as of two chimeric mRNAs, *PYK/URA3* and *PYK/LacZ* (Santiago et al, 1986; Purvis et al 1987 and Brown et al, 1988). On this basis it is tempting to suggest that in general, that the longer the RNA the larger is the target for endonucleolytic cleavage. However, doubt as to the generality of these observations arises from

experiments in which half-life was measured by a method involving the use of a strain with a temperature sensitive lesion in the large subunit of RNA polymerase II (*rpb1-1;* **R**NA polymerase **B**), that at the restrictive temperature causes transcription of PolII genes to cease immediately (Herrick and Jacobson, 1988). Half-lifes of 11 mRNAs of known function determined by this method showed no correlation with length. The use of a different method may be one reason for the discrepancy; alternatively the relationship between length and stability may only hold in some instances.

Even should there not be a general relationship between stability and length, the observation that there are two populations of mRNA in yeast with regard to stability is still an interesting one. Might there be a general mechanism underlying the observation? For example the populations might differ in the degree to which they are protected by ribosomes, either due to differential recruitment for translation or for some other reason. It has been shown that nonsense mutations in the proximal regions of *URA1* and *URA3* destabilise the respective transcripts, possibly because the distal portions are no longer protected by ribosomes; nonsense mutations located towards the 3'-end of these genes have little effect (Losson and Lacroute, 1979; Pelsy and Lacroute, 1984). In contrast, when proximal stop codons were introduced into the coding regions of *PYK1* and a *PYK/LacZ* fusion, no significant changes were observed in mRNA stability (Purvis et al, 1987). Further, in cases where the polysome profiles of stable and unstable mRNAs have been examined, no great differences in ribosome loading have been observed (Santiago, 1986; Santiago et al, 1987). Thus, while ribosome loading may affect mRNA half-life to a minor extent, it does not seem to account for the dramatic differences in stability that are observed.

Yeast mRNAs may have one of two cap structures, $m^7G(5)ppp(5')A$ and $m^7G(5)ppp(5')G$ (Scripati et al, 1976). Could it be this difference that determines whether an mRNA belongs to the stable or unstable population? Apparently not since, when *S. cerevisiae* mRNAs were labelled with [^{32}P] orthophosphate *in vivo*, and the two cap structures were separated from one another and from mononucleotides, the ratio of the two structures after different periods of labelling was the same (Piper et al, 1987). Shorter labelling periods will preferentially label mRNAs that are turned over rapidly while long labelling periods allow labelling of more stable mRNAs.

In a recent extension of the work of Herrick and Jacobson (1988; see above), and using the same method, Herrick et al (1990) investigated 20 mRNAs; 18 of known function, plus cDNA74 and cDNA90 (Santiago, 1986; Table 3.1). Again, no specific relationship between length and stability was apparent, nor were two distinct populations of mRNA. Herrick et al (1990) did, however, categorise the mRNAs as members of three arbitrary stability groups: stable, half-life > 25 minutes; moderately stable, half-life 10 to 20

minutes; and unstable, half-life < 7 minutes. Especially pertinent was their measurement of the half-lifes of mRNAs74 and 90. mRNA90 was estimated to decay with a half-life of 7 minutes, in good agreement with the 6.6 minutes determined by Santiago (1986; Table 3.1.). mRNA74 on the other hand, which Santiago had assigned to the stable population (half-life approx. 83.4 minutes; Fig.3.1. and Table 3.1.), was estimated to have a half-life of only only 17 minutes and thus, by their criteria, to be one of the moderately stable mRNAs. In an attempt to validate their own method, Herrick et al (1990) compared their results with those obtained using incorporation kinetics (Appendix 4). Several of the measurements were consistent, although several were not. In the present context, the half-life of mRNA74 was found to be approx. 37 minutes, in between their own results and Santiago's. Herrick et al (1990) also analysed the effect of cycloheximide on mRNAs from each stability group; in every case the treatment resulted in increased mRNA stability. This could imply either that ribosome translocation is required for normal degradation of an mRNA (see below), or that a factor required for degradation of each class, possibly an RNase(s), is metabolically unstable.

Both Santiago et al (1986) and Herrick et al (1990) examined poly(A) tail length, since experiments on higher eukaryotes have suggested that the degradation of some mRNAs is initiated by deadenylation (Brewer and Ross, 1988; Wilson and Triesman, 1988). In neither study was an obvious correlation between stability and poly(A) tail length observed. Moreover, deadenylation of two stable mRNAs, including one represented by cDNA39 (see Table 3.1.), is not the rate limiting step, since the shortening or loss of the poly (A) tails is significantly faster than the degradation of the mRNAs (Herrick et al, 1990) .

Since neither differences in the poly(A) tail, nor the cap, seem to account for the differences observed between individual mRNA half-lifes in yeast, signals internal to the mRNA must be involved as already discussed for mammalian mRNAs. Studies of *MATα1* mRNA are consistent with this. The mRNA encoded by *MATα1* is very unstable (half-life, approx. 5 minutes; Herrick et al, 1990). Various chimeric mRNAs have been produced by in frame fusion of either *ACT1* or *PGK1* (both stable mRNAs) to regions of *MATα1*. Although the 3'-untranslated region of *MATα1* does not affect the stability of either type of chimeric mRNA, a 42nt element from the coding region does (Parker and Jacobson, 1990). This element has no striking homology at either the nucleotide or amino acid levels with sequences found in other unstable yeast mRNAs. However, 8 out of the 14 codons within it are rare (occurring fewer than 13 times per 1000 yeast codons; see Section 3.1.), and 6 out of 7 of these are consecutive. Moreover, the percentage of rare codons has been noted to be, in general, higher in unstable than in stable mRNAs (Herrick et al, 1990). This could be taken to suggest that

translational-pausing can cause destabilisation, perhaps by allowing the folding of an mRNA into a structure that is the target of a ribonuclease. On the other hand, the fact that the codons are rare may be a red herring. Whatever the case, any model must account for the observation that an in-frame stop codon placed upstream of the 42nt region increases mRNA half-life, suggesting that rapid decay promoted by the 42nt sequence is dependent on ribosome translocation (Parker and Jacobson, 1990). This would be consistent with the effects of cycloheximide on *MATα1* mRNA half-life (Herrick et al, 1990). Among those mRNAs designated as being unstable by Herrick et al (1990) were others involved in mating type switching.

1.3. The Ribosome

Ribosomes, ubiquitous cellular organelles, are sites of protein synthesis. The concept of the ribosome evolved during the nineteen-fifties, as it became apparent that sub-fractions of the "microsome" (a term used to describe fractions consisting of endoplasmic reticulum and what are now known to be ribosomes) retained synthetic ability. In particular, Littlefield et al (1955) were able to dissociate rat liver reticulum and ribosomes, following *in vivo* labelling with radioactive amino acids.

In fact, ribosomes had been visualised even earlier. For example, Luria, Delbruck and Anderson (1943), observed bacterial ribosomes while using electron microscopy to study the lysis of *E. coli* by phage. They saw, "besides the particles of virus, also granular material of very regular units ..... their size corresponds to a molecular weight of the order 10^6. These particles are liberated from the cell in great abundance, and seem to constitute the bulk of the cytoplasm." However, it was not until the experiments of McQuillen et al (1959) that bacterial ribosomes were unequivocally shown to be the sites of protein synthesis, and thus to have similar properties to ribosomes of higher eukaryotes.

By the beginning of the nineteen-sixties, there was general agreement that protein synthesis in a variety of organisms and tissues occurs on rather spherical particles with a diameter of 200-300 Å (somewhat smaller in the case of bacteria), consisting of RNA and protein in roughly equal proportion. Since then, progress in our understanding of the structure and function of the ribosome has been considerable, particularly in the case of the *E. coli* ribosome, which is composed of 3 RNA species and 52 proteins (Nomura et al, 1984). Although the eukaryotic ribosome, composed of 4 RNA species and 75 or more proteins is less well understood, considerable progress is now being made, particularly in the case of the yeast ribosome. The advantage of yeast, over most other

eukaryotic organisms, lies in the degree to which genetic analysis can be used to support more traditional approaches.

1.3.1. Ribosomal RNA (rRNA)

The three rRNAs of *E. coli* are generated as a single precursor molecule with the order 5'-16S-23S-5S-3' (Ginsburg and Steitz, 1975). There are seven rRNA operons in *E. coli* K12, all of which have been cloned and their chromosomal locations determined (see Nomura et al, 1984). A few are clustered around the origin of replication, and since initiation of DNA replication occurs more frequently at higher growth rates, tend to be over-represented in fast-growing cells. Thus the relative contribution of each operon to the synthesis of rRNA depends upon growth rate. Further, the direction of transcription of all of the rRNA operons is the same as the direction of DNA replication. It has been speculated that this arrangement has evolved to prevent the DNA replication machinery and RNA polymerases from colliding with one another (Nomura et al, 1977).

Genes encoding the several RNA components of the ribosome of *Saccharomyces cerevisiae* are present on a 9.1 kb unit, which is tandemly repeated within chromosome XII (Phillippsen et al, 1978a; Petes, 1979). The copy number per haploid genome normally ranges from 100-120 (Phillippsen et al, 1978b), representing approx. one third of the chromosome. The organisation of the repeated unit is shown in Fig. 1.1. (Phillippsen et al, 1978b). 5S rRNA, and a 35S precursor rRNA containing 5.8S, 18S and 25S sequences, are both encoded within the repeat, but by separate transcription units (Udem and Warner, 1972; Trapman and Planta, 1975). This differs from the arrangement seen in many higher eukaryotes, where the genes encoding 5S rRNA and the 35S rRNA precursor occupy different chromosomal locations (Wimber and Steffensen, 1970; Pardue et al, 1973; Wimber et al, 1974; Attardi and Amaldi, 1970). In *S. cerevisiae*, sequences encoding the precursor rRNA are transcribed in the order 5'-18S-5.8S-25S-3', the same as in higher eukaryotes (Kramer et al, 1978; Hackett and Sauerbier, 1975; Liau and Hurlbert, 1975; Dawid and Wellauer 1976). The 5S rRNA gene is transcribed in the opposite direction (Kramer et al, 1978). A complete sequence for the 9.1 kb repeat unit is now available (Rubin, 1973; Maxam et al, 1977; Rubstov et al, 1980; Georgiev et al, 1981; Mankin et al, 1986; Skryabin et al, 1986). It should be remembered, however, that restriction site polymorphisms have been observed between repeats, and that deletions and insertions can occur (Swanson et al, 1985; Jemtland et al, 1986).

In eukaryotes, 35S precursor rRNA is synthesised by RNA polymerase I (reviewed by Sollner-Webb and Tower, 1986), while 5S rRNA together with tRNAs are synthesised

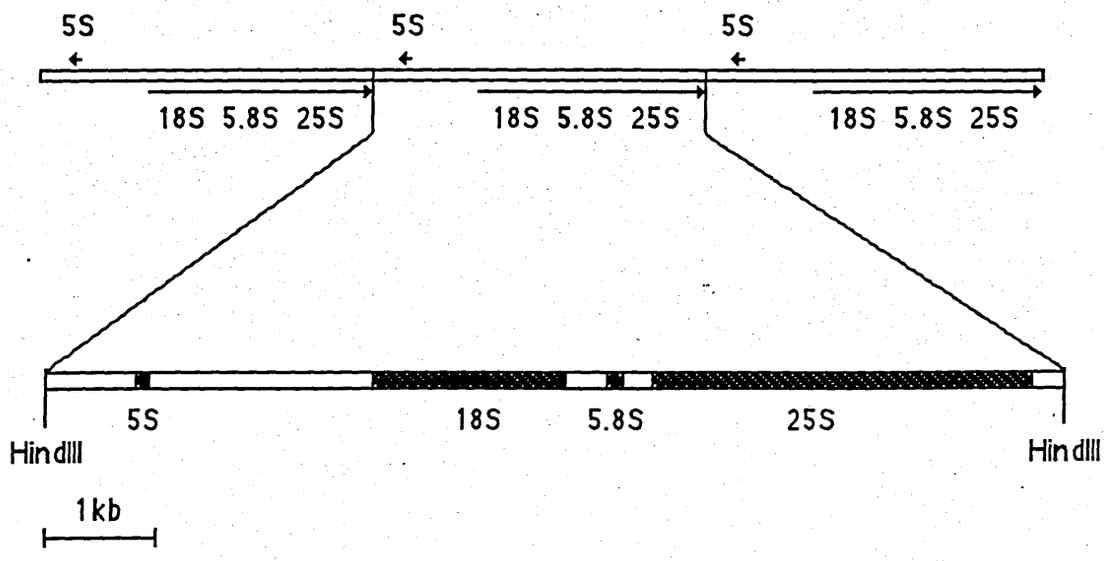


Figure 1.1 Map of yeast rDNA . The upper portion represents 3 tandemly repeated rDNA units. The arrows indicate the direction of transcription. The lower portion represents one repeating unit. Filled boxes represnt the location of segments that come to form the mature rRNA species. Taken from Phillippsen et al (1978b).

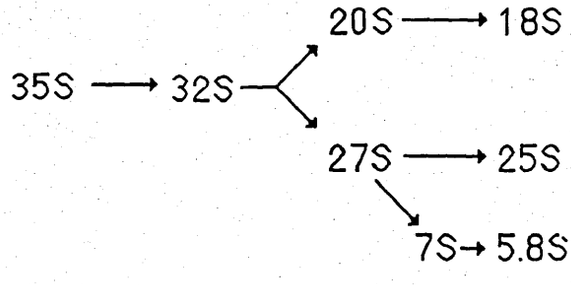


Figure 1.2 The sequence of rRNA processing steps. Taken from Warner (1982).

by RNA polymerase III (reviewed by Geiduschek and Tocchini-Valentini, 1988). 35S rRNA is modified by methylases, pseudouridylase(s), endonucleases and exonucleases, giving rise to mature 25S, 18S and 5.8S rRNA species (see Warner, 1989; Fig. 1.2.). Processing results in loss of 10-20% of the original precursor molecule (Udem and Warner, 1972). Mature rRNAs greatly exceed 35S rRNA in abundance, suggesting that the precursor is processed rapidly. Following processing, 5.8S rRNA becomes non-covalently bound to 25S rRNA, from which it can be released by heat or urea treatment (Udem et al, 1971). Approximately 60 methyl groups are added to 35S rRNA both during and after transcription, mostly to the 2' OH of the ribose moiety, but also to the bases (Veinot-Drebot et al, 1988; Klootwijk and Planta, 1973). Pseudouridylation occurs prior to processing (Brand et al, 1979). 5.8S, 25S and 5S rRNAs, together with associated ribosomal proteins form the large (60S) subunit, while 18S rRNA together with associated ribosomal proteins form the small (40S) subunit.

Mutations in two *S. cerevisiae* genes are known to affect the processing of rRNA. One of the genes, *rna82*, appears to encode an endonuclease that both processes the 3' end of 5S rRNA, and has a role in 35S rRNA 3' end formation (Piper et al, 1983; Kempers-Veenstra et al, 1986). Mutations in the second gene, *rrp1* (**r**ibosomal **R**NA **p**rocessing), block the conversion of 27S rRNA to 25S and 5.8S rRNAs (see Fig. 1.2.), and slow conversion of 20S rRNA to 18S rRNA. Although the exact function of the gene is not known, its product is not a ribosomal protein (Fabian and Hopper, 1987).

Elucidation of the details of rRNA processing thus awaits the identification and characterisation of most of the enzymes involved. However, it is clear that ribosomal proteins themselves play a role (Warner and Udem, 1972). A recent example of this concerns mutations in the gene encoding ribosomal protein L29. Most processing events, as well as ribosome assembly, take place in the nucleus (the exception to this being processing of 20S to 18S rRNA, which takes place in the cytoplasm; Udem and Warner, 1973). Mutations that interfere with nuclear localisation of L29 lead to accumulation of 35S and 27S species (Underwood and Fried, 1990).

(Please refer to Section 1.3.7. for a discussion of *S. cerevisiae* ribosomal protein nomenclature)

As mentioned above, genes encoding eukaryotic 5S rRNA and tRNAs are transcribed by RNA polymerase III. A *Xenopus* transcription factor, TFIIIA, has been characterised that is specific for 5S rRNA genes, and that by virtue of binding internally has the unique ability to bind both the gene and its RNA product. *In vitro* transcription of

5S rRNA genes can be inhibited by excess 5S rRNA, while the transcription of tRNA genes remains unaffected, providing a feedback mechanism that is likely to be important in 5S rRNA gene regulation *in vivo*. A similar form of feedback inhibition exists in *S. cerevisiae*, again thought to be mediated by a TFIIIA-like factor (Klekamp and Weil, 1982). There is also a yeast ribosomal protein, YL3, that binds to 5S rRNA. In an *in vitro* system from which YL3 has been depleted (by immunoprecipitation), feedback inhibition by 5S rRNA is more severe than when the YL3 is present (Brow and Geiduschek, 1987). This suggests one mechanism by which the stoicheiometry of ribosomal proteins and rRNA might be controlled, i.e. when YL3 is limiting, the transcription factor will bind to newly synthesised 5S rRNA, thus inhibiting further synthesis of 5S rRNA, without affecting the transcription of tRNA genes. This hypothesis is supported by the finding that rat ribosomal protein L5 (the counterpart of YL3), and *Xenopus* TFIIIA protect overlapping regions of 5S rRNA from nuclease (Huber and Wool, 1986a; Huber and Wool, 1986b).

1.3.2. Newly-Synthesised Ribosomal Proteins are Targetted to the Nucleus

Assembly of the ribosome takes place in the nucleus. Thus ribosomal proteins, which are produced in the cytoplasm, must be transported to the nucleus for assembly to take place. In HeLa cells, newly synthesised ribosomal proteins accumulate in the nucleus, more specifically within the nucleolus, within 3 to 5 minutes of their synthesis (Warner, 1979).

Specific elements of their amino acid sequences facilitate the transport of ribosomal proteins to the nucleus. In the case of chimeric proteins formed by fusing regions of *S. cerevisiae* ribosomal protein L3 to *E. coli* β-galactosidase, for example, only the 21 N-terminal residues of L3 are required for the fusion protein to become located in the nucleus, as determined by staining with antibodies to β-galactosidase (Moreland et al, 1985). However, there is no apparent homology between this sequence and previously characterised nuclear localisation determinants. Nuclear localisation of another ribosomal protein, L29, is also determined by the sequence at the N-terminus. Initial experiments suggested that the first 16 amino acids are important (Underwood and Fried, 1990). In fact, it now appears there is some redundancy of information. The amino acid sequence of L29 between residues 6 and 13 is identical in five out of seven positions with the sequence between residues 23 and 29. Insertion of oligonucleotides encoding either of these heptapeptides (NLS-1 = KTRKHRG and NLS-2 = KHRKHPG: NLS stands for nuclear localisation signal; underlined amino acids differ between the two) into the amino-terminal region of β-galactosidase results in localisation of β-galactosidase to the

nucleus. Underwood and Fried (1990), investigated the effects of single and double amino acid substitutions on NLS-2 function. In general, substitution of any of the basic amino acids (K=lys, H=his and R=arg) by non-charged or acidic residues resulted in fusion proteins being maintained in the cytoplasm, although there were exceptions.

The actual mechanism of nuclear localisation remains to be elucidated, although presumably the localisation determinants described above act as ligands either for a nuclear pore receptor or for a chaperone.

1.3.3. Assembly of the Ribosome

Equimolar amounts of each ribosomal protein, and of each rRNA species, are present within the *S. cerevisiae* ribosome. Except for ribosomal protein L7/L12, which is present in fourfold excess over the other ribosome constituents, the same is true of *E. coli* (Hardy, 1975). In *E. coli* there is also another slight anomaly, in that proteins L26 and S20 (components of the large and small subunits respectively) are identical. The stoicheiometries of S20 (0.8) and L26 (0.2), correspond to a total of one polypeptide per 70S ribosome (Weber, 1972). Thus, it is likely that the protein is distributed between the two subunits when the 70S ribosome dissociates.

Assembly of active 30S and 50S subunits of the *E. coli* ribosome *in vitro* has aided the elucidation of the process in this organism (reviewed by Zimmermann, 1980). In contrast, details of the assembly of eukaryotic ribosomes must be largely inferred, either from the results of *in vivo* labelling experiments, or from studies in which the ribosome is gradually disassembled by increasing the salt concentration such that proteins are stripped off in sequence.

The course of assembly of ribosomal proteins into *S. cerevisiae* ribosomes was studied by Kruiswijk et al (1978). During, or immediately after, transcription of the rRNA precursor, many ribosomal proteins, together with several non-ribosomal proteins, associate with the primary transcript. The resulting ribonucleoprotein complex contains most of the ribosomal proteins present in the mature ribosome. During maturation the non-ribosomal proteins, like non-rRNA sequences, are removed from the pre-ribosomal particle in a number of discrete steps. However, in the case of both the large and small subunits, a number of ribosomal proteins do not associate with ribonucleoprotein particles until a relatively late stage during assembly. Indeed, several of the late associating proteins of the large (60S) subunit could not be found in 66S precursor

ribosomal fractions isolated from the nucleus, thus suggesting that they assemble into the ribosome in the cytoplasm. Late assembly of certain ribosomal proteins is also seen in higher eukaryotes.

Ribosomal proteins L4, L8, L10, L16 and L25 appear to be the first to assemble with rRNA *in vivo* (Kruiswijk et al, 1978). Further, a small subset of ribosomal proteins, that includes most of the above, remains part of a complex with 25S and 5.8S rRNAs in the presence of 0.5M LiCl, a treatment that causes most the large subunit ribosomal proteins to dissociate. This subset of core proteins can be stripped from rRNA by further increasing the LiCl concentration (L25 is bound most tightly), and will reassociate in the reverse order as the salt concentration is reduced (El-Baradi et al, 1984). The region of 25S rRNA to which L25 binds has been determined by RNase protection experiments. Strikingly, the sequence/structure of this region is virtually identical to the sequence of *E. coli* 23S rRNA to which *E. coli* protein L23 binds, in fact so much so that L25 can bind there too. Thus L25 is likely to be the eukaryotic counterpart of L23, in spite of the limited degree of similarity between the two proteins at the amino acid level (El-Baradi et al, 1985). Similar results have been obtained for L15 of *S. cerevisiae* and L11 of *E. coli*. Both proteins bind to the same regions of either *S. cerevisiae* or mouse 25S rRNAs (El-Baradi et al, 1987). These results provide strong evidence for the evolutionary conservation of the ribosome.

In the *E. coli* ribosome, and presumably in the eukaryotic ribosome (although this has yet to be confirmed), rRNA appears to be at the core with the proteins forming the periphery. At the interface between the large and small subunits, which is where protein synthesis takes place, there are relatively few proteins, and direct contact between rRNAs present in each subunit has been inferred. The active ribosome is highly flexible, a feature that would appear to be a prerequisite for function (Nygard and Nilsson, 1990)

1.3.4. Transcriptional Regulation of Ribosomal Protein Genes

The overall rate of synthesis of *S. cerevisiae* ribosomal proteins is modulated such that ribosomal proteins accumulate only to the extent that is needed. In particular, transcription rates are modulated in response to changes in the availability of nutrients, and in response to heat-shock (Pearson and Harbor, 1980; Keif and Warner, 1981; Gorenstein and Warner, 1976). Moreover, rates of transcription of most ribosomal protein genes are modulated in a coordinated manner (see Section 4.1.). Several families of coordinately expressed *S. cerevisiae* genes have family-specific short regulatory motifs upstream of the co-regulated genes (upstream activation sequences or UAS). Examples of this are genes involved in amino acid biosynthesis (Hinnenbusch, 1988),

and the expression of mating type (Miller et al, 1985). Similarly, coordinately expressed ribosomal protein genes have common upstream elements. For example, Teem et al (1984), comparing sequences upstream of eight ribosomal protein genes, found various regions of sequence similarity. The most notable of these was designated HOMOL 1, an element closely related to a 12 bp consensus, that was found upstream of 6 of the 8 genes. The distance of HOMOL 1 from the respective translation initiation codons ranged between 250 and 350 bp. Moreover, it was always followed closely by a T-rich domain. Neither feature was seen in any of the non-ribosomal protein genes examined. Of the two ribosomal protein genes that showed no evidence of an HOMOL 1 motif, the sequences of one extended only 147 bp 5' of the initiation codon. However, the second gene (*TCM1*) had been sequenced as far as 444 bp 5' of the initiation codon (see Section 3.3.2.2.).

Later comparison of 20 ribosomal protein gene sequences revealed a second conserved upstream element, the 12 bp RPG-box (**r**ibosomal **p**rotein **g**ene-box; Leer et al, 1985). When both HOMOL 1 and the RPG-box are present upstream of a ribosomal protein gene, the RPG-box is always 3' to HOMOL 1, and they are usually separated by only a few bp. The RPG-box can occur in both orientations, and may be present in more than one copy.

In order to determine the functional significance of these observations, Rotenberg and Woolford (1986) analysed in detail the promoter of the *RP39a* gene. *RP39a* upstream sequences, variously manipulated, were fused to a reporter gene, *lacZ* of *E. coli*. Following transformation of yeast, levels of β-galactosidase activity, as well as mRNA levels were measured. A sequence from between -256 and -170 upstream of the initiation codon, and containing HOMOL 1, an RPG-box and a T-rich domain, was sufficient to confer *RP39a*-like expression on *lacZ* (Rotenberg and Woolford, 1986). Similar sequence elements were found to be important for the expression of the ribosomal protein gene, *CYH2* (Schwindinger and Warner, 1987). There also seems to be some redundancy in that the T-rich domain alone can facilitate transcription of the ribosomal protein gene *CRY1*, although at significantly lower levels (Larkin et al,1987).

Woudt et al (1986, 1987) re-evaluated upstream regions of yeast ribosomal protein genes in the context of a larger data set. They found the elements originally designated as HOMOL 1 and RPG to be related to one another. On this basis, a new consensus sequence known as UAS_{rpg} (**u**pstream **a**ctivation **s**equence of **r**ibosomal **p**rotein **g**enes) was deduced. Most ribosomal protein genes contain two UAS_{rpg} elements, although some contain only one. Critically, a consensus UAS_{rpg} alone, introduced by means of a synthetic oligonucleotide into a transcriptionally-inactive reporter gene, is functional (Herruer et al, 1987).

HOMOL 1- and RPG-like elements (one of each) are found upstream of *TEF1* and *TEF2*, both of which encode the translation elongation factor EF-1α. A 150 kD protein that protects each element has been characterised and named TUF (Huet et al, 1985; Huet and Sentenac, 1987). This is further evidence that HOMOL 1 and RPG are variants of the same element. Moreover, UAS_{rpg}s are seen to be not only responsible for coordinate synthesis of ribosomal proteins, but for synthesis of the yeast translation apparatus in general.

Sequences with homology to UAS_{rpg} have been identified upstream of several other yeast genes, including some genes encoding glycolytic enzymes, and *MATα*. Proteins that protect these sequences have been variously described as GRF1 (general regulatory factor 1) or RAP1 (repressor/activator factor). It now seems clear that TUF, RAP1 and GRF1 are the same protein, suggesting that regulation of ribosomal protein gene expression is merely one facet of a potentially wide-ranging ability to respond to environmental conditions (see Warner, 1989). However, not all ribosomal protein genes appear to be subject to control by TUF. For example *TCM1* lacks a UAS_{rpg}, and is controlled instead by a so-called UAS_T element, the target of a protein known as TAF (Hamil et al, 1988). TAF seems also to be involved in control of more than ribosomal protein genes, since the UAS_T sequence is identical to the consensus ABF1 binding site (ARS-binding factor; ARS=autonomously replicating sequence). Moreover, the apparent molecular weights and chromatographic properties of the two proteins are the same (see Brindle et al, 1990). As well as binding to ARS sequences, ABF1 binds to a different UAS element upstream of one of the genes that encodes enolase, *ENO2* (Brindle et al, 1990; Section 3.3.1.). Interestingly RAP1/TUF/GRF1 also binds to a UAS element upstream of *ENO2*, as well as to one upstream of *ENO1* (Brindle et al, 1990). Finally, RAP1 and ABF1 (and therefore probably TUF and TAF) show extensive sequence similarity, supporting a notion that the two regulatory proteins have similar functional roles (see Brindle et al, 1990).

1.3.5. Post-Transcriptional Regulation of Ribosomal Protein Gene Expression

The 52 ribosomal protein genes of *E. coli* are organised into at least 20 operons comprising from 1 to 11 genes. Several of the operons contain, in addition, genes that encode relevant transcription and translation factors (see Nomura et al, 1984). Rates of synthesis of ribosomal proteins respond coordinately to changes in environmental conditions. However, comparison of sequences from the promoters of ribosomal protein operons does not suggest conserved response elements. Moreover, although provision

of additional copies of individual ribosomal protein genes can result in a proportionate increased in the respective mRNAs, ribosomal proteins levels do not inevitably follow suit. Ribosomal protein abundance is thus regulated post-transcriptionally in *E. coli*. The effect is predominantly at the level of translation, and occurs by an autoregulatory mechanism. Autoregulation is mediated by only a subset of the ribosomal proteins, as can be demonstrated by adding individual proteins to a coupled transcription-translation cell-free extract, in which ribosomal protein operons are used as templates. Each regulatory protein affects the translation only of the polycistronic mRNA by which it is encoded.

A common feature of the regulatory proteins is that they are among the first proteins to bind to rRNA *in vitro*, and therefore presumably *in vivo*, and do so strongly. It has thus been postulated that translational regulation is dependent upon the balance between incorporation of a regulatory protein into the ribosome, and its binding to the polycistronic mRNA that encodes it. Indeed, translational repression of the bicistronic L11 operon *in vitro* by ribosomal protein L1 is abolished by excess 23S rRNA (Nomura et al, 1980), and the target for L1 within the L11 mRNA shares common structural motifs with the binding site for L1 within 23S rRNA (Thomas and Nomura, 1987; Kearney and Nomura, 1987). Further, both *Serratia marcescens* and *Proteus vulgaris* have an L1-like protein that can regulate expression of a chimeric operon containing the L11 target site, indicating considerable evolutionary conservation of the regulatory mechanism (Sor and Nomura, 1987).

How does binding at a unique site within a polycistronic mRNA affect translation of every gene within it? The L11 gene precedes the L1 gene within the L11 operon. The binding site for L1 lies upstream of both genes, and is intimately associated with a Shine-Delgarno sequence (Thomas and Nomura, 1987). Thus, binding of L1 can inhibit translation of L11. Moreover, deletions that disrupt L11 translation initiation also abolish translation of L1 (Baughman and Nomura, 1983). Thus, it would appear that translation of L1 is dependent upon previous translation of L11. Such translational coupling has been demonstrated for other ribosomal protein operons (see Nomura et al, 1984). The various interactions that occur between the autoregulatory ribosomal proteins, their mRNAs and rRNA have been discussed in detail by Draper (1989).

In *S. cerevisiae* too, there is evidence for post-transcriptional regulation of ribosomal protein genes, in that protein levels do not necessarily parallel mRNA levels. For example, although introduction of extra copies of *tcm1*, encoding ribosomal protein L3, causes an increase in L3 mRNA abundance that corresponds to the gene dosage, little

effect on L3 protein levels is observed (Pearson et al, 1982). This was taken initially as evidence of a translational control mechanism analogous to that occurring in *E. coli*, although polysome profiles were not investigated (see below).

Results similar to the above have been obtained in the case of a second ribosomal protein gene, *RP51A*. Here, however, the location of the sequences that mediate post-transcriptional regulation has been identified. Instead of being in the 5'-untranslated region, as would have been expected if translational regulation were involved, they occur within the coding region, between codons 1 and 63, and it has been suggested that regulation of rp51a levels is mediated by proteolytic cleavage (Gritz et al, 1985). Further, no enhancement of translation of *RP51b* mRNA is observed in cells deleted for *RP51a* (both genes encode the same protein; Abovich and Rosbash, 1984; Abovich et al, 1985). Maicas et al (1988) compared the polysome distribution of *tcm1*, *cyh2* and *cry1* mRNAs (encoding L3, L29 and rp59 respectively), in strains carrying extra copies of each gene, with the distribution in normal cells. No differences were observed, again indicating that the rate of translation initiation is unaffected by gene dosage. Although this conflicts with the data of Warner et al (1985), who did observe an altered polysome distribution of *tcm1* mRNA in a similar experiment, Maicas et al (1988) point out that they analysed more fractions per polysome gradient, resulting in greater resolution and hence sensitivity.

It is now generally believed that translational regulation of ribosomal protein levels is relatively insignificant and that control of ribosomal protein turnover plays the major role (Warner, 1989). Very short periods of labelling are necessary to measure the effect successfully, since the half-lifes of some ribosomal proteins (e.g. L29) fall to as low as a few seconds at high gene dosage (Maicas et al, 1988). The ability of the cell to detect so rapidly that a protein is being over-produced is quite remarkable, although it must be pointed out that decay as defined by Maicas et al (1988) is any event that prevents a protein migrating to its normal position in a 2-D polyacrylamide gel. This could result from a single proteolytic cleavage, or a chemical modification that marks a protein out for decay. The breakdown of a protein to small peptides or to amino acids would most likely occur more slowly . Neither the mechanism by which decay takes place, nor the compartment in which it occurs, is known, although it seems plausible that a protein should be at least marked for degradation in the nucleus, where ribosome assembly takes place. Otherwise it is difficult to imagine how a cell could know that a protein was in excess.

In the case of at least one ribosomal protein (L32), extra copies of the gene result in an inhibition of splicing of the corresponding pre-mRNA (Dabeva et al, 1986; see also

Section 4.3.). Since splicing is a nuclear event, it is again possible to envisage control dependent upon assembly.

1.3.6. Transport of the Ribosomes to the Cytoplasm

Assembled ribosome subunits are transported from the nucleus to the cytoplasm, presumably through nuclear pores. Whether this constitutes yet another level at which regulation of ribosome function occurs is unknown.

1.3.7. Ribosomal Protein Nomenclature

Nomenclatures for ribosomal proteins of different organisms do not necessarily correspond. In addition, three separate nomenclatures for *S. cerevisiae* ribosomal proteins are in common use. At risk of confusion, the nomenclature used here is always that of the authors of the literature cited. The different nomenclatures have arisen due to the use of different 2-D gel systems for the separation of ribosomal proteins, and unfortunately they have only been partially correlated. The most recent comparison of the different nomenclatures was made by Planta et al (1986), and is reproduced in Appendix 2. In general, the prefix S indicates a component of the small subunit, L a component of the large subunit. Where Y is used, this is merely to indicate a yeast protein. Data concerning ribosomal protein genes that have been sequenced, including information as to whether they are duplicated, whether they contain introns, whether they have UAS_{rpg} elements, and whether there are known counterparts in other organisms, are given in Appendix 3.

CHAPTER 2

MATERIALS AND METHODS

2.1. *E. coli* Strains

The *E. coli* strains used in this work are all derivatives of *E. coli* K12. They are listed below with their genotypes.

Strain	Genotype	Reference/ Source
JM101	*sup*E, *thi*, Δ(*lac-proAB*),F' *traD36*, *proAB*$^+$, *lacI*^{q}, *lac* ZΔM15.	Messing (1979)
NM522	*supA*, *thi*, Δ(*lac-proAB*), Δ*hsd5*(R$^-$, M$^-$),F' *proAB*, *lacI*^{q} ZΔM15.	Gough and Murray (1983)
NM621	*hsdR*, *mcrA*, *mcrB*, *supE44*, *recD1009*.	Whittaker et al (1988)
P2PLK-17	*hsdR*$^-$*M*$^+$, *mcrA*, *mcrB*, *lac*, *supE*, *gal*, P2.	Stratagene USA

2.2. *S. cerevisiae* Strains

The *S. cerevisiae* strains used in this work are shown below with their genotypes.

Strain	Genotype	Source
DBY746	α, *his*3, *ura*3-52, *leu*2-3, *leu*2-112, *trp*1-289	David Botstein
DJy36	a, *ura*3-1 or *ura*3-52, *ade*, *prp*2-1, Gal$^+$	Derek Jamieson
DJy89	a, *leu*2, *ura*3-52, *his*3, *ade*1/2, Gal$^+$	Derek Jamieson

2.3. Plasmids and Bacteriophages

Plasmids and bacteriophages used in this study, other than those whose construction is described elsewhere in this thesis, are listed below.

Plasmid/ Bacteriophage	Description	Source/Reference
pSP65R	A fragment of the *S. cerevisiae* 18S rRNA gene in pSP65	Santiago et al (1986)
pSPACT9	A fragment of the *S. cerevisiae* actin gene in pSP64	Bettany et al (1989)
Bluescript IIKS$^+$		Stratagene USA
M13mp18/19		Messing (1983)
EMBL3		Frischauf et al (1983)

2.4. Oligonucleotides

Oligonucleotides used in this study are shown below, along with their application.

Oligonucleotide (5' to 3')	Use
GTGTGTTGATAAGCAGTTGCTTGGTT	Probe for *ENO1*
TCATGGTCATAGCTGTT	Probe-primer (strand-specific M13 probes; BRL)
GTAAAACGACGGCCAGT	Universal sequencing primer (-20)
GTTTTCCCAGTCACGAC	Univeral sequencing primer (-40)
AACAGCTATGACCATG	Reverse sequencing primer
GTGAACACCTTCGTTAGC	Primers for sequencing of YL6a
GGTAGAGTTGACAAACC	"
GGGTGTGATGACATTGC	"
CAACCATGTGATTGTCG	"
CCTTCAATATCATTACC	"
CACGGTTACATTCGTGG	"
ATTGTTTGCTTTGAGG	"
CAACTCTGTACTTGTGG	"

2.5. *E. coli* Growth Media and Culture

2.5.1. Media

D & M Salts: 28g K_2HPO_4, 8g KH_2PO_4, 4g $(NH_4)_2SO_4$, 1g sodium citrate, 0.2g $MgSO_2.H_2O$ made up to 1 litre with distilled H_2O.

Minimal Agar: 20g Bacto-agar (Difco), 25ml D & M salts, made up to 1 litre with distilled H_2O.

L-Broth: 10g Bacto-tryptone (Difco), 5g yeast extract (Difco), 10g NaCl, made up to 1 litre with distilled H_2O and adjusted to pH7.0 with NaOH.

L-Agar: As L-broth with the addition of Bacto-agar (Difco) to 1.5%.

BBL Broth: 10g trypticase peptone (BBL), 5g sodium citrate, made up to 1 litre with distilled H_2O.

BBL agar: As BBL broth with the addition of Bacto-agar to 1.5%.

BBL top agarose: As BBL broth with the addition of gel quality agarose to 0.7%.

2xYT Broth: 10g Bacto-tryptone (Difco), 10g yeast extract (Difco), 5g NaCl made up to 1 litre with distilled H_2O.

All growth media were sterilised by autoclaving at 120°C for 15 minutes at 15 psi.

Where required, minimal media were supplemented with sterile glucose (2mg/ml) and filter sterilised thiamine hydrochloride (vitamin B1; 20µg/ml). L-broth, L-agar and BBL-agar were supplemented with 10mM $MgSO_4$ for growth of bacteriophage lambda and its derivatives.

2.5.2. Antibiotics

When required either ampicillin, at a final concentration of 100µg/ml, or tetracycline, at a final concentration of 10µg/ml, were added to broth or agar.

2.5.3. Growth Conditions

L-broth was used for growth of most *E. coli* strains Incubation was at 37°C, with vigorous shaking.

To obtain single colonies bacterial cultures were streaked on L-agar, antibiotics being added as required, and incubated for approximately 16 hours at 37°C.

E. coli JM101 or *E. coli* NM522 were used for the propagation of M13 derivatives. These strains were grown in 2xYT for either single-stranded DNA preparation or for preparation of competent cells. Transformants of these strains were plated on Minimal agar containing glucose and thiamine.

E. coli NM621 or *E.coli* P2PLK-17 were used for the propagation of the bacteriophage vector EMBL3 and its derivatives. These strains were grown in L-broth and plated on either L-, or BBL-agar. The former results in small-, and the latter in large-plaque formation.

For long term storage, 500µl of a mid-log culture was added to 1ml of sterile 40% glycerol/2% peptone, and kept at -70°C indefinitely.

2.6. *S. cerevisiae* Growth Media and Culture

Complete medium (YPG)	2% glucose, 2% Bacto-peptone (Difco), 1% yeast extract (Difco).
Glucose minimal medium (GMM)	2% glucose, 0.65% yeast nitrogen base (without amino acids).
Minimal medium (MM)	0.65% yeast nitrogen base (without amino acids).
Solid medium	As either YPG or GMM, but with the addition of Bacto-agar (Difco) to 2%

All growth media were sterilised by autoclaving at 120°C for 15 minutes at 15psi. Ethanol was added to MM at a final concentration of 2% v/v to give EMM.

2.6.1. Nutrient Supplements

Amino acids (Sigma) and adenine (Sigma) were added to minimal media where required (at a final concentration of 50µg/ml), for growth of auxotrophic mutants.

2.6.2. Growth Conditions

Cultures were streaked on YPG or GMM agar with appropriate additions and grown for 3 days at 30°C. Plates could be stored for several weeks at 4°C.

Unless otherwise stated either YPG or GMM were inoculated with a small volume of a 2-day stationary culture and grown at 30°C, with vigorous shaking overnight, or until the appropriate phase of growth was attained. Growth was monitored by absorbance at 600nm.

For long term storage, 500µl of a mid-log culture was added to 1ml of sterile 40%glycerol / 2% peptone and kept at -70°C indefinitely.

2.7. Buffers and Reagents

Buffers and reagents, when appropriate, were autoclaved at 120°C for 10 minutes. Except where stated, all reagents were Analar grade, or ultra-pure reagents for molecular biology.

2.7.1. Electrophoresis

10 x TBE buffer	109g Tris-HCl, 55g boric acid, 9.3g $Na_2EDTA.2H_2O$ made up to 1 litre with distilled H_2O (pH8.3).
10 x Sequencing TBE	121.1g Tris-HCl, 51.36g boric acid, 3.72g $Na_2EDTA.2H_2O$ made up to 1 litre with distilled H_2O (pH8.3).
10 x TAE	48.4g Tris-HCl, 16.4g sodium acetate, 3.6g $Na_2EDTA.H_2O$ made up to 1 litre with distilled H_2O (pH8.2).
10 x MOPS	41.8g 3-[N-morpholino]-propane-sulphonic acid, 4.1g sodium acetate, 1.89g Na_2EDTA made up to 1 litre with distilled H_2O (pH7.2).
5 x gel loading buffer	10% Ficoll, 0.5% SDS, 0.06% bromophenol blue, 0.06% xylene cyanol FF.

2.7.2. Isolation of Nucleic Acids

TE	10mM Tris-HCl, 1mM Na_2EDTA (pH8.0).
Phenol	All phenol used in the purification of DNA or RNA was distilled. For DNA isolation it was then equilibrated, first with 1M Tris-HCl (pH8), then with TE. For RNA isolation it was equilibrated with distilled H_2O. Short-term storage was at 4°C; long-term storage at -20°C.
Phenol/Chloroform	A 1:1 mixture of phenol:chloroform.
TEN	10mM Tris-HCl, 1mM EDTA, 10mM NaCl (pH7.6.).
SB	0.2M Tris-HCl, 1M sorbitol, 0.1M EDTA, 0.1M β-mercaptoethanol (pH 7.5).
RNA denaturing solution	5 parts de-ionised formamide, 1.6 parts formaldehyde, 1 part 10 x MOPS buffer.
RNA extraction buffer	0.1M LiCl, 0.1M Tris-HCl pH 7.5, 0.01M DTT (freshly added).

2.7.3. DNA Manipulation

Restriction enzyme buffers	DNA cleavage was normally carried out in the presence of the appropriate BRL React™ buffer.
10 x Ligation buffer	0.5M Tris-HCl, 100mM $MgCl_2$, 100mM DTT, 500µg/ml BSA (fraction V; Sigma) (pH7.6).
10 x Kinase buffer	0.5M Tris-HCl, 100mM $MgCl_2$, 50mM DTT, 1mM spermidine, 1mM $Na_2EDTA.2H_2O$ (pH7.6).

10 x Nick-translation buffer	0.5M Tris-HCl, 100mM $MgSO_4$, 1mM DTT, 500µg/ml BSA (fraction V; Sigma) (pH7.5).
10 x Random-priming buffer	500mM Tris-HCl; 100mM $MgSO_4$; 1mM DTT; 0.6mM each of dATP, dGTP and dTTP (Pharmacia); 3µg/ml random $dp(N)_6$ (Pharmacia) (pH6.2).
Sucrose gradients	10% sucrose in 1M NaCl, 20mM Tris-HCl, 5mM $Na_2EDTA.2H_2O$ (pH8.0).
	40% sucrose in 1M NaCl, 20mM Tris-HCl, 5mM $Na_2EDTA.2H_2O$ (pH8.0).

2.7.4. DNA Hybridisation

100 x Denhardt's solution	2% BSA (fraction V; Sigma), 2% Ficoll (MW 400,000), 2% polyvinylpyrolidone (MW 360,000).
20 x SSC	3M NaCl, 0.3M sodium citrate (pH7.0).
20 x SSPE	3.6M NaCl, 200mM sodium phosphate, 20mM $Na_2EDTA.2H_2O$ (pH7.4).
Blot-denaturing solution	1.5M NaCl, 0.5M NaOH.
Blot-neutralising solution	1.5M NaCl, 0.5M Tris-HCl pH 7.2.

2.8. Nucleic Acid Isolation

2.8.1. Plasmid DNA

Large scale plasmid isolation was carried out by the alkaline-lysis method of Birnboim and Doly (1979). Further purification by CsCl/EtBr equilibrium centrifugation was as described by Sambrook et al (1989). On a small scale, the boiling method of Holmes and Quigley (1981) was used.

Plasmid DNA to be used for dideoxy sequencing was further purified after CsCl-gradient centrifugation. The DNA was treated with RNase A (pre-boiled for 10 minutes) at a final concentration of 20µg/ml for 30 minutes at room temperature. The solution was then extracted once with phenol/chloroform and twice with chloroform. The aqueous phase was made 0.3M in sodium acetate (pH5.2) and the DNA was precipitated by the addition of 2 volumes of ethanol.

2.8.2 M13 DNA

Isolation of single- and double-stranded M13 DNA was performed as described by Sambrook et al (1989).

2.8.3. Bacteriophage Lambda DNA

2.8.3.1. Large Scale

Large scale isolation of λ DNA was performed by a modification of the protocol of R. Buckland (EMBL, Heidelberg). A high titre stock was prepared from a plate lysate, using a fresh single plaque as starting material. This stock was used to infect a 200ml culture of the appropriate host cells ($OD_{660} \approx 0.3$) in a 2 litre flask in the presence of 10mM $MgSO_4$ and 0.2% maltose. The culture was incubated at 37°C with vigorous shaking until lysis was apparent (4-5hrs.), 1ml of chloroform and 8g NaCl were added, and shaking was continued for a further 10-15 minutes. Debris was pelleted (15000 x g, 15 minutes) and 20g of PEG 8000 was added to the supernatant. The PEG was dissolved at room temperature and the samples were placed at 4°C overnight. Precipitated phage were pelleted (15000 x g, 15 minutes) and resuspended in 11ml of phage buffer.

This phage suspension was vortexed with an equal volume of chloroform, the phases were separated by centrifugation (27000 x g, 10 minutes), and 10ml of the aqueous phase was added to 7.2g of CsCl. Once the CsCl had dissolved, the samples were decanted into Quick-seal tubes and spun for 16 hours at 49,000 rpm in a Beckmann Ti70 rotor at 20°C. The phage band was removed from the side through an 18-gauge syringe needle, and dialysed against TE (3 changes of 1 litre) overnight at 4°C. DNA was extracted from the phage by extracting sequentially with phenol, phenol/chloroform and chloroform. The DNA was dialysed against TE as above, and stored at 4°C. Yields for EMBL3 derivatives were generally 100-200μg.

2.8.3.2. Small Scale

Small scale isolation of λ DNA for the analysis of recombinants was by a plate-lysate method (Sambrook et al, 1989).

2.8.4. *S. cerevisiae* DNA

High molecular weight *S. cerevisiae* DNA was prepared by a modification of the procedure of Lautenberger and Chen (1987):

A 1 litre stationary-phase culture of DBY746 (more than 2 x 10^8 cells/ml) in YPG was harvested by centrifugation at 15,000 x g for 10 minutes at 4°C. The cells were resuspended in 10ml of TEN, and the pellet obtained after re-centrifugation at 15,000 x g for 10 minutes was weighed and resuspended in SB buffer at a concentration of 0.25g per ml. After mixing with 8mg of zymolyase 20000, the cell-suspension was incubated at 30°C, with occasional gentle mixing, for approx. 1 hour, or until more than 80% of the cells were spheroplasts as determined by light microscopy. At this stage it is of the utmost importance that mixing should be very gentle and that pipetting be carried out with wide-bore, cut-tip, pipettes. The spheroplasts were harvested by centrifugation at 2,500 x g for 5 minutes and gently resuspended in 20ml of SB. 40ml of TEN, 5ml of 10% SDS and 5mg of pre-boiled RNase A were added, and after incubation at 37°C for 2 hours, 2mg of Proteinase K was added. Incubation was continued for a further 2 hours at 37°C, followed by heating to 65°C for 30 minutes and then cooling to room temperature. An equal volume of phenol/chloroform was added and the tubes were rocked gently for 15 minutes. The phases were separated by centrifugation at 15,000 x g for 10 minutes after which the aqueous phase, containing the DNA, was extracted with an equal volume of chloroform. The DNA was then dialysed against TE (3 changes of 1 litre) overnight at 4°C. Following dialysis, the DNA was treated again with pre-boiled RNase A, at a final concentration of 10µg/ml, for 1 hour at room temperature; again extracted once with phenol/chloroform and once with chloroform; and again dialysed. If the concentration was not high enough at this stage, the volume was reduced by placing the dialysis bag on dry PEG.

2.8.5. *S. cerevisiae* RNA

In order to isolate undegraded RNA, all chemicals and receptacles for RNA work were made RNase free: glassware was baked at 300°C for a minimum of 4 hours and plasticware and reagents (except those containing Tris) were treated with 0.1% DEPC overnight and then autoclaved at 120°C for 15 minutes at 15psi. Tris-containing solutions were made by adding new ingredients to DEPC-treated distilled water that had been autoclaved to break down the DEPC.

RNA was isolated according to the procedure of Linquist (1981). Cultures in mid-log phase were harvested by centrifugation at 960g for 5 minutes at 4°C. Alternatively 2

volumes of ethanol, previously cooled to -70°C, were added to the culture and the mixture was stored at -20°C for up to 1 week, after which the cells were harvested as above. In either case the pellet was rapidly resuspended in 5ml of RNA extraction buffer and transferred to tubes containing 14g of glass beads (40 mesh), 1ml of 10% SDS, and 10ml of phenol/chloroform. This mixture was vortexed continuously for 5 minutes and then centrifuged at 12,000 x g for 5 minutes at 4°C. The aqueous phase was removed, re-extracted with 10ml of phenol/chloroform by vortexing for 1 minute, and separated by centrifugation at 12,000 x g for 5 minutes at 4°C. It was then made 0.3M in sodium acetate (pH5.2), 2.5 volumes of ethanol were added, and precipitation was allowed to occur for at least 1 hour at -20°C. The precipitate was recovered by centrifugation at 30,000 x g for 15 minutes and resuspended in 2ml of water. RNA was further purified by centrifugation through a CsCl cushion according to the method of Chirgwin et al (1979). The final RNA pellet was resuspended in water and stored at -70°C indefinitely.

2.9. *In Vitro* Manipulation of DNA

Restriction-endonuclease cleavage of DNA was carried out using conditions recommended by the manufacturers. Sub-cloning of DNA into plasmid and M13 vectors, and introduction of recombinant plasmids into *E. coli*, were performed essentially as described by Sambrook et al (1989). Isolation of DNA fragments was carried out by electro-elution, from LMT (low melting temperature) agarose gels in TAE gel buffer (Sambrook et al, 1989).

2.10. DNA Sequencing

Sequencing of either single- or double-stranded DNA was carried out by the dideoxy chain-termination method first described by Sanger et al (1977). The polymerase was usually Sequenase, a modified T7 DNA polymerase (Tabor and Richardson, 1987). Sequencing reactions were carried out using the conditions recommended in the Sequenase™ Version 2.0 manual supplied by the manufacturers (United States Biochemical Corporation). Reaction products were separated on 6% polyacrylamide gels, as described below. Where problems were encountered with secondary structure, TAQuence™, a modified Taq DNA polymerase (Innis et al, 1988) was used, since extension reactions can be carried out at a higher temperature: conditions used were again those recommended by the manufacturer (United States Biochemical Corporation).

2.11. Gel Electrophoresis

2.11.1. Agarose Gels

For most purposes the gel buffer was TBE. When purifying DNA fragments, however, TAE was used preferentially. A range of agarose concentrations (0.3%-2%) was used in order to resolve molecules across a range of lengths. Markers were usually wild-type λ DNA cleaved with *Hin*dIII and/or *Hin*dIII-*Eco*RI. DNA was visualised by UV fluorescence after staining the gel with EtBr (0.6μg/ml in running buffer).

Prior to electrophoresis, RNA was denatured by heating to 65°C for 15 minutes in RNA denaturing solution. It was separated on 1.0%-1.5% agarose formaldehyde gels (Sambrook et al, 1989), using MOPS buffer, with constant circulation from anode to cathode chambers in order to maintain a constant pH.

2.11.2. Polyacrylamide Gels

Products of DNA sequencing reactions were separated on denaturing polyacrylamide wedge gels: 6% acrylamide (acrylamide:bisacrylamide, 19:1), 7 M urea, in sequencing TBE. Polymerisation was initiated by the addition of 1ml of 10% ammonium persulphate and 20μl of TEMED to 150ml of 6% acrylamide/urea. Gels were allowed to polymerise for at least 2 hours before use and were then pre-run at 60W for 30-45 minutes in order to heat them to 45°C-50°C. Samples were denatured by boiling for 2 minutes in the buffer supplied by the manufactures, and then loaded onto gels with the aid of a sharks-tooth comb. Gels were run for 2-10 hours, soaked in 10% acetic acid/10% methanol for 30-60 minutes, and then dried onto Whatmann 3MM paper under vacuum. Autoradiography was carried out without intensifying screens at room temperature when ^{35}S was used, and with intensifying screen at -70°C when ^{32}P was used.

2.12. Nucleic Acid Hybridisation

2.12.1. Blotting

Genomic and plasmid DNAs, and RNA, were transferred to nylon membranes (Hybond-N), either by capillary action or vacuum blotting, as instructed by the manufacturer (Amersham UK). Hybond-N was also used for plaque-lifts of M13 and EMBL3 derivatives, again as recommended by Amersham. DNA was fixed to the membrane by UV treatment, RNA by baking for 2 hours at 80°C.

2.12.2. Radio-labelling of Nucleic Acids

^{32}P-labelled plasmid and phage DNAs were prepared by nick-translation (Rigby et al, 1977); labelled gel-purified DNA fragments by random-priming (Hodgson and Fisk, 1987). End-labelled oligonucleotide probes were prepared as follows: kinase buffer containing 100-200ng of the oligonucleotide, 50μCi γ[^{32}P]-dATP, and 1 unit of T4 kinase (final volume 10μl), was incubated at 37°C for 30 minutes. Probes were purified by Sephadex G50 chromatography, or G25 chromatography in the case of oligonucleotides, in columns prepared from disposable 1ml syringes. A fraction of each probe was counted to determine the specific activity in cpm/μg. On average, nick-translated probes had a specific activity of 10^7-10^8 cpm/μg; random-primed probes and oligonucleotide probes a specific activity of 10^8-10^9 cpm/μg.

In order to use single-stranded M13 derivatives directly as probes for Northern blots, and thereby to determine which cDNA strand contained coding information (see Chapter 3), single-stranded circular DNAs were made partially double-stranded by primer-extension (Hu and Messing, 1982). In the presence of limiting radioactive nucleotides, extension from a "probe-primer" (BRL) generates a radioactive DNA molecule in which the insert DNA remains single-stranded. Such a probe is **not** denatured before use. In practice, 50ng of a 17 nt probe-primer and 2μg of single-stranded DNA were resuspended in 11μl of React 2™ (BRL), and heated for 2 minutes at 90°C. Slow cooling to room temperature allowed annealing to occur. The primer was then extended for 1 hour at room temperature following the addition of 1μl of 0.1M DTT; 1μl of a mixture of dCTP, dGTP and dTTP (0.5M each); 1μl (10 μCi) of α-[^{32}P]-dATP (800 Ci/mmol), and one unit of Klenow DNA polymerase. Reactions were terminated by the addition of 1μl of 0.5M EDTA. The probe was separated from unincorporated nucleotides as above.

2.12.3. Hybridisation

Southern blots and plaque lifts to be hybridised with either nick-translated or random-primed DNA probes were pre-hybridised at 68°C in 5 x SSPE, 5 x Denhardt's solution, 0.5% SDS and 20μg/ml sonicated and denatured salmon sperm DNA, for at least 1 hour before addition of the probe. Northern blots to be hybridised with either type of probe were treated similarly, with the exception that the pre-hybridisation solution contained 50% (v/v) formamide, and the temperature was 42°C. Hybridisation was allowed to take place at the same temperature as for pre-hybridisation, and for a minimum of 16 hours.

Northern blots to be probed with the end-labelled *ENO1* oligonucleotide were treated essentially as above, except that the pre-hybridisation solution contained 40% (v/v) formamide and the hybridisation temperature was 37°C. Hybridisation was for only 2 hours.

Blots were washed in 2 x SSC, 0.1% SDS, at room temperature alone where low stringency was necessary, or with additional washes in 0.2 x SSC, 0.1% SDS at 68°C when higher stringency was required.

Blots probed with the *ENO1* oligonucleotide were washed in 6 x SSC at room temperature for 15 minutes and then again in 6 x SSC at 42°C for a further 15 minutes.

CHAPTER 3

SEQUENCING OF 12 UNIDENTIFIED cDNAs

3.1. Introduction

3.1.1. Overview of Analytical Strategy

My initial aim in this project was to sequence the cDNAs analysed by Santiago (1986), and to attempt to identify their encoded products. The cDNAs were originally cloned, as described in Section 1.1., by C-tailing and insertion into a G-tailed *Pst*I site of the plasmid cloning vector pBR322. Since I desired to use a single-stranded DNA sequencing strategy, appropriate fragments of the cDNAs were sub-cloned into the vectors M13mp18 and M13mp19. Since M13 and its derivatives replicate as double-stranded DNA molecules, these can be used for subcloning purposes and for restriction-analysis of recombinant molecules, while the single-stranded form can be used for DNA sequencing by the Sanger dideoxy method (Sanger et al, 1977). *Pst*I fragments representing each cDNA contained terminal homopolymer tracts, which present problems for DNA polymerases *in vitro*. Thus each cDNA was mapped with respect to restriction enzyme sites compatible with those in the M13mp18/19 polylinker, so that suitable sub-fragments could be cloned. Single stranded DNA prepared from individual sub-clones was used as the substrate for dideoxy sequencing.

Computer analysis of the sequences was carried out using the Wisconsin (GCG) programmes (Devereux et al, 1984). In particular, nucleotide sequences were compared with sequences in the GenEMBL DNA database. Also, where an open reading frame was detected, the putative amino acid sequence was compared with sequences in the NBRF protein data base. These comparisons were carried out using the GCG programme "Wordsearch", which uses an algorithm similar to that of Wilbur and Lipman (1983). The output is a list of significant diagonals that can be displayed as alignments by the use of the programme "Segments". Alternatively, an Apple Macintosh-compatible programme, "DNA Strider", devised by Christian Marck (Gif-Sur-Yvette), was used. Of particular importance was open reading frame analysis, where each of the six frames was searched for methionine and stop codons. The former are indicated by half bars and the latter by full bars across the window.

In many cases the strand that contains coding information was deduced by using single-stranded probes (Hu and Messing, 1982; see Section 2.12.2.), prepared from the same DNA templates that were used for DNA sequencing. This involved making the single-stranded circular DNAs partially double-stranded over the vector component by extension from a "probe primer" in the presence of limiting radioactive nucleotides. The cDNA component remains in single-stranded form. Such probes were tested for their

ability to hybridise to Northern blots of RNA prepared from *S. cerevisiae* DBY746, the strain from which Santiago (1986) prepared his cDNA library. Where hybridisation occurred, this indicated that the sequence obtained from the corresponding clone should be that of the coding strand. The sequences presented here are those of the putative coding strands, sometimes the complement of the sequence obtained: the state of affairs with respect to any particular cDNA is indicated in the text.

S. cerevisiae has a distinct codon bias: more than 96% of the amino acids in highly expressed genes are encoded by a subset of 25 of the possible 61 codons (Bennetzen and Hall, 1982). This is reflected in the relative abundances of specific tRNAs (Ikemura, 1982). A similar phenomenon has been observed in *E. coli* (Ikemura, 1981). An example of the potential significance of codon bias for *S. cerevisiae* is given by the *PGK1* gene, which encodes the glycolytic enzyme phosphoglycerate kinase, and which is expressed at very high levels. *PGK1* has a pronounced codon bias, almost all of its codons being chosen from the restricted set. When 39% of the codons within a region encoding the N-terminal domain of the protein were replaced with more rarely used codons, levels of phosphoglycerate kinase dropped ten-fold (Hoekema et al, 1987). Thus, analysis of open reading frames for codon bias can help to distinguish which frame is most likely to represent the true one, and may also highlight frame-shifts due to sequencing errors. On the basis of their study of 110 individual sequences, Sharp et al (1986) compiled codon usage tables for *S. cerevisiae*. In such tables codon usage is expressed with respect to the total number of available codons for each amino acid. For example, in the case of an amino acid for which there are n possible codons, a codon that is used 50% of the time has a relative synonymous codon usage (RSCU) value of n/2. I incorporated the data of Sharp et al (1986) into our local GCG data-base, but modified the RSCU values such that the sum of all RSCUs for a particular amino acid is unity. Cluster analysis of codon usage in *S. cerevisiae* reveals that the majority of genes fall into one or other of two distinct groups (Sharp et al,1986). The group with the most distinct codon bias consists in the main of genes expressed at high levels. It was the RSCU values for this group of genes (see Appendix 5) that was used in this study. The GCG programme "codonpreference" (Gribskov et al, 1984) provides a graphical representation of codon usage for any specific open reading frame. Below each graph are shown the locations of rare codons, as determined from the codon frequency table for highly expressed genes, along with the open reading frames.

In the case of cDNAs representing short-half life mRNAs, the distribution of acidic and basic residues within the predicted amino acid sequences was examined. This is because ribosomal proteins for which mRNA half-lifes have been determined (Kim and Warner, 1983b) appear to fall within the unstable population (Brown, 1989). Further, with few

exceptions, ribosomal proteins range from basic to very basic, as determined by 2-dimensional electrophoresis, and roughly half are more basic than the core histones (Warner and Gorenstein, 1978).

Most of the cDNAs were sequenced on only one strand. This was thought to be adequate, since the reason for sequencing the cDNAs was to attempt to identify them using the available databases. However, although this approach is less time consuming than obtaining sequence information from both strands, it will inevitably result in the occasional mistake. The worst possible scenario is that false stop codons will be introduced, and therefore that regions of amino acid sequence will not be used to search the protein database. Thus sequencing only on one strand may preclude identification of a cDNA at the amino acid level.

Where I have estimated mRNA lengths during the course of this study, I have generally found them to be slightly larger than reported by Santiago (1986). This is most likely due to Santiago having used rRNAs as size markers [25S=3360 nt, 18S=1710 nt, 5.8S=158 nt and 5S=120 nt (Phillippsen et al, 1978b)]. Thus the size range within which most of the mRNAs lie was not covered in his studies (see Table 3.1). In this study BRL RNA size markers were used. These give a better distribution.

In the following Section I deal the cDNAs in turn, beginning with those derived from mRNAs with long half-lifes (Fig. 3.1.; Table 3.1.).

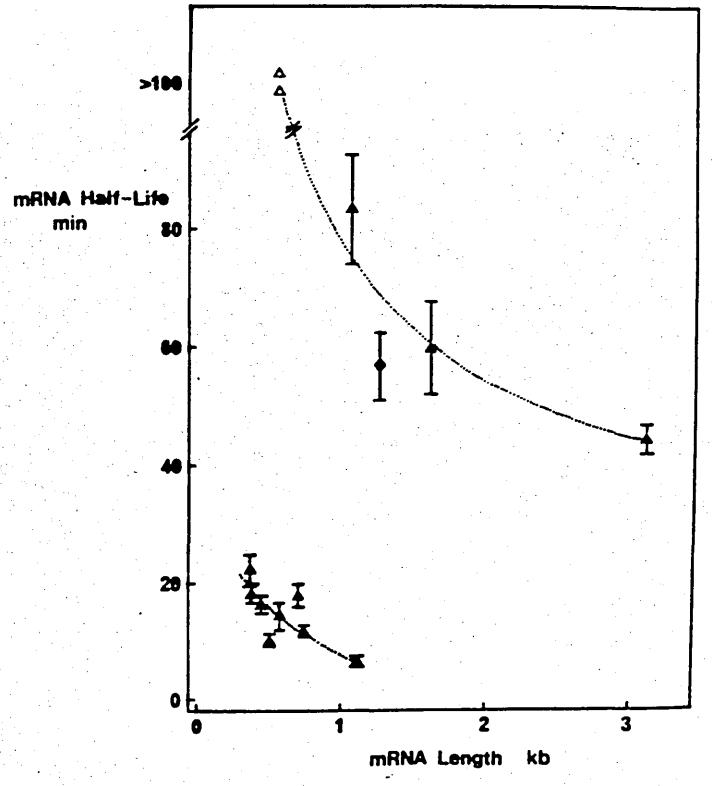


Figure 3.1. The relationship between mRNA length and half-life. Each point represents a different mRNA. Bars indicate standard deviations. mRNAs represented by open triangles have half-lifes too long to be accurately determined by the method of Santiago (1986).

mRNA	mRNA length (nt)	mRNAhalf-life (minutes)
Stable:-		
10	1250	56.9 +/- 6.0
22	3100	44.7 +/- 2.4
46	550	>100
74	1050	83.4 +/- 9.2
85	550	>100
Unstable:-		
9	440	16.5 +/- 1.5
11	700	18.0 +/- 2.2
13	740	12.1 +/- 1.0
19	550	15.3 +/- 2.7
39	380	18.3 +/- 1.5
82	370	22.3 +/- 3.1
90	1100	6.6 +/- 0.67
100	500	10.4 +/- 1.1

Table 3.1. The lengths and half-lifes of the mRNAs studied by Santiago (1986). The mean half-life of each mRNA is given in minutes together with the standard error. The half-lifes of mRNA46 and mRNA85 could not be measured accurately since they were greater than 100 minutes.

3.2. Results

3.2.1. cDNAs Representing Long Half-Life mRNAs

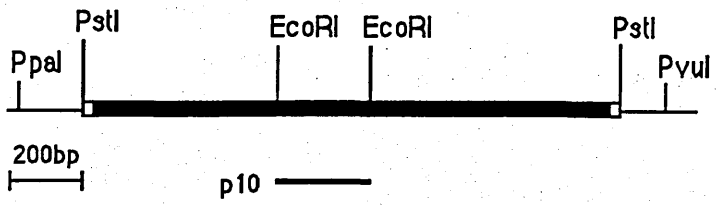
3.2.1.1. cDNA10

The mRNA from which cDNA10 is derived is estimated to have a half-life of 56.9 +/- 6.6 minutes and a length of 1250 nt (Santiago, 1986). The length of the cDNA insert is approx. 1.4 kb; the disparity is presumably due to inaccuracies in measuring the size of the mRNA, the cDNA, or both.

A restriction map of cDNA10 is shown in Fig.3.2.a. The internal 250bp *Eco*RI fragment was subcloned into M13mp19 and the subclone was named p10. The nucleotide sequence of the *Eco*RI fragment, together with a putative amino acid sequence deduced from it, is shown below the restriction map (Fig. 3.2.a.). The DNA sequence was used to search the GenEMBL database and was found to be very similar to those of the two yeast enolase genes, *ENO1* and *ENO2* (Holland et al, 1981). *ENO1* and *ENO2* are 95% identical at the DNA level over their coding regions. The differences between cDNA10 and each of these genes are indicated in Fig. 3.2.a. (see legend for details; the sequence presented is the complement of that obtained). At the DNA level, cDNA10 is 99.6% identical to *ENO2* and 98.0% identical to *ENO1* over the sequenced region; at the amino acid level it is 100% identical to *ENO2* and 99% identical to *ENO1*, due to a single amino acid difference, methionine for leucine, at amino acid 183 of *ENO1* (Fig. 3.2.a.). Thus cDNA10 would appear to encode *ENO2*. It can be further deduced that cDNA10 encodes *ENO2* by comparing the restriction maps of both enolase genes (Holland et al, 1981) with that of cDNA10 (Fig. 3.2.b.). There is significant restriction fragment polymorphism between the two enolase genes and it is evident that cDNA10 is equivalent to *ENO2* rather than *ENO1*. The slight differences in restriction fragment sizes observed between cDNA10 and *ENO2* are presumably due to inaccuracies in measurements in one or both studies.

3.2.1.2. cDNA22

The mRNA from which cDNA22 is derived is estimated to have a half-life of 44.7 +/- 2.4 minutes and a length of approx. 3100 nt (Santiago, 1986). The length of the cDNA insert is, however, only approx. 780 bp.




```
EcoRI
GAATTCATGATTGCTCCAACTGGTGCTAAGACCTTCGCTGAAGCCATGAGAATTGGTTCC      60
GluPheMetIleAlaProThrGlyAlaLysThrPheAlaGluAlaMetArgIleGlySer

GAAGTTTACCACAACTTGAAGTCTTTGACCAAGAAGAGATACGGTGCTTCTGCCGGTAAC     120
GluValTyrHisAsnLeuLysSerLeuThrLysLysArgTyrGlyAlaSerAlaGlyAsn

GTCGGTGACGAAGGTGGTGTTGCTCCAAACATTCAAACCGCTGAAGAAGCTTTGGACTTG     180
ValGlyAspGluGlyGlyValAlaProAsnIleGlnThrAlaGluGluAlaLeuAspLeu

ATTGTTGACGCTATCAAGGCTGCTGGTCACGACGGTAAGGTCAAGATCGGTTTGGACTGT     240
IleValAspAlaIleLysAlaAlaGlyHisAspGlyLysValLysIleGlyLeuAspCys
      EcoRI
GCTTCCT                                                           247
AlaSer
```

Figure 3.2.a. Restriction map (see also Fig. 3.2.b.) and sequence of cDNA10. Narrow lines represent pBR322; the bold line indicates the cDNA insert. Open boxes represent the GC tails at the end of the insert. p10 is the M13 subclone from which sequence information was obtained. The overlined*Eco*RI site in the DNA sequence represents the junction between the insert and the vector. The predicted amino acid sequence is shown below the DNA sequence. Underlined nucleotides mark the differences between the cDNA10 sequence and that reported for *ENO1,* in which C at position 6 is T; CA at positions 45 and 46 is TT; C at position 159 is T; and T at position 201 is C. The overlined C, at position 102, is a T residue in the reported sequence of *ENO2*. All of the differences are conservative, except for the one at position 46, which results in a methionine residue here instead of the leucine residue in *ENO1*. The sequence of *ENO1* and *ENO2* are from Holland et al (1981).

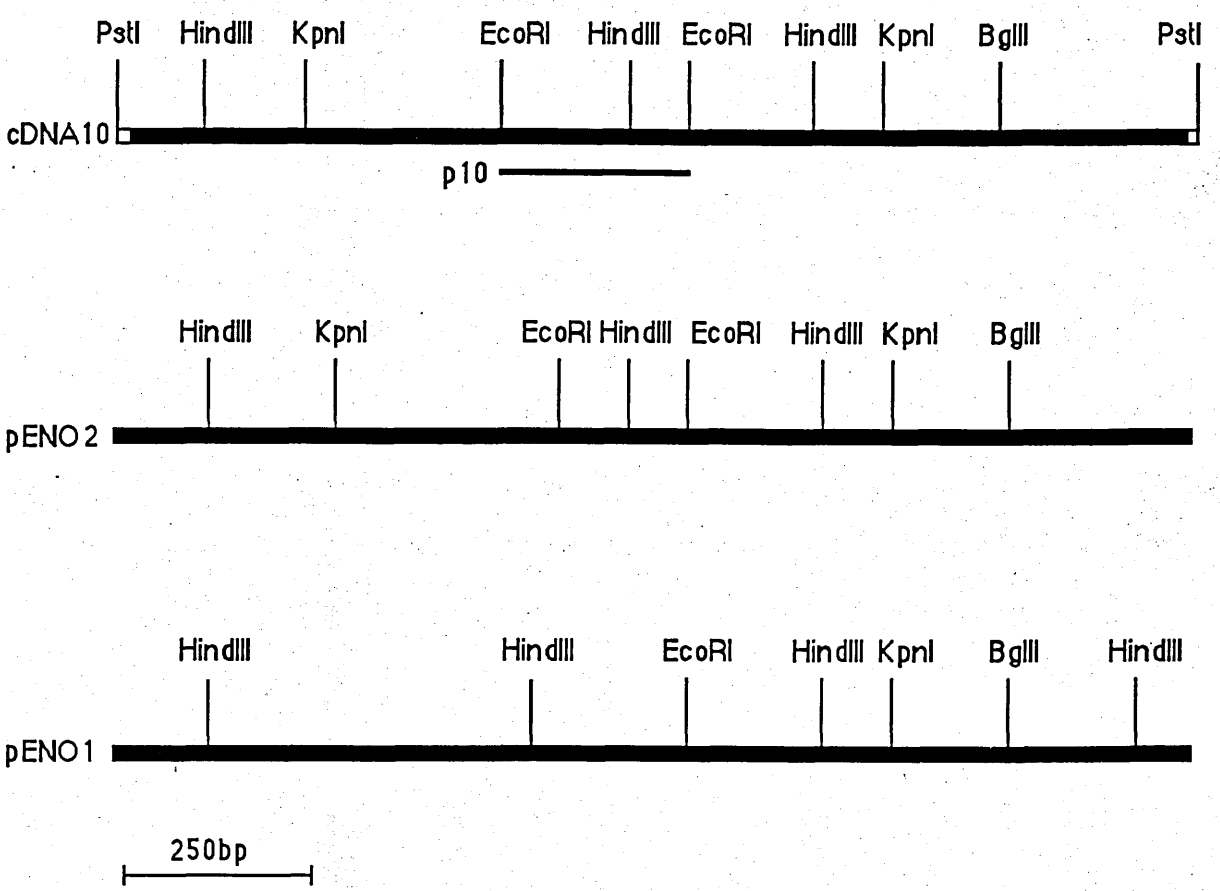
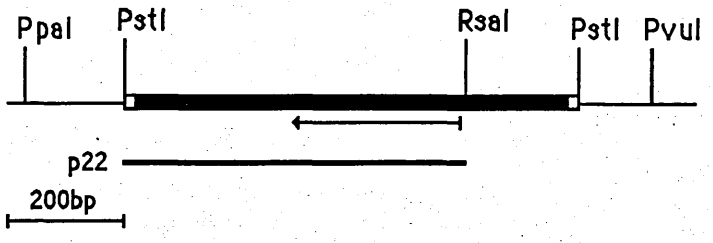


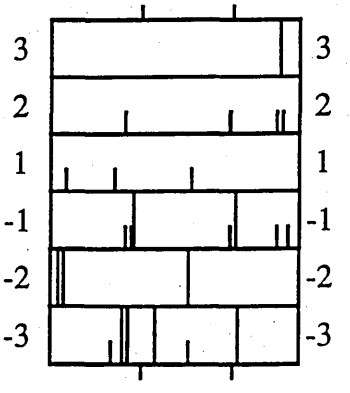
Figure 3.2.b. Restriction maps of cDNA10, *ENO1* and *ENO2*. Open boxes in the cDNA10 map represent the GC tails of the insert. Restriction maps of *ENO1* and *ENO2* are taken from Holland et al (1981).




```
GCTTTGTTCAAATTTATGTATTTTCGGAACGCCGGCGGGCGCGCCGCCCGTTGTCGAACC        60
AlaLeuPheLysPheMetTyrPheArgAsnAlaGlyGlyArgAlaAlaArgCysArgThr
 LeuCysSerAsnLeuCysIlePheGlyThrProAlaGlyAlaProProValValGluPro
  PheValGlnIleTyrValPheSerGluArgArgArgAlaArgArgProLeuSerAsnPro

CAGGCACGCATGCGGAGGCTTCATGTTCATTCACAACAAGAAACTGCAATACACGGTCAG        120
GlnAlaArgMetArgArgLeuHisValHisSerGlnGlnGluThrAlaIleHisGlyGln
 ArgHisAlaCysGlyGlyPheMetPheIleHisAsnLysLysLeuGlnTyrThrValArg
  GlyThrHisAlaGluAlaSerCysSerPheThrThrArgAsnCysAsnThrArgSerGly

GTCGCCCGACCGAATCCCGTCCTCGCGAGTTTCATGCTGGAGCAGTTCGGCGGCGCAGGG        180
ValAlaArgProAsnProValLeuAlaSerPheMetLeuGluGlnPheGlyGlyAlaGly
 SerProAspArgIleProSerSerArgValSerCysTrpSerSerSerAlaAlaGlnGly
  ArgProThrGluSerArgProArgGluPheHisAlaGlyAlaValArgArgArgArgAla

CGAGCTGGCCGCGGCCATGCGCTATTTCACCCAGGCGCTCGGGGAGGTCGATCCCGGCCG        240
ArgAlaGlyArgGlyHisAlaLeuPheHisProGlyAlaArgGlyGlyArgSerArgPro
 GluLeuAlaAlaAlaMetArgTyrPheThrGlnAlaLeuGlyGluValAspProGlyArg
  SerTrpProArgProCysAlaIleSerProArgArgSerGlyArgSerIleProAlaAla
                                  RsaI
CAAGGACATGCTGATGGACATCGCCACCGA                                      270
GlnGlyHisAlaAspGlyHisArgHisArg
 LysAspMetLeuMetAspIleAlaThr
  ArgThrCys *
```

Figure 3.3.a. Restriction map and sequence of cDNA22. Narrow lines represent pBR322 sequences; the bold line indicates the cDNA insert. Open boxes represent the GC tails at the end of the insert. p22 is the M13 subclone from which sequence information was obtained; arrows show the extent and direction of the sequence. Three open reading frames were found (Fig.3.3.b.); the putative amino acid sequence of each is shown below the DNA sequence.

Figure 3.3.b. Open reading frame map of the sequence obtained from cDNA22. Full bars represent stop codons and short bars methionine codons. Bars above the diagram represent 100 bp intervals.



The cDNA was devoid of sites for common hexanucleotide recognition-sequence enzymes. Thus an *Rsa*I-*Pst*I fragment (Fig.3.3.a) was subcloned into M13mp19 cleaved with *Hinc*II and *Pst*I. The complement of the sequence derived from this subclone (p22) is shown below the restriction map. Fig. 3.3.b. shows the open reading frames (ORFs) within the sequence; three major ORFs were identified. Their deduced amino acid sequences are shown in Fig.3.3.a., below the DNA sequence. The ORF in the third frame has a stop codon at position 252, which would indicate a 3' untranslated region of approx. 200 bases, assuming that the cDNA extends to the true 3' end.

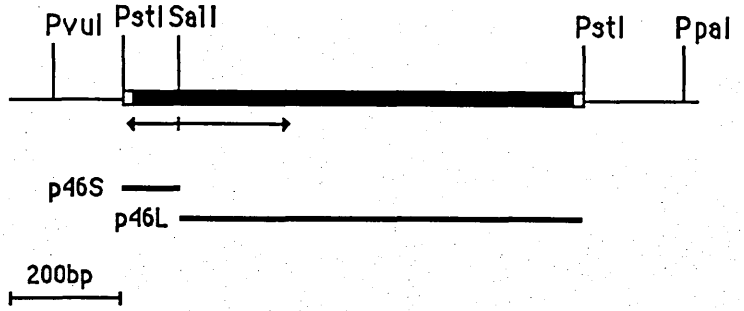
Further evidence that the sequence presented in Fig. 3.3.a. is that of the coding strand, is that single-stranded p22 DNA failed to hybridise to a Northern blot of yeast RNA. However, adequate controls were not carried out. All three major ORFs were analysed for their codon usage; no significant codon bias was observed in any (data not shown).

No significant homology with any DNA or amino acid sequence in the GenEMBL and NBRF databases was observed. Thus the identity and function of this gene remains to be established.

3.2.1.3. cDNA46

The mRNA from which cDNA46 is derived is estimated to a have a half-life of >100 minutes (Santiago, 1986), and a length of approx. 800 nt (this study). The length of the cDNA insert is approx. 840 bp; this disparity is presumably due to inaccuracies in measuring the size of the mRNA, the cDNA, or both.

A restriction map of cDNA46 is shown in Fig. 3.4.a. The two *Sal*I-*Pst*I fragments were subcloned into M13mp19 (p46S and p46L). Single-stranded p46L DNA hybridised to an mRNA on a Northern blot, while no such hybridisation occurred with p46S DNA (Fig.3.4.c.). This indicates that the sequence obtained from clone p46L is that of the coding strand, while that from p46S is non-coding. A sequence assembled from those of both sub-clones is shown in Fig. 3.4.a. The sequence is shown as being continuous across the *Sal*I boundary, but as indicated in the legend, the junction was not sequenced directly. The positions of methionine and stop codons, in all six frames, are shown in Fig. 3.4.b. Fig. 3.4.d. shows the codon bias within the three frames of the predicted coding strand. Frame 1 is the most obvious candidate for a protein-coding sequence, containing a complete open reading frame and the highest overall codon bias. However, the data are also suggestive of a frame shift from strand 2 to strand 1. The T of the stop codon in frame 2, which roughly coincides with where the frame shift would have occurred, is only 9 bp downstream of the *Sal*I site. The gel was very easy to read in this




```
CAACCAACCACAACTACATACACATACATACACAATGGTCGCTCAAGTTCAAAAGAAGCT      60
GlnProThrThrThrThrTyrThrTyrIleHisAsnGlyArgSerSerSerLysGluAla
                          SalI
CCAACTTTTAAGAAAACTGGGTGGTCGACGGTGTCTTTGACGAAGTCCTTGGACAAATAC     120
ProThrPheLysLysThrGlyTrpSerThrValSerLeuThrLysSerLeuAspLysTyr

AAGGGTAAGTACGTTGTCCTAGCCTTTATTCCATTGGCCTTCACTTTCGTCTGTCCAACC     180
LysGlyLysTyrValValLeuAlaPheIleProLeuAlaPheThrPheValCysProThr

GAAATCATTGCTTTCTCAGAGGCTGCTAAGAAATTCGAAGAACAAGGCGCTCAAG          235
GluIleIleAlaPheSerGluAlaAlaLysLysPheGluGluGlnGlyAlaGln
```

Figure 3.4.a. Restriction map and sequence of cDNA46. Narrow lines represent pBR322; the bold line indicates the cDNA insert. Open boxes represent the GC tails at the end of the insert. p46 is the M13 subclone used for sequencing; the direction and extent of sequence is indicated by arrows. The *Sal*I site used in cloning is indicated by overlining. The sequence is shown as continuous, even though no sequence data were obtained across the *Sal*I junction. However, when the two sequences were juxtaposed, an open reading frame was observed to run the full length of the combined sequence (Fig.3.4.b). Further, when Northern blots were probed with p46S and p46L, only the latter hybridised, indicating that the sequence obtained from p46L is that of the coding strand.

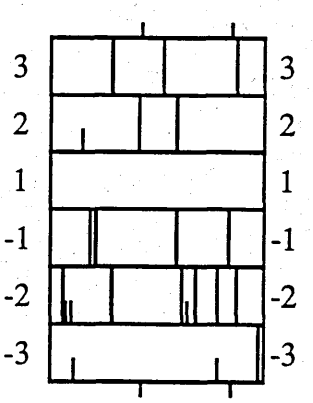


Figure 3.4.b. Open reading frame map of sequence obtained from cDNA46. Full bars represent stop codons and short bars methionine codons. Bars above the diagram represent 100 bp intervals.

Figure 3.4.c. Northern blot of total yeast RNA probed with p46S and p46L DNAs.

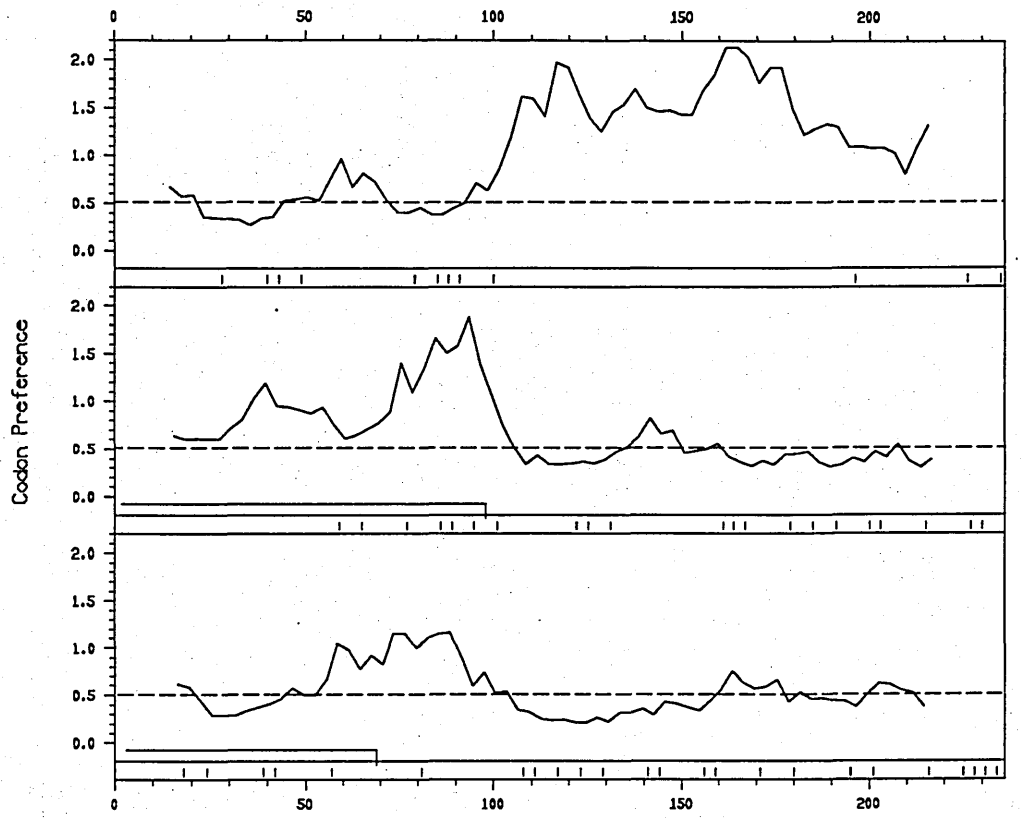


Figure 3.4.d. Codon preference plots for the sequence obtained from cDNA46. Upper, frame 1; middle, frame 2; lower, frame 3. The degree of departure fom the dotted line indicates the extent of deviation from random codon usage. The window length was set at 10 codons. Beneath each plot, open reading frames are indicated as boxed regions. Rare codons, ≤ 5% of a synonymous codon family, are indicated by short bars. Numbers along the top and bottom indicate distance in nucleotides.

```
  3 TTTTYTYIHNGRSS.............SKEAPTFKKTGWSTVSLTK.SLDK 39
    |.. :.   .|:|.              ...|| || |:  ...:.. |||.
 41 TDILWSASAQGKSAFSTSSSFHTPAVTQHAPYFKGTAVVNGEFKELSLDD 90

 40 YKGKYVVLAFIPLAFTFVCPTEIIAFSEAAKKFEEQGAQ 78
    :||||:|| |.||.||||||||||:|||: |..|.: ..:
 91 FKGKYLVLFFYPLDFTFVCPTEIVAFSDKANEFHDVNCE 129
```

Figure 3.4.e. Similarity between the putative amino acid sequence derived from cDNA46 (top) and that derived from the mouse gene, *MER5* (bottom). Full bars join identical amino acids. Two dots join amino acids that are not identical, but for which a pair of possible codons can be found that are related in two out of three positions. One dot joins amino acids that are not identical, but for which a pair of possible codons can be found that are related in one out of three positions. cDNA46 and *MER5* are 44.7% identical at the amino acid sequence level, or 57.9% when conservative changes are allowed, as determined by the GCG programme "Gap". Clearly, the identity is considerably higher over the second half of the sequence derived from cDNA46.

region. Thus, rather than a sequencing error, this may indicate that there is a small *Sal*I fragment missing, although such a fragment was not detected by restriction analysis. Unfortunately, this was not investigated, for example by end-labelling *Sal*I cleaved cDNA46 and separation of the fragments on a polyacrylamide gel. However, the possibility that the sequence presented (Fig. 3.4.a.) is contiguous cannot be ruled out (see below).

The amino acid sequence shown in Fig. 3.4.a., that of frame 1, was used to search the NBRF database. The closest match is shown in Fig. 3.4.e. The amino acid sequence derived from cDNA46 is 44.7% identical, or 57.9% similar when conservative changes are allowed, to that derived from *MER5*, a putative housekeeping protein preferentially expressed in mouse erythroleukemia cells (Yamamoto et al, 1989). Bearing the possibility of a frame-shift in mind, the homology between the two amino acid sequences is most significant after a large gap between amino acids 16 and 17 in the sequence derived from cDNA46, while the *Sal*I site overlaps codons 28-30. Thus, if there had been a frame shift, the similarity in the region between residues 17 and 28 must be fortuitous. However, when the amino acid sequence of frame 2 is compared with that derived from *MER5*, no significant homology is observed (data not shown).

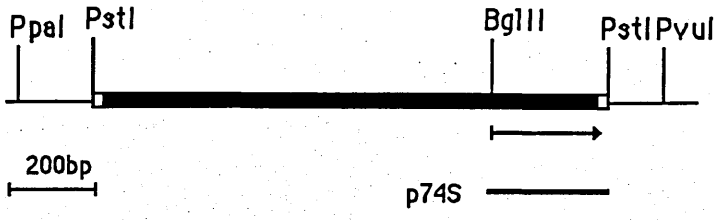
3.2.1.4. cDNA74

The mRNA from which cDNA74 is derived is estimated to have a half-life of 83.4 +/- 9.2 minutes (Santiago, 1986), and a length of approx. 1350 nt (this study). The length of the cDNA insert is approx. 1 kb.

A restriction map of cDNA74 is shown in Fig.3.5.a. The small *Bgl*II-*Pst*I fragment was subcloned into M13mp19 cleaved with *Bam*HI and *Pst*I (p74S). The extent of sequence information obtained from p74S is displayed below the restriction map. In the DNA sequence a complete *Bgl*II site is shown, although the actual sequence read was a composite of the 5' side of a *Bam*HI site and the 3' side of a *Bgl*II site.

The sequence obtained from subclone p74S is presumably that of the coding strand since a) single-stranded p74S DNA hybridises to a Northern blot of *S. cerevisiae* RNA (Fig. 3.5.c.), and b) the sequence runs into a polyA tail (data not shown).

Below the DNA sequence of p74S (Fig.3.5.a.) the three putative amino acid sequences are shown. Fig. 3.5.b. demonstrates the open reading frames diagrammatically. The




```
  BglII
  AGATCTGGGAACACTTACAACCTCATTTTACTACAAATGACGGGTTGGAAGTTATCTTCA      60
- ArgSerGlyAsnThrTyrAsnLeuIleLeuLeuGlnMetThrGlyTrpLysLeuSerSer
-  AspLeuGlyThrLeuThrThrSerPheTyrTyrLys *
-   IleTrpGluHisLeuGlnProHisPheThrThrAsnAspGlyLeuGluValIlePheLys

  AAGACGAAGAAGAAAAGGACCATCCTGTAAGAAAGTTGACCAACGCTAAGGGCGAATCAT     120
- LysThrLysLysLysArgThrIleLeu *
-
-   AspGluGluGluLysAspHisProValArgLysLeuThrAsnAlaLysGlyGluSerPhe

  TCAAGGTTGCTAGTATTGCTAATGCTCAAGTCCGTTAAAAAGATATAATATCACGAATTT     180
-
-
-   LysValAlaSerIleAlaAsnAlaGlnValArg *

  GTTACTAATTTAACATGCGTAAATACTTATTTCACATATATAATTCAAAGATTTATAAGT     240
  AAATGTCTTGTGCTAAT                                                257
```

Figure 3.5.a. Restriction map and sequence of cDNA74. Narrow lines represent pBR322 sequences; the bold lines indicates the cDNA insert. Open boxes show the GC tails at the end of the insert. p74S is the M13 subclone from which sequence information was obtained; the arrows show the extent and direction of the sequence. The *Bgl*II site used in cloning is indicated by overlining. The sequence obtained is presumably that of the coding strand since it has a poly(A) tract at its 3' end (data not shown) and, in addition, p74S DNA hybridises to *S. cerevisiae* RNA (Fig. 3.5.c.). Three open reading frames were found (Fig. 3.5.b.). The putative amino acid sequence of each is shown below the DNA sequence.

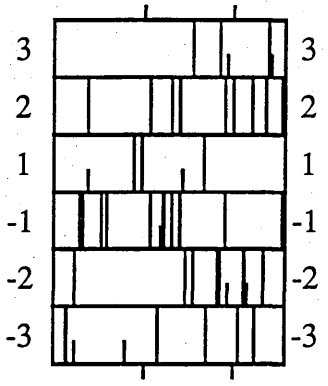


Figure 3.5.b. Open reading frame map of the sequence obtained from cDNA74. Full bars represent stop codons and short bars methionine codons. Bars above the diagram represent 100 bp intervals.

Figure 3.5.c. Northern blot of total yeast RNA probed with p74S DNA.

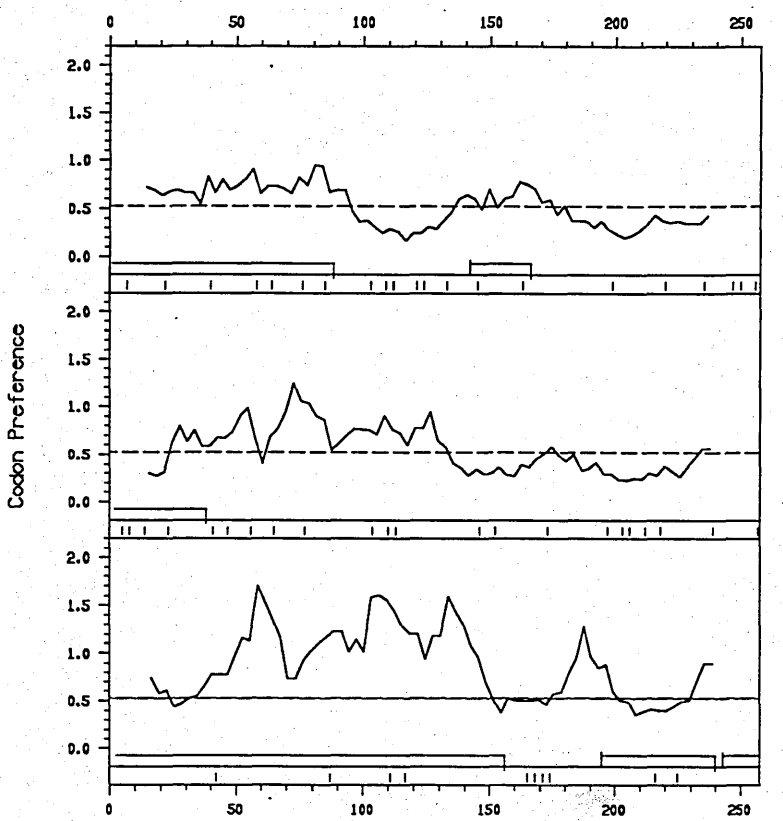


Figure 3.5.d. Codon preference plots for the sequence obtained from cDNA74. Upper, frame 1; middle, frame 2; lower, frame 3. The degree of departure fom the dotted line indicates the extent of deviation from random codon usage. The window length was set at 10 codons. Beneath each plot, open reading frames are indicated as boxed regions. Rare codons, ≤ 5% of a synonymous codon family, are indicated by short bars. Numbers along the top and bottom indicate distance in nucleotides.

codon bias of all three reading frames was examined (Fig.3.5.d.). The most striking is frame 3, where there is a reasonably high bias towards *S. cerevisiae* codon usage up to the position of the stop codon.

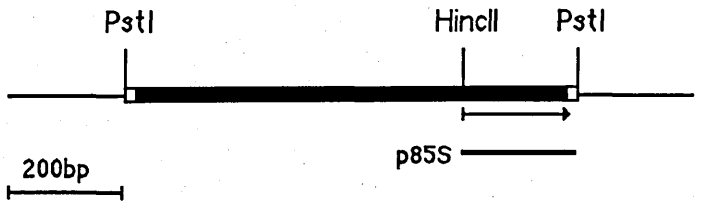
No significant homology with any DNA or amino acid sequence in the GenEMBL and NBRF databases was observed. Thus the identity and function of this gene remains to be established.

3.2.1.5. cDNA85

The mRNA from which cDNA85 is derived is estimated to have a half-life of >100 minutes (Santiago, 1986), and a length of approx. 620 nt (this study). The length of the cDNA insert is approx. 720 bp; this disparity is presumably due to inaccuracies in measuring the size of the mRNA, the cDNA, or both.

A restriction map of cDNA85 is shown in Fig. 3.6.a.. The orientation of the cDNA in the vector was not determined. The smaller of the two *Hinc*II-*Pst*I fragments was subcloned into M13mp19 (p85S) cleaved with the same enzymes. The sequence derived from p85S is presumably that of the coding strand since p85S hybridises with mRNA (Fig.3.6.c.), although no polyA tail was observed. This may be due to internal priming by oligo dT having occurred during first-strand cDNA synthesis. Thus there may be a longer 3'-untranslated sequence than is apparent from Fig. 3.6.a. Below the DNA sequence of p85S three ORFs are shown. These ORFs are shown diagrammatically in Fig.3.6.b. The strongest bias towards yeast codon usage was found in frame 1 (Fig.3.6.d.). Significantly the bias disappears after the putative stop codon. Thus it is possible that this frame contains the coding information, although the others cannot be ruled out.

No significant homology with any DNA or amino acid sequence in the GenEMBL and NBRF databases was observed. Thus the identity and function of this gene remains to be established.




```
 HincII
GTCGACTTTTCTGGTCAAAAGGCTGGTTTGATTGCTGCCAGAAGAACCGGTTTACTACGT      60
- ValAspPheSerGlyGlnLysAlaGlyLeuIleAlaAlaArgArgThrGlyLeuLeuArg
-  SerThrPheLeuValLysArgLeuVal *
-   ArgLeuPheTrpSerLysGlyTrpPheAspCysCysGlnLysAsnArgPheThrThrTrp
GGTTCTCAAAAGACCCAAGATTAATCTTTTTAATTTTGGTTTCTTTCCTTCTGTCATATT     120
- GlySerGlnLysThrGlnAsp *
-
-   PheSerLysAspProArgLeuIlePheLeuIleLeuValSerPheLeuLeuSerTyrTyr
ATTTTATCAATTTTCTTAAATAATTATATAATTTAACGACGTCCTATA                  168
-
-
-   PheIleAsnPheLeuLys *
```

Figure 3.6.a. Restriction map and sequence of cDNA85. The narrow lines represent pBR322 sequences; the bold lines indicate the cDNA insert. Open boxes show the GC tails at the end of the insert. Note that the orientation of the insert was not determined. p85S is the M13 subclone from which sequence information was obtained; the arrow shows the extent and direction of the sequence. The *Hinc*II site used in cloning is indicated by overlining. The sequence presented was concluded to be that of the coding strand on the basis of Northern blot analysis (Fig. 3.6.c.). Three open reading frames were found (Fig. 3.6.b.); the putative amino acid sequence of each is shown below the DNA sequence.

Figure 3.6.b. Open reading frame map of the sequence obtained from cDNA85. Full bars represent stop codons and short bars methionine codons. Bars above the diagram represent 100 bp intervals.

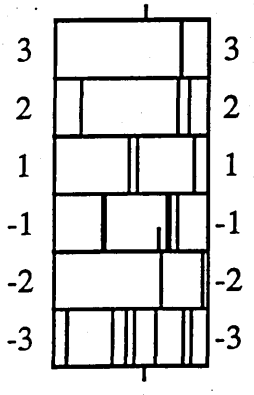


Figure 3.6.c. Northern blot of total yeast RNA probed with p85S DNA.

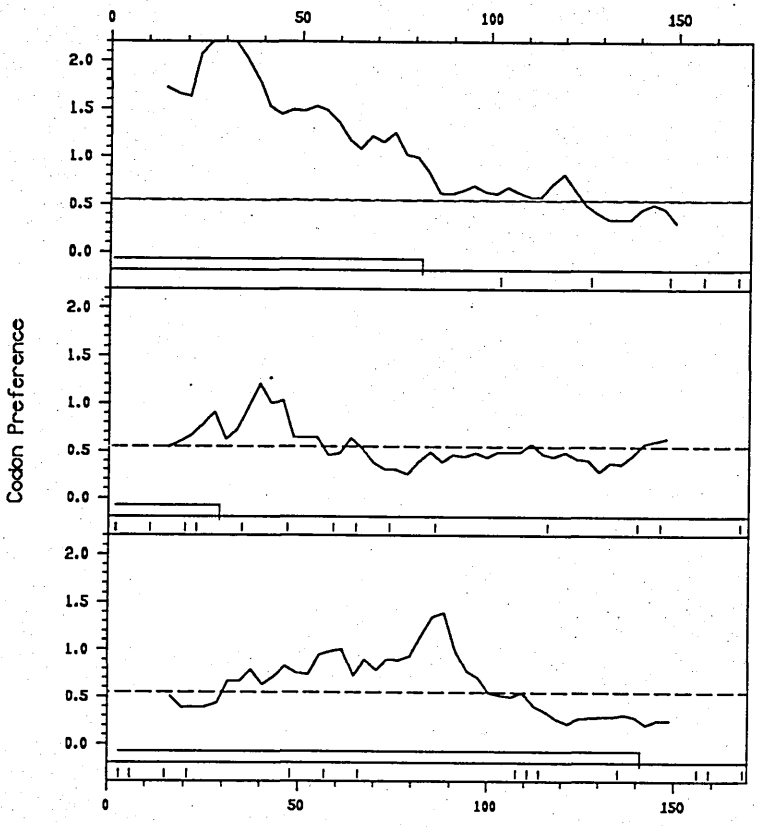


Figure 3.6.d. Codon preference plots for the sequence obtained from cDNA85. Upper, frame 1; middle, frame 2; lower, frame 3. The degree of departure fom the dotted line indicates the extent of deviation from random codon usage. The window length was set at 10 codons. Beneath each plot, open reading frames are indicated as boxed regions. Rare codons, ≤ 5% of a synonymous codon family, are indicated by short bars. Numbers along the top and bottom indicate distance in nucleotides.

3.2.2. cDNAs Representing Short Half-Life mRNAs

3.2.2.1. cDNA9

The mRNA from which cDNA9 is derived is estimated to have a half-life of 16.5 +/- 1.5 minutes (Santiago, 1986) and a length of approx. 620 nt (this study). The length of the cDNA insert is 700 bp; this disparity is presumably due to inaccuracies in measuring the size of the mRNA, the cDNA, or both.

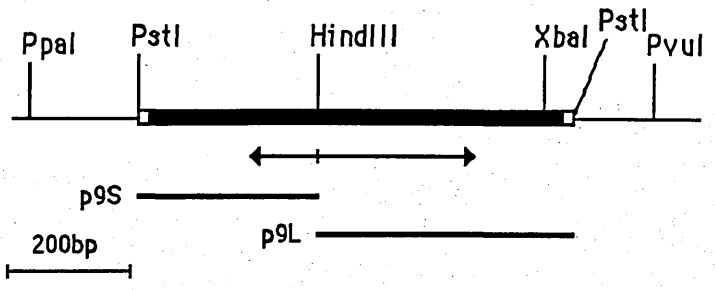
A restriction map of the cDNA is shown in Fig.3.7.a. The two *Hin*dIII-*Pst*I fragments were subcloned into M13mp18. The coding strand was determined by hybridising single-stranded DNA from clones p9S and p9L to Northern blots (Fig.3.7.c). Since only single-stranded p9L DNA hybridises, the sequence derived from this clone must contain coding information, and that from p9S non-coding. The combined sequence obtained from p9S and p9L is shown below the restriction map. Although the sequence is presented as being continuous across the *Hin*dIII site, this junction was not sequenced, as shown in the sequencing strategy (Fig. 3.7.a.). However, a continuous open reading frame is observed in frame 1 (Fig. 3.7.b.). The putative amino acid sequence is shown under the DNA sequence (Fig.3.7.a.). The codon bias of this ORF was determined and although it is high in general, a low value was observed between approx. 70 to 130 nt (Fig. 3.7.d.). Since sequencing was only carried out on one strand, this might indicate a frame shift.

The distribution of acidic and basic residues within the putative amino acid sequence is shown in Fig.3.7.e. Over the region sequenced it would appear to be relatively basic protein, providing this is the correct ORF. There are 9 acidic residues (4 aspartic acid residues and 5 glutamic acid residues), and 18 basic residues (9 lysine and 9 arginine residues).

No significant homology with any DNA or amino acid sequence in the GenEMBL and NBRF databases was observed. Thus the identity and function of this gene remains to be established.

3.2.2.2. cDNA11

The mRNA from which cDNA11 is derived was estimated to have a half-life of 18.0 +/- 2.2 minutes (Santiago, 1986), and a length of approx. 960 nt (this study). The length of the cDNA insert is only approx. 380 bp, however.




```
TTTGCAAGGCAGGAGAACCGCTTTGGTTCACGATGGTCTAGCCTAAGAGGTTTGAGAGAA      60
PheAlaArgGlnGluAsnArgPheGlySerArgTrpSerSerLeuArgGlyLeuArgGlu
                          HindIII
TCTAGGAAGGCTATATCTATAGTTAAGCTTTATTGGTTGTTTTGGTCAGCTCTGTTACTG     120
SerArgLysAlaIleSerIleValLysLeuTyrTrpLeuPheTrpSerAlaLeuLeuLeu

AAGCTAACATTATCAAGTTGGTTGAAGGGTTTGGCTAACGACCCAGAAAACAAGGTTCCA     180
LysLeuThrLeuSerSerTrpLeuLysGlyLeuAlaAsnAspProGluAsnLysValPro

TTGATCAAGGTTGCTGATGCTAAGCAATTAGGTGAATGGGCTGGTTTGGGTCAAGACGAC     240
LeuIleLysValAlaAspAlaLysGlnLeuGlyGluTrpAlaGlyLeuGlyGlnAspAsp

CGTGAAGGTAACGCCAGAAAGGTTGTCGGTGCCTCCGTTGTTGTTGTCAAGAACTGG      297
ArgGluGlyAsnAlaArgLysValValGlyAlaSerValValValValLysAsnTrp
```

Figure 3.7.a. Restriction map and sequence of cDNA9. Narrow lines represent pBR322 sequences; bold lines indicate the cDNA insert. Open boxes show the GC tails at the ends of the insert. p9S and p9L are the M13 subclones from which sequence information was obtained; arrows show the extent and direction of the sequence. The *Hin*dIII site used in cloning is indicated by overlining. The sequence is shown as continuous, even though no sequence data were obtained across the *Hin*dIII junction. However, when the two sequences were juxtaposed, an open reading frame was observed to run the full length of the combined sequence (Fig. 3.4.b.). Further, when Northern blots were probed with p9S and p9L DNAs, only the latter hybridised, indicating that the sequence obtained from p9L is that of the coding strand.

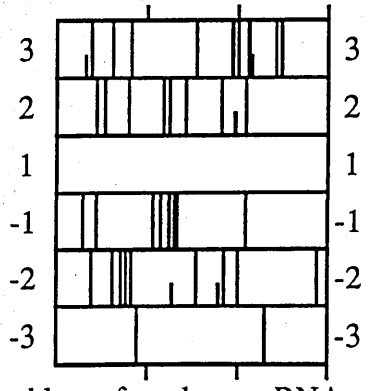


Figure 3.7.b. Open reading frame map vof the sequence obtained from cDNA9. Full bars represent stop codons and short bars methionine codons. Bars above the diagram represent 100 bp intervals.

Figure 3.7.c. Northern blots of total yeast RNA probed with p9S and p9L DNAs.

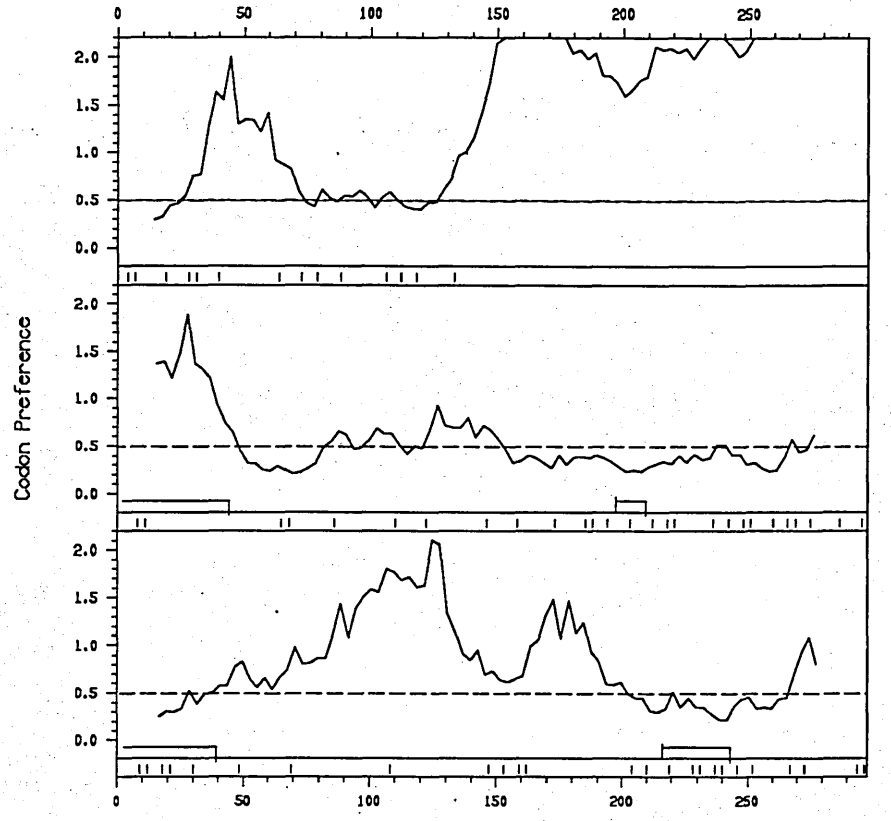


Figure 3.7.d. Codon preference plots for the sequence obtained from cDNA9. Upper, frame 1; middle, frame 2; lower, frame 3. The degree of departure fom the dotted line indicates the extent of deviation from random codon usage. The window length was set at 10 codons. Beneath each plot, open reading frames are indicated as boxed regions. Rare codons, ≤ 5% of a synonymous codon family, are indicated by short bars. Numbers along the top and bottom indicate distance in nucleotides.

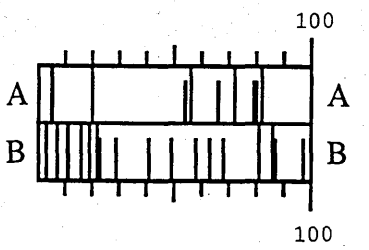
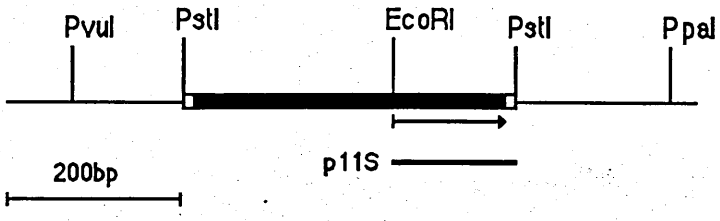


Figure. 3.7.e. Distribution of the acidic (A) and basic (B) amino acids within the putative amino acid sequence presented in Fig. 3.7.a. The acidic residues are glutamic acid (full bar) and aspartic acid (intermediate bar); the basic residues are histidine (small bar), lysine (intermediate bar) and arginine (full bar). The numbers indicate the position of the amino acids stating with leucine (Fig. 3.7.a.)




```
 EcoRI
GAATTCATCAACAAATTGGTTAAGTCCATGGAACGAAACTGGTTGAAGTCACATTACATA      60
—GluPheIleAsnLysLeuValLysSerMetGluArgAsnTrpLeuLysSerHisTyrIle
— AsnSerSerThrAsnTrpLeuSerProTrpAsnGluThrGly *
—  IleHisGlnGlnIleGly *

TAACTACTTATTTTATCTAAGTCCTATATAGTTCTGTTATTCACTAGTTTGTATATTGTA     120
 *

CCTCTA                                                            126
```

Figure 3.8.a. Restriction map and sequence of cDNA11. Narrow lines represent pBR322 sequences; bold lines indicate the cDNA insert. Open boxes show the GC tails at the ends of the insert. p11S is the M13 clone from which sequence information was obtained; arrows show the extent and direction of the sequence. The sequence derived from p11S is probably that of the coding strand since it has a poly(A) tail at its 3'-end (data not shown). Three open reading frames were found (Fig. 3.8.b.); the putative amino acid sequence of each is shown below the DNA sequence.

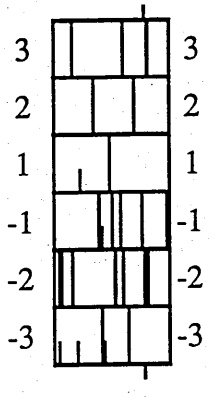


Figure 3.8.b. Open reading frame map of the sequence obtained from cDNA11. Full bars represent stop codons and short bars methionine codons. Bars above the diagram represent 100 bp intervals.

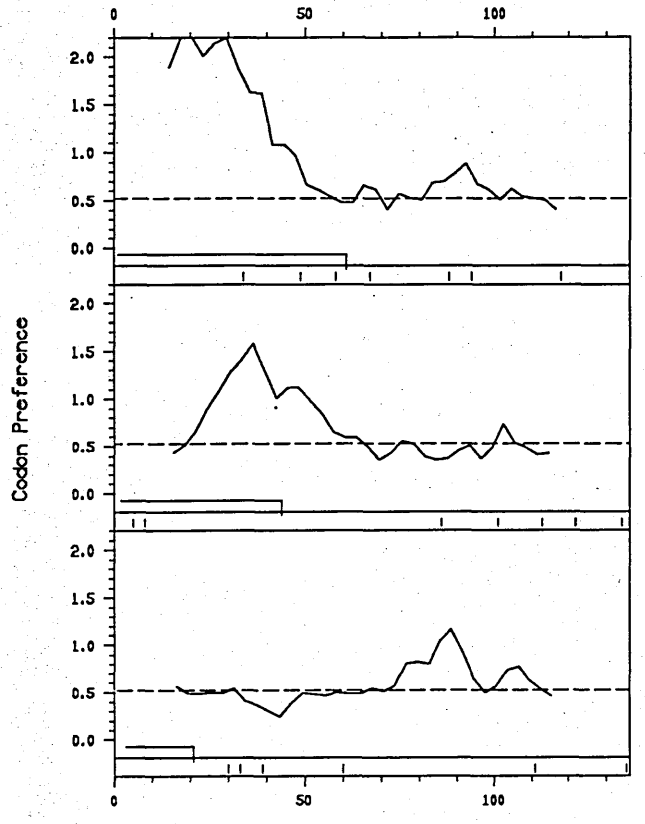


Figure 3.8.c. Codon preference plots for the sequence obtained from cDNA11. Upper, frame 1; middle, frame 2; lower, frame 3. The degree of departure fom the dotted line indicates the extent of deviation from random codon usage. The window length was set at 10 codons. Beneath each plot, open reading frames are indicated as boxed regions. Rare codons, ≤ 5% of a synonymous codon family, are indicated by short bars. Numbers along the top and bottom indicate distance in nucleotides.

A restriction map of cDNA11, together with the sequencing strategy, is shown in Fig.3.8.a. p11S contains the smaller of the two *Eco*RI-*Pst*I fragments in M13mp19. The sequence obtained from p11S is probably that of the coding strand, since a polyA tail is present at its end (data not shown). The DNA sequence is shown below the map, together with putative amino acid sequences in all three frames. The open reading frames are shown diagrammatically in Fig. 3.8.b.

The most significant codon bias within any of these ORFs is in frame 1, although with such a short sequence any such analysis is very dubious (Fig. 3.8.c.). However, codon bias is very high along the full length of the ORF, falling abruptly at the putative termination codon. Analysis of the distribution of acidic and basic residues within frame 1 shows that over these 20 amino acids there are 5 basic and 2 acidic residues (data not shown). Again, however, the sequence is too short to draw meaningful conclusions.

No significant homology with any DNA or amino acid sequence in the GenEMBL and NBRF databases was observed. Thus the identity and function of this gene remains to be established.

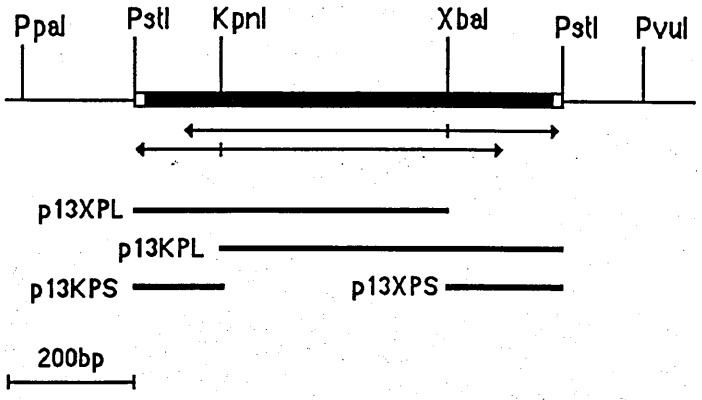
3.2.2.3. cDNA13

The mRNA from which cDNA13 is derived is estimated to have a half-life of 12.1 +/- 1.0 minutes (Santiago, 1986) and a length of approx. 980 nt (this study). The length of the cDNA insert is approx. 660 bp.

A restriction map of the cDNA is shown in Fig. 3.9.a. Subclones p13XPL and p13XPS are *Xba*I-*Pst*I fragments in M13mp19; p13KPL and p13KPS are *Kpn*I-*Pst*I fragments in the same vector. The sequence derived from these subclones is displayed under the restriction map together with the putative amino acid sequence (frame 1; Fig. 3.9.b.).

The sequence was used to search the GenEMBL database with the result shown in Fig. 3.9.e. cDNA13 is highly homologous to the gene encoding ribosomal protein K37 of *S. pombe* (Nischt et al, 1986 and 1987). Within the coding region 78% identity is observed at the DNA level, and 70.7% identity at the amino acid level. When conservative changes are allowed the similarity at the amino acid level is 82.9% (Fig. 3.9.f.).

Ribosomal proteins are in the main very basic proteins, though there are exceptions (Remacha et al, 1988; Mitsui and Tsurugi, 1988 a,b,c.). The deduced amino acid




```
TACAAGTACAGATTACGTGAAGAAATCTTCATTGCTAACGAAGGTGTTCACACTGGTCAA      60
TyrLysTyrArgLeuArgGluGluIlePheIleAlaAsnGluGlyValHisThrGlyGln

TTCATTTACGCTGGTAAGAAGGCTTCTTTGAACGTCGGTAACGTCTTGCCATTGGGTTCT     120
PheIleTyrAlaGlyLysLysAlaSerLeuAsnValGlyAsnValLeuProLeuGlySer

          KpnI
GTCCCAGAAGGTACCATTGTCTCCAACGTTGAAGAAAAGCCAGGTGACAGAGGTGCCCTA     180
ValProGluGlyThrIleValSerAsnValGluGluLysProGlyAspArgGlyAlaLeu

GCCAGAGCTTCTGGTAACTACGTTATCATCATTGGTCACAACCCAGATGAAAACAAAACC     240
AlaArgAlaSerGlyAsnTyrValIleIleIleGlyHisAsnProAspGluAsnLysThr

AGAGTCAGATTACCATCCGGTGCCAAGAAGGTTATCTCTTCTGACGCCAGAGGTGTCATC     300
ArgValArgLeuProSerGlyAlaLysLysValIleSerSerAspAlaArgGlyValIle

GGTGTCATTGCCGGTGGTGGTAGAGTTGACAAACCATTGTTGAAGGCTGGTCGTGCTTTC     360
GlyValIleAlaGlyGlyGlyArgValAspLysProLeuLeuLysAlaGlyArgAlaPhe

CACAAGTACAGATTGAAGAGAAACTCTTGGCCAAAGACCCGTGGTGTTGCCATGAATCCA     420
HisLysTyrArgLeuLysArgAsnSerTrpProLysThrArgGlyValAlaMetAsnPro

                                                    XbaI
GTTGATCACCCTCACGGTGGTGGTAACCATCAACATATTGGTAAGGCTTCTACTATCTCT     480
ValAspHisProHisGlyGlyGlyAsnHisGlnHisIleGlyLysAlaSerThrIleSer

AGAGGTGCCGTTTCTGGTCAAAAGGCTGGTTTGATTGCTGCCAGAAGAACCGGTTTACTA     540
ArgGlyAlaValSerGlyGlnLysAlaGlyLeuIleAlaAlaArgArgThrGlyLeuLeu

CGTGGTTCTCAAAAGACCCAAGATTAATCTTTTTAATTTTGGTTTCTTCCTTCTGTCATA     600
ArgGlySerGlnLysThrGlnAsp *

TTATTTTATCAATTTTCTTAAATATTATATAATTTAATCCGAAACGTTCCTTATAA         656
```

Figure 3.9.a. Restriction map and sequence of cDNA13. Narrow lines represent pBR322 sequences; bold lines represent the cDNA insert. Open boxes denote the GC tails at the ends of the insert. p13XPL, p13KPL, p13KPS and p13XPS are the M13 clones from which sequence was obtained. Arrows below the restriction map indicate the extent and direction of the sequence. The *Kpn*I and *Xba*I sites used in cloning are indicated by overlining. The putative amino acid sequence is shown below the DNA sequence.

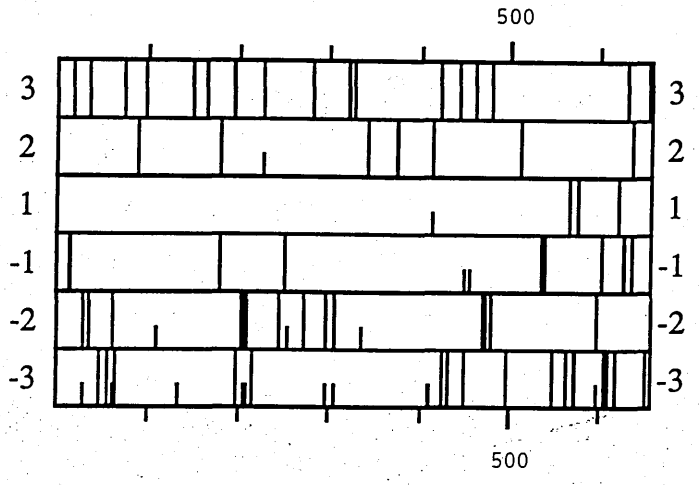


Figure 3.9.b. Open reading frame map of the sequence obtained from cDNA13. Stop codons are represented by full bars and short bars methionine codons. Bars above the diagram represent 100 bp intervals.

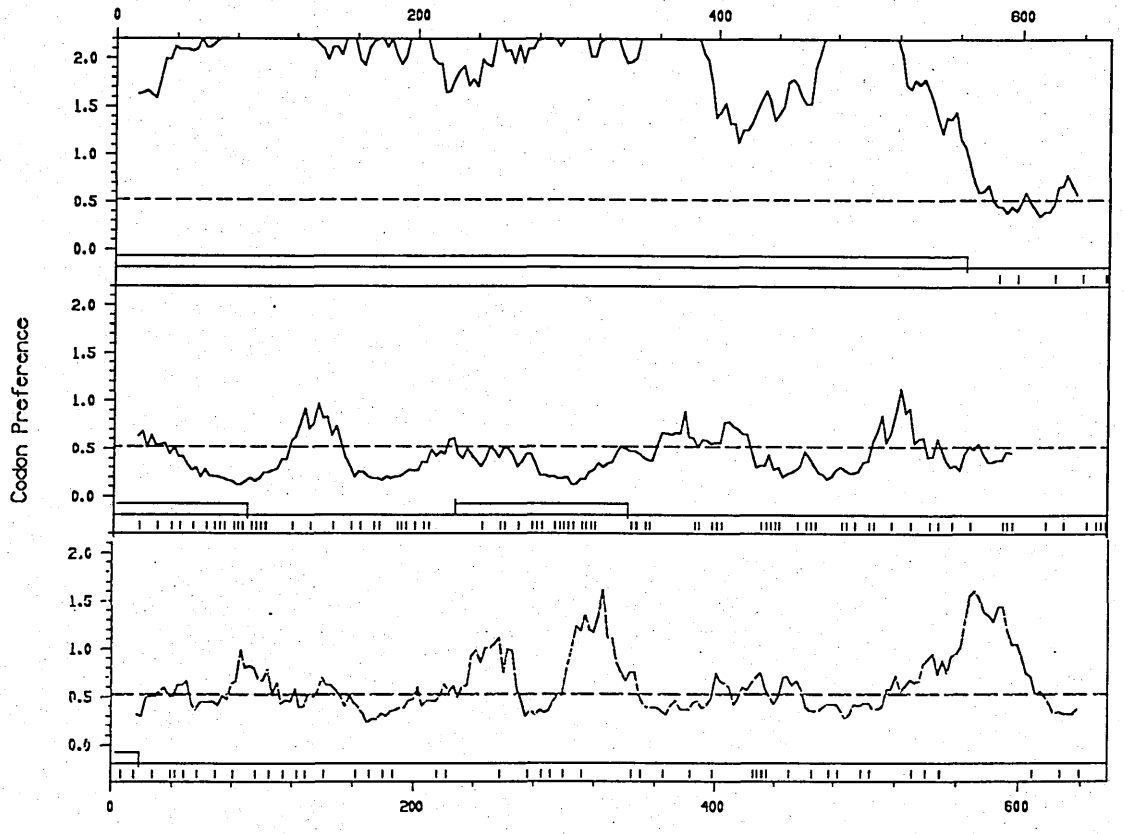


Figure 3.9.c. Codon preference plots for the sequence obtained from cDNA13. Upper, frame 1; middle, frame 2; lower, frame 3. The degree of departure fom the dotted line indicates the extent of deviation from random codon usage. The window length was set at 10 codons. Beneath each plot, open reading frames are indicated as boxed regions. Rare codons, ≤ 5% of a synonymous codon family, are indicated by short bars. Numbers along the top and bottom indicate distance in nucleotides.

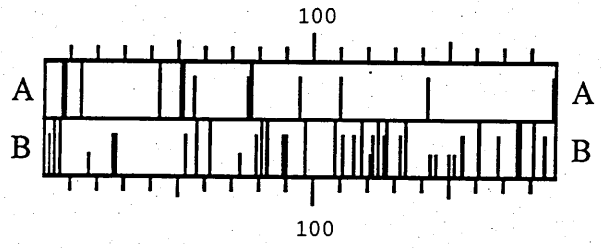


Figure 3.9.d. Distribution of the acidic (A) and the basic (B) residue within the amino acid sequence presented in Fig. 3.9.a. The acidic residues are glutamic acid (full bar) and aspartic acid (intermediate bar); the basic residues are histidine (small bar), lysine (intermediate bar) and arginine (full bar). The numbers indicate the position of the amino acids starting with tyrosine (Fig. 3.9.a.).

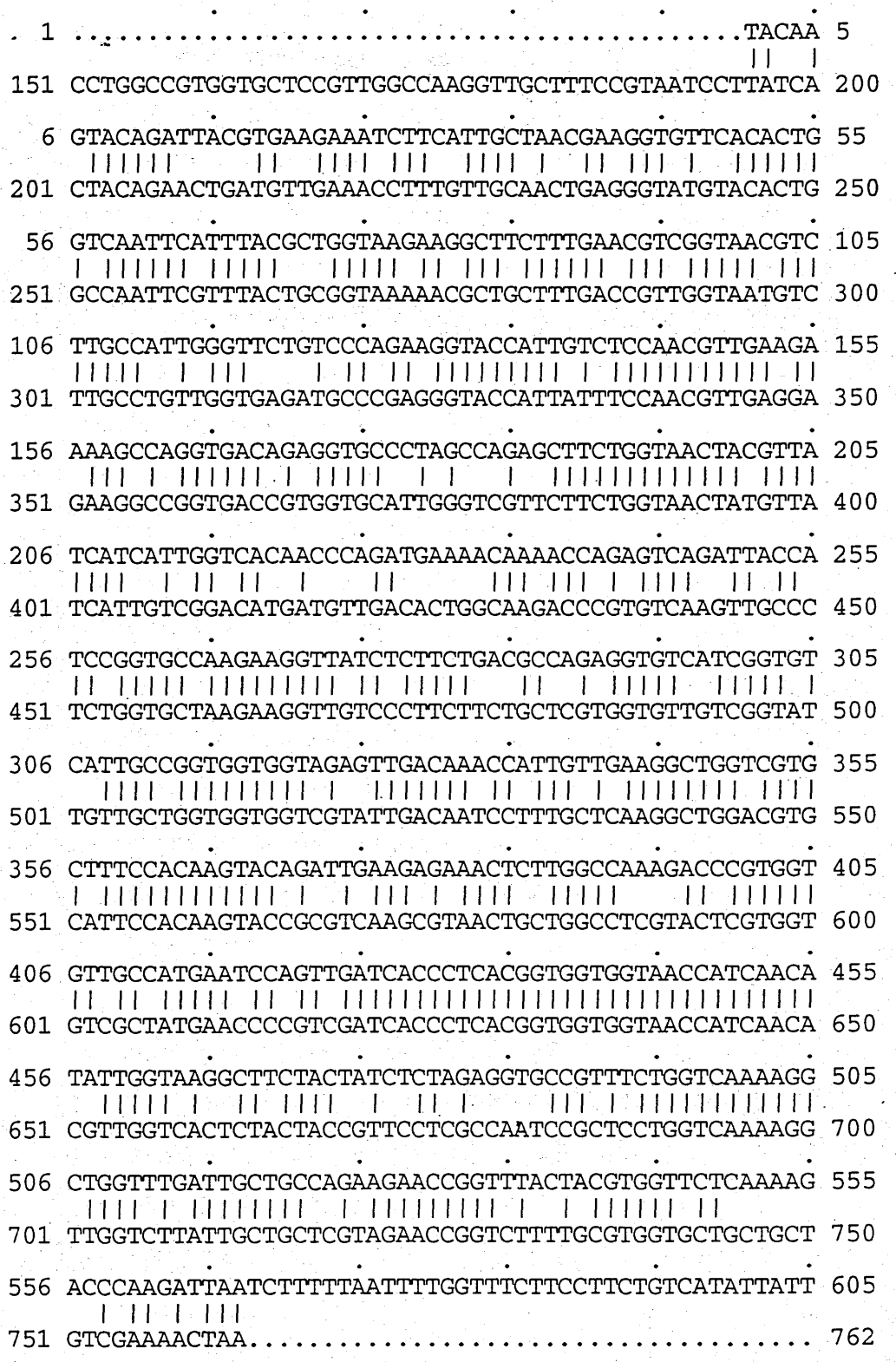


Figure 3.9.e. Identity between cDNA13 (upper) and the gene encoding ribosomal protein K37 (lower) of *S. pombe*. The two genes show considerable homology over the coding region but diverge in the 3'-untranslated region. Over the coding regions they are 78% identical.

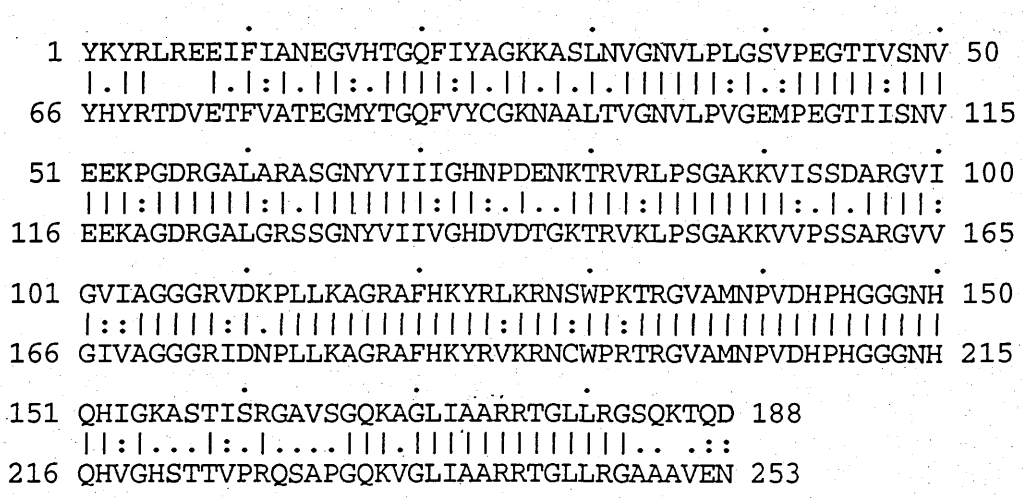


Figure 3.9.f. Similarity between the putative amino acid sequence derived from cDNA13 (upper) and that of ribosomal protein K37 (lower). Full bars join identical amino acids. Two dots join amino acids that are not identical, but for which a pair of possible codons can be found that are related in two out of three positions. One dot joins amino acids that are not identical, but for which a pair of possible codons can be found that are related in one out of three positions. The amino acid sequence derived from cDNA13 and that of K37 are 71% identical, or 83% similar when conservative changes are allowed, as determined by the GCG programme "Gap".

sequence of cDNA13 is indeed that of a basic protein (Fig.3.9.d.). There are 13 acidic residues (6 aspartic acid and 7 glutamic acid residues) and 38 basic residues (15 lysine, 7 histidine and 16 arginine residues) in a total length of 197 amino acids.

All ribosomal protein genes sequenced to date from *S. cerevisiae* show a strong bias towards *S. cerevisiae* codon usage. From the data presented in Fig. 3.9.c. it can be seen that the same holds true for the mRNA represented by cDNA13: codon bias is high over the open reading frame but drops significantly after the UAA stop codon at position 565.

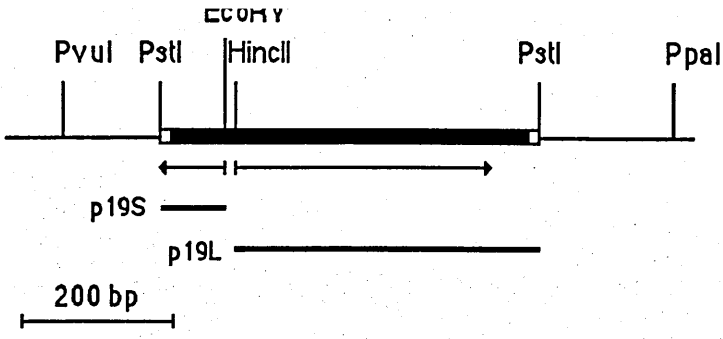
3.2.2.4. cDNA19

The mRNA from which cDNA19 is derived is estimated to have a half-life of 15.3 +/- 2.7 minutes (Santiago,1986) and a length of approx. 680 nt (this study). The length of the cDNA is approx. 500 bp.

A restriction map of cDNA19 is shown in Fig. 3.10.a. The small *Eco*RV-*Pst*I fragment and the large *Hinc*II-*Pst*I fragment were subcloned into M13mp19 cleaved with *Hinc*II and *Pst*I. Unfortunately this does not provide a continuous sequence. However, it has proved very difficult to subclone the large *Eco*RV-*Pst*I fragment since rearrangements at the M13-cDNA boundary, and possibly rearrangements elsewhere in the sequence, were invariably observed to occur. One anomaly in the sequence of p19L is the presence of an *Eco*RI site at position 25 that was not detected when RF p19L DNA was cleaved with this enzyme. This strongly suggests that a rearrangement might have taken place.

The amino acid sequences presented below the DNA sequences of each subclone are those of the major ORFs (Fig.3.10.b.). On this basis the sequence obtained from p19L is presumably that of the coding strand and the sequence obtained from p19S that of the non-coding strand. Further support for the sequence obtained from p19S being that of the non-coding strand is that no evidence of a polyA tail is found at its end. This does not constitute proof, however.

The distribution of acidic and basic residues within the putative amino acid sequence derived from frame 1 of p19L is shown in Fig. 3.10.d. If this is the correct ORF the protein would appear to be quite basic. There are 12 acidic residues (5 glutamic acid and 7 aspartic acid), compared with 30 basic residues (10 lysine and 20 histidine) in a total of 130 amino acid residues. No significant yeast codon bias was observed in any ORF (data not shown).




```
TCTTTGGGTTCCGGGGGGTCTCAGTTGCGGTTAGCATGATTAACACAGAACATACTATAGAC      62
  PheGlyPheArgGlyValSerValAlaValSerMetIleAsnThrGluHisThrIleAsp

                        EcoRV
AAATATGATTATTTACAAGGATATC                                          87
LysTyrAspTyrLeuGlnGlyTyr

 HincII
GTCGACGAGCTGGATGTTGGACTTGAATTCCTCGAGAACCGGCAGCCCAGCAGCGTCGAG      60
ValAspGluLeuAspValGlyLeuGluPheLeuGluAsnArgGlnProSerSerValGlu

AACGGGGCCGGGTTGTCCAGGCCGGGGGTGACGAAAGAGGTGTGCTCGGTCTCGGCGACG      120
AsnGlyAlaGlyLeuSerArgProGlyValThrLysGluValCysSerValSerAlaThr

ATCGGCGCCAGCGGCTCATACTCCTCCAGGCGGGCGCTCACCTCCTTGGCGATGCCCATG      180
.IleGlyAlaSerGlySerTyrSerSerArgArgAlaLeuThrSerLeuAlaMetProMet

AACGTCCGCAGCGTCACACGGAATTGCCCTGGCGCACGGCCTCGGCAAATGGCGGGCGCG      240
AsnValArgSerValThrArgAsnCysProGlyAlaArgProArgGlnMetAlaGlyAla

TTTTGCGGCTTGACGCGCCTTGACGATCAAGGCTTTCCAGGTCGGCGCTTTCCGCGTTGG      300
PheCysGlyLeuThrArgLeuAspAspGlnGlyPheProGlyArgArgPheProArgTrp

CGGCGTGACCAGACTGATTTCATGATCCGCGACTTGGTTTCACCGCGTCGAGGCGCGAGG      360
ArgArgAspGlnThrAspPheMetIleArgAspLeuValSerProArgArgGlyAlaArg

ATTGCTCCTCTGCGGCAAAGTCGCGTGATGA                                    391
IleAlaProLeuArgGlnSerArgValMet
```

Figure 3.10.a. Restriction map and sequence of cDNA19. Narrow lines represent pBR322 sequences; bold lines indicate the cDNA insert. Open boxes show the GC tails at the ends of the insert. p19S and p19L are the M13 subclones from which sequence information was obtained; arrows show the extent and direction of sequence. The sequence derived from p19L, rather than its complement, is thought to be that of the coding strand since it contains a complete open reading frame (Fig.3.10.c). Thus the sequence derived from p19S will be that of the non-coding strand. The complement of the p19S sequence is presented above. There is an open reading frame in the latter sequence (Fig.3.10.b.). The putative amino acid sequences of each of these open reading frames are shown below the DNA sequences.

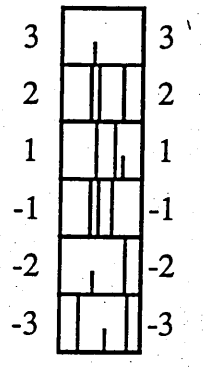


Figure 3.10.b. Open reading frame map of the complement of the p19S sequence. Full bars represent stop codons and short bars methionine codons.

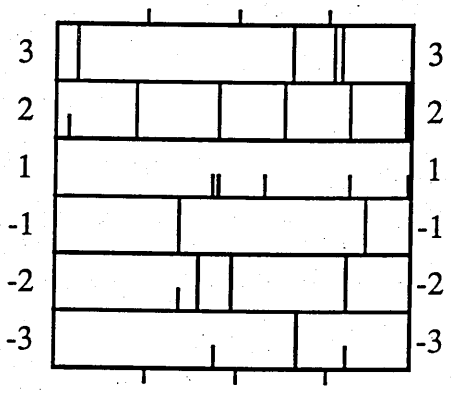


Figure 3.10.c. Open reading frame map of the sequence obtained from p19L Full bars represent stop codons and short bars methionine codons. Bars above the diagram represent 100 bp intervals.

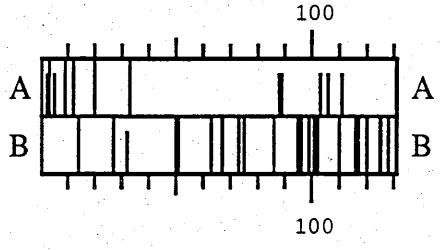


Figure 3.10.d. Distribution of the acidic (A) and basic (B) residues within the amino acid sequence obtained from p19L. The acidic residues are glutamic acid (full bar) and aspartic acid (intermediate bar); the basic residues are lysine (intermediate bar) and arginine (full bar). The numbers above the diagram indicate the position of the amino acids starting with valine.

No significant homology with any DNA or amino acid sequence in the GenEMBL and NBRF databases was observed. Thus the function of this gene remains to be identified.

3.2.2.5. cDNA39

The mRNA from which cDNA39 is derived is estimated to have a half-life of 18.3 +/- 1.5 minutes (Santiago, 1986) and a length of approx. 570 nt (this study). The length of the cDNA is approx. 550 bp, and thus is almost full length.

A restriction map of cDNA39 is shown in Fig. 3.11.a. together with the sequencing strategy. Unfortunately, since subclones representative of the entire cDNA could not be obtained, sequence information was derived only from the indicated *Bgl*II-*Pst*I and *Eco*RI-*Pst*I fragments, cloned into M13mp19.

When the sequences of p39S and p39L were used to search the GenEMBL database, both showed a high degree of homology with an *S. cerevisiae* gene, *CYH2*, that encodes the ribosomal protein L29 (Kaufer et al, 1983).

The sequence of *CYH2* is shown in Fig. 3.11.a. Start and stop codons are indicated as is the position of the single intron. Sequences equivalent to those obtained from cDNA39 are boxed. The one difference between cDNA39 and *CYH2* is indicated by an arrow at position 782; this G residue was not found in the sequence of cDNA39. Here, however, it must be emphasised that cDNA39 was sequenced on only one strand, while *CYH2* sequence was sequenced on both strands. In any case, the difference is in the 3'-untranslated region rather than the coding region and may represent a strain difference. Finally, the residue at the 3'-end of the cDNA39 sequence (which precedes a poly(A) tract; data not shown), is one of three transcription termination sites used in *CYH2* (Kaufer et al, 1983).

The codon bias of *CYH2* is shown graphically in Fig. 3.11.b. Like most other ribosomal protein genes, codon usage is that of a highly expressed gene. The distribution of acidic and basic residues within the coding region is shown in Fig. 3.11.c. Like most other ribosomal proteins L29 is very basic, containing 10 acidic residues (5 each of aspartic and glutamic acid) and 40 basic residues (18 lysine, 10 histidine and 12 arginine residues) in a total of 249.

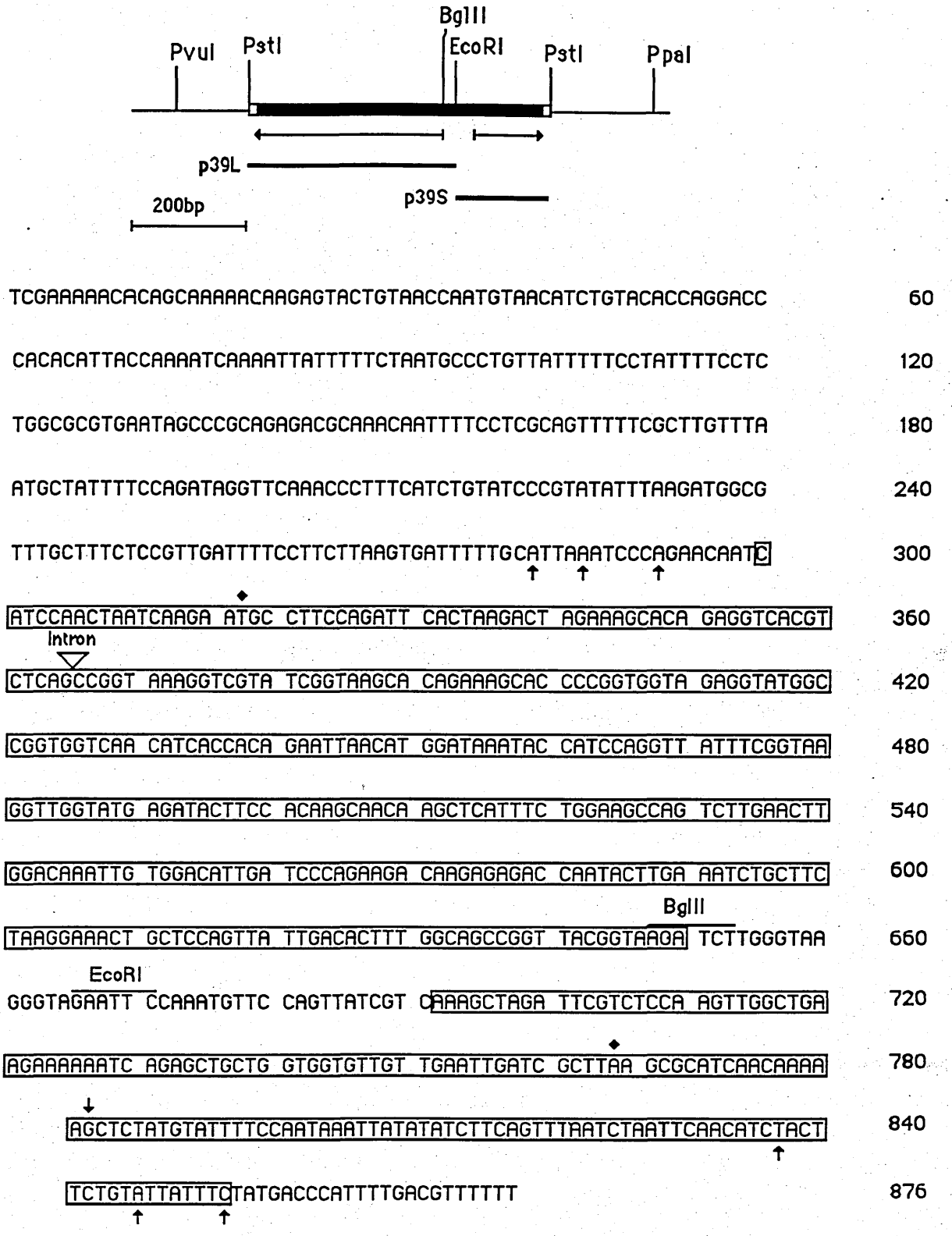


Figure 3.11.a Restriction map of cDNA39 and sequence of *CYH2* (taken from Kaufer et al, 1983). Narrow lines represent pBR322 sequences; bold lines represent the cDNA insert. Open boxes show the GC tails at the ends of the insert. p39L and p39S are the M13 clones from which sequence information was obtained. Arrows below the restriction map indicate the extent and direction of the sequence. Stop and start codons are highlighted by triangles above the sequence. The position of the intron is marked and the *Bgl*II and *Eco*RI sites used in subcloning are overlined. The boxed region represents the sequenced available from cDNA39. The one difference is highlighted by an arrow above the sequence at position 782. This G is absent in cDNA39. Arrows below the 5' and 3' flanking regions mark the transcription initiation and termination sites, determined by Kaufer et al (1983).

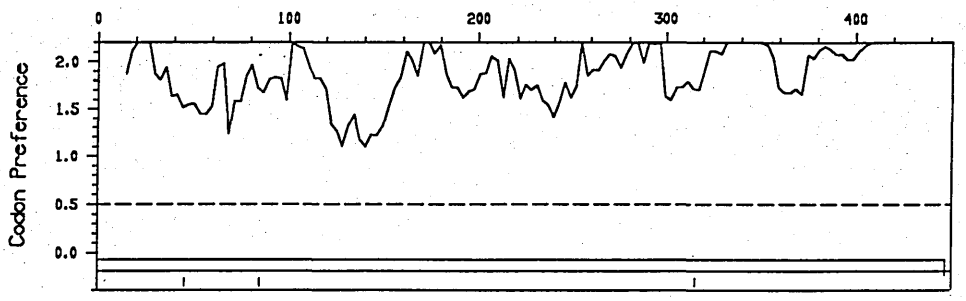


Figure 3.11.b. Codon preference plot of the *CYH2* sequence. The degree of departure from the dotted line indicates the extent of deviation from random codon usage. The window length was set at 10 codons. Beneath the plot, the open reading frame is indicated a boxed region. Rare codons, ≤ 5% of a synonymous codon family, are indicated by short bars. Numbers along the top and bottom indicate distance in nucleotides.

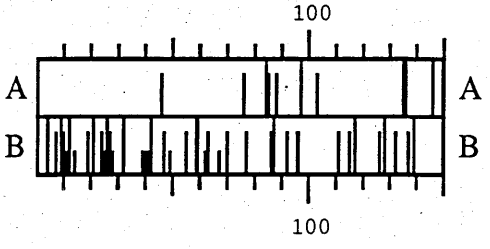


Figure 3.11.c. Distribution of the acidic (A) and basic (B) residues within the amino acid sequence of L29. The acidic residues are glutamic acid (full bar) and aspartic acid (intermediate bar); the basic residues are histidine (small bar), lysine (intermediate bar) and arginine (full bar). The numbers indicate the position of the amino acids starting from methionine.

3.2.2.6. cDNA90

The mRNA from which cDNA90 is derived is estimated to have a half-life of 6.6 +/- 0.67 minutes (Santiago, 1986) and a length of approx. 1300 nt (this study). The length of the cDNA insert is approx. 1.2-kb, and thus it is almost full length.

A restriction map of cDNA90 is shown Fig. 3.12.a. Sequence information was obtained from the indicated *Sal*I-*Pst*I fragment, which was subcloned into M13mp19 (p90). The sequence of p90 was used to search the GenEMBL database, and was found to be 99.6% identical with that of the *S. cerevisiae* gene *tcm1*, a mutant allele of the gene that encodes the ribosomal protein L3 (Shultz and Friesen, 1983). At the amino acid level, only one difference was observed between *tcm1* and cDNA90 over the area sequenced.

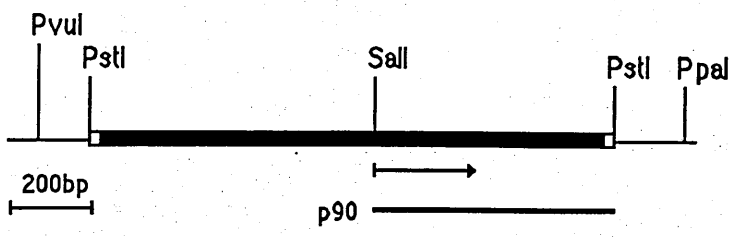
The sequence of the coding region of *tcm1*, together with the deduced amino acid sequence, is shown in Fig. 3.12.a. The region equivalent to that derived from p90 is boxed. The indicated nucleotide in *tcm1* at position 765 is a G residue in cDNA90. This would lead to a tryptophan codon at this position in cDNA90, instead of the cysteine codon in *tcm1*. The difference may have been the result of misincorporation during first-strand cDNA synthesis, or could be a sequencing artifact. Alternatively, since the *tcm1* sequence is that of a mutant gene, the difference may be genuine.

The codon bias of *tcm1* is shown graphically in Fig. 3.12.b. and the distribution of acidic and basic residues of L3 is shown in Fig.3.12.c.: L3 is a very basic protein, like most other ribosomal proteins from eukaryotes and prokaryotes alike. There are 36 acidic residues (16 aspartic and 20 glutamic acid residues) and 93 basic residues (17 histidine, 43 lysine and 33 arginine residues) in a total of 387.

3.2.2.7. cDNA100

The mRNA from which cDNA100 is derived is estimated to have a half-life of 10.4 +/- 1.1 minutes (Santiago, 1986) and a length of approx. 760 nt (this study). The length of the cDNA insert is approx. 600 bp.

A restriction map of cDNA100 is shown in Fig. 3.13.a. The small *Eco*RI-*Pst*I fragment was subcloned into M13mp19 (p100S). Since the sequence obtained from p100 shows no evidence of a poly(A) tail, even though sequencing was carried out up to the cDNA/vector boundary, the complement is shown below the restriction map. That the latter sequence contains coding information is in accord with ORF and codon bias analysis (Fig. 3.13.b. and Fig. 3.13.c.).




```
ATGTCTCACA GAAAGTACGA AGCACCACGT CACGGTCATT TAGGTTTCTT GCCAAGAAAG      60
AGAGCTGCCT CCATCAGAGC TAGAGTTAAG GCTTTTCCAA AGGATGACAG ATCCAAGCCA     120
GTTGCTCTAA CTTCCTTCTT GGGTTACAAG GCTGGTATGA CCACCATTGT CAGAGATTTG     180
GACAGACCAG GTTCTAAGTT CCACAAGCGT GAAGTTGTCG AAGCTGTCAC CGTTGTTGAC     240
ACTCCACCAG TTGTCGTTGT TGGTGTTGTC GGTTACGTCG AAACCCCAAG AGGTTTGAGA     300
TCTTTGACCA CCGTCTGGGC TGAACATTTG TCTGACGAAG TCAAGAGAAG ATTCTACAAG     360
AACTGGTACA AGTCTAAGAA GAAGGCTTTC ACCAAATACT CTGCCAAGTA CGCTCAAGAT     420
GGTGCTGGTA TTGAAAGAGA ATTGGCTAGA ATCAAGAAGT ACGCTTCCGT CGTCAGAGTT     480
TTGGTCCACA CTCAAATCAG AAAGACTCCA TTGGCTCAAA AGAAGGCTCA TTTGGCTGAA     540
ATCCAATTGA ACGGTGGTTC CATCTCTGAA AAGGTTGACT GGGCTCGTGA ACATTTCGAA     600
                SalI
AAGACTGTTG CTGTCGACAG CGTTTTTGAA CAAAACGAAA TGATTGACGC TATTGCTGTC     660
ACCAAGGGTC ACGGTTTCGA AGGTGTTACC CACAGATGGG GTACTAAGAA ATTGCCAAGA     720
                                    ◆
AAGACTCACA GAGGTCTAAG AAAGGTTGCT TGTATTGGTG CTTGCCATCC AGCCCACGTT     780
ATGTGGAGTG TTGCCAGAGC TGGTCAAAGA GGTTACCATT CCAGAACCTC CATTAACCAC     840
AAGATTTACA GAGTCGGTAA GGGTGATGAT GAAGCTAACG GTGCTACCAG CTTCGACAGA     900
ACCAAGAAGA CTATTACCCC AATGGGTGGT TTCGTCCACT ACGGTGAAAT TAAGAACGAC     960
TTCATCATGG TTAAAGGTTG TATCCCAGGT AACAGAAAGA GAATTGTTAC TTTGAGAAAG    1020
TCTTTGTACA CCAACACTTC TAGAAAGGCT TTGGAAGAAG TCAGCTTGAA GTGGATTGAC    1080
ACTGCTTCTA AGTTCGGTAA GGGTAGATTC CAAACCCCAG CTGAAAAGCA TGCTTTCATG    1140
GGTACTTTGA AGAAGGACTT GTAA                                           1164
```

Figure 3.12.a. Restriction map of cDNA90 and coding sequence of the *tcm1* gene (taken from Shultz and Friesen, 1983). Narrow lines represent pBR322 sequences; bold lines represent the cDNA insert. Open boxes show the GC tails at the ends of the insert. p90 is the M13 clone from which sequence information was obtained. Arrows below the restriction map indicate the extent and direction of the sequence. The *Sal*I site used in subcloning is overlined. The boxed region represents the sequenced available from cDNA90. The one difference observed between cDNA90 and *tcm1* is highlighted by a triangle at position 765; in cDNA90 this residue is a G. This results in a cysteine in L3 rather than the tryptophan in the sequence deduced from cDNA90 at amino acid position 255.

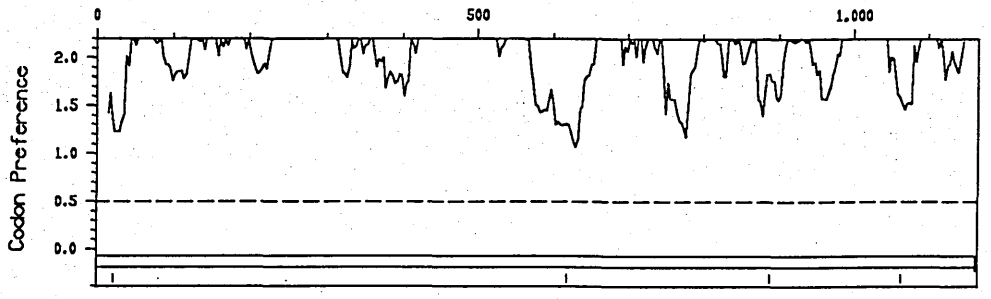


Figure 3.12.b. Codon preference plot of the *tcm1* sequence. The degree of departure from the dotted line indicates the extent of deviation from random codon usage. The window length was set at 10 codons. Beneath the plot, the open reading frame is indicated a boxed region. Rare codons, ≤ 5% of a synonymous codon family, are indicated by short bars. Numbers along the top and bottom indicate distance in nucleotides.

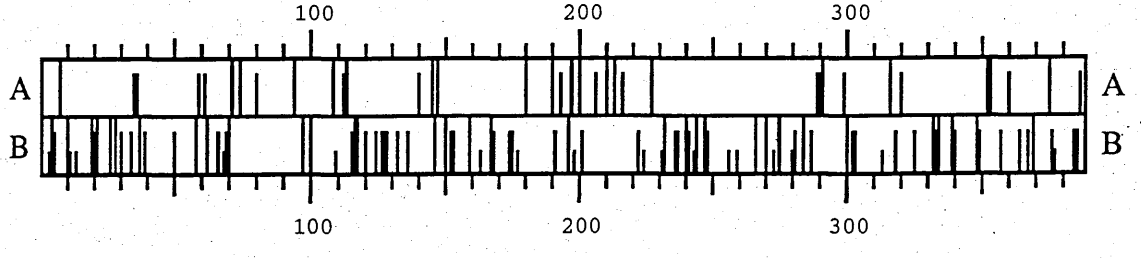
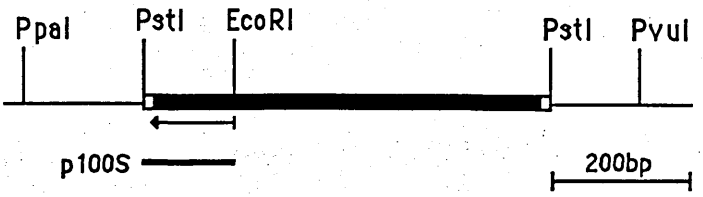


Figure 3.12.c. Distribution of the acidic (A) and basic (B) residues within the amino acid sequence of ribosomal protein L3. The acidic residues are glutamic acid (full bar) and aspartic acid (intermediate bar); basic residues are histidine (small bar), lysine (intermediate bar) and arginine (full bar). The numbers indicate the position of the amino acids starting with methionine.




```
AATACAACTAAGACTAAGCAACAATGCCAAAGAGCTCCAAGACTTACTCTAAGAACTTAC      60
AsnThrThrLysThrLysGlnGlnCysGlnArgAlaProArgLeuThrLeuArgThrTyr

TCTACACCAAAGAGACCTTACGAATCTTCTCGTTTGGACGCAGAATTGAAGTTGGCNNGT     120
SerThrProLysArgProTyrGluSerSerArgLeuAspAlaGluLeuLysLeuAla ?

 EcoRI
GAATTC                                                            126
GluPhe
```

Figure 3.13.a. Restriction map and sequence of cDNA100. Narrow lines represent pBR322 sequences; bold lines represent the cDNA insert. Open boxes denotes the GC tails at the ends of the inserts. p100S is the M13 subclone from which sequence information was obtained. Arrows indicate the extent and direction of sequence. The *Eco*RI site used in subcloning is shown by overlining. Two nucleotides adjacent to the *Eco*RI site were not determined and are denoted by an N. One significant open reading frame was found (Fig. 3.13.b.), and the putative amino acid sequence is shown below the DNA sequence.

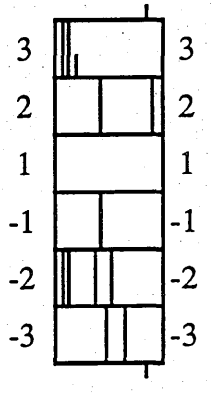


Figure 3.13.b. Open reading frame map of the sequence presented in Fig.3.13.a.. Full bars represent stop codons and short bars methionine codons.

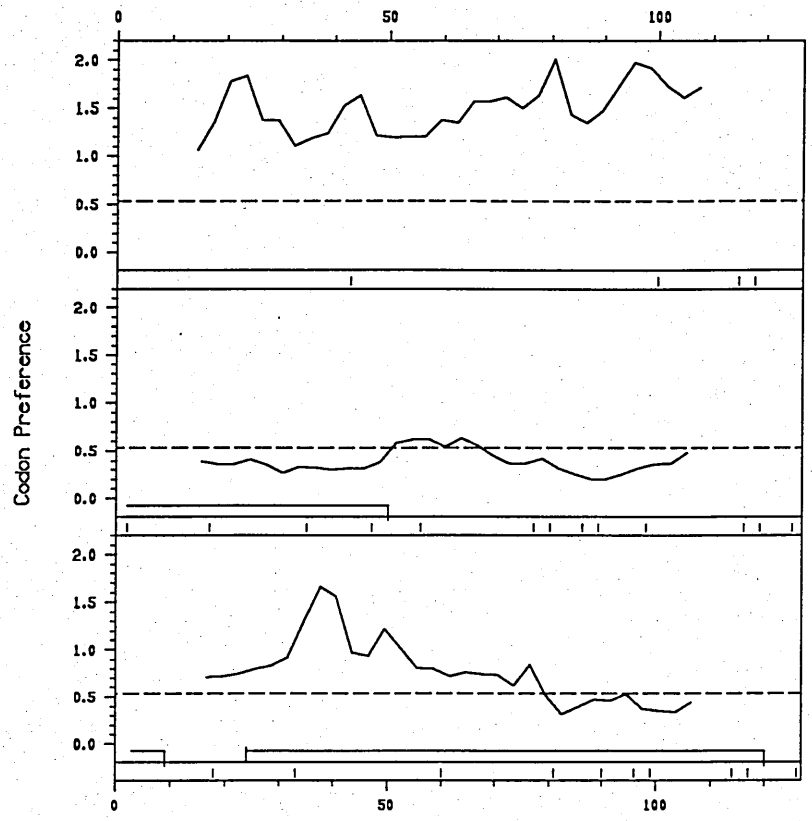


Figure 3.13.c. Codon preference plot of the sequence obtained from cDNA100. Upper, frame 1; middle, frame 2; lower, frame 3. The degree of departure from the dotted line indicates the extent of deviation from random codon usage. The window length was set at 10 codons. Beneath each plot, open reading frames are indicated as boxed regions. Rare codons, ≤ 5% of a synonymous codon family, are indicated by short bars. Numbers along the top and bottom indicate distance in nucleotides.

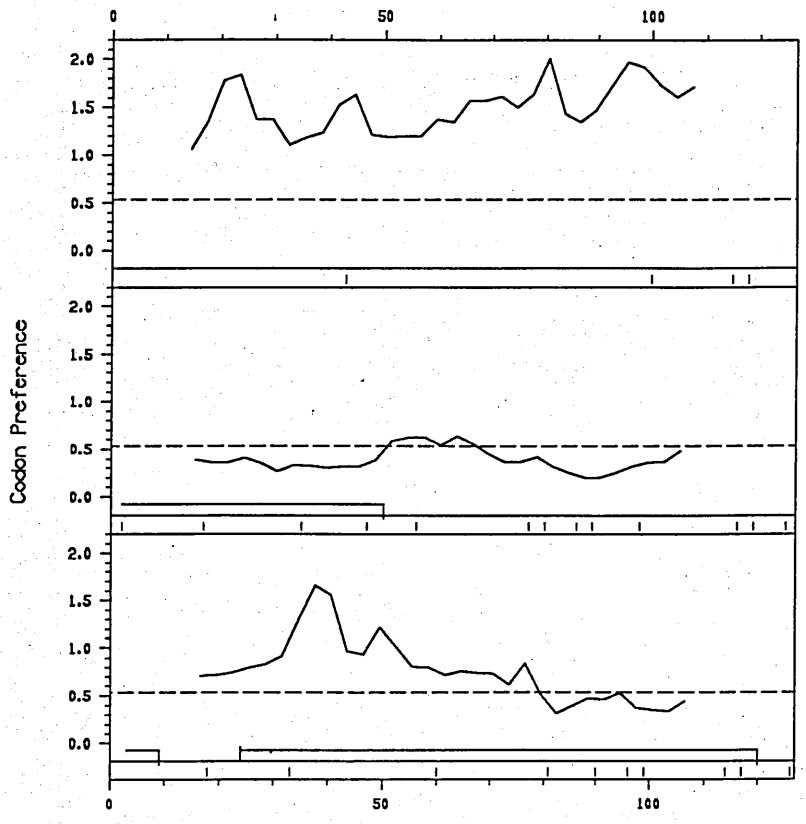


Figure 3.13.c. Codon preference plot of the sequence obtained from cDNA100. Upper, frame 1; middle, frame 2; lower, frame 3. The degree of departure from the dotted line indicates the extent of deviation from random codon usage. The window length was set at 10 codons. Beneath each plot, open reading frames are indicated as boxed regions. Rare codons, ≤ 5% of a synonymous codon family, are indicated by short bars. Numbers along the top and bottom indicate distance in nucleotides.

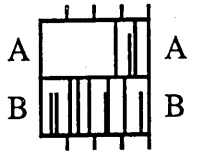


Figure 3.13.d. Distribution of the acidic (A) and basic (B) residues within the putative amino acid sequence presented in Fig. 3.13.a. The acidic residues are glutamic acid (full bar) and aspartic acid (intermediate bar); the basic residues are histidine (small bar), lysine (intermediate bar) and arginine (full bar). The numbers indicate the position of the amino acids starting with isoleucine.

The deduced amino acid sequence of this frame 1 is shown below the DNA sequence in Fig. 3.13.a. As well as being the largest ORF, frame 1 has the highest yeast codon bias. Analysis of the distribution of acidic and basic residues encoded by frame 1, shows that there are 3 acidic residues (2 glutamic acid; 1 aspartic acid) and 9 basic residues (5 arginine; 4 lysine), in a total of 42 residues. However, the length of the sequence available for analysis is too short for meaningful conclusions to be drawn.

No significant homology with any DNA or amino acid sequence in the GenEMBL and NBRF databases was observed. Thus the identity and function of this gene remains to be established.

3.3. Discussion

Of the 13 cDNAs studied by Santiago (1986), twelve have been sequenced, one completely, the remainder incompletely. Three have been identified as known *S. cerevisiae* genes, one as a putative counterpart of a known *S. pombe* gene, and one as being related to a putative mouse housekeeping gene. To the remaining seven I have been unable to assign a function.

3.3.1. cDNAs Representing Long Half-Life mRNAs

Very little can be concluded with respect to the mRNAs for which no similarity to known genes has been observed. The codon bias data are perhaps the most enlightening, as these can give some indication as to the level of expression of a gene (Sharp et al, 1986). From the analyses of cDNAs 74 and 85, it would appear that the frames most likely to encode proteins are frames 3 and 1 respectively (Fig. 3.5.a.; Fig. 3.6.a.). Although neither frame has a codon bias as high as those of most ribosomal proteins (examples of highly expressed genes), they are suggestive of moderately expressed genes. Unfortunately, no significant codon bias was observed in any frame of the cDNA22 sequence. Whether this is because the gene is expressed at low levels remains to be seen.

3.3.1.2. cDNA10

cDNA10 encodes the glucose-inducible glycolytic enzyme, enolase (Holland et al, 1981). Enolase (E C. 4.2.1.11) catalyses the conversion of 2-phosphoglycerate to phosphoenolpyruvate. There are two enolase genes per haploid genome in *S. cerevisiae* (Holland et al, 1981), each at a different genomic location. Santiago (1986) prepared the RNA that he used in the construction of his cDNA library from cells growing exponentially in rich medium containing 2% glucose. It has been estimated that the *ENO2* product is 20-fold more abundant than the glucose-constitutive *ENO1* product, under such conditions. Only when cells are grown on ethanol, or glycerol plus lactate, are there similar amounts of each protein. *In vitro* translation of total cellular mRNA suggests that the relative levels of *ENO1* and *ENO2* mRNAs reflect those of the respective proteins (McAlistair and Holland, 1982). Thus it is not surprising that cDNA10 represents *ENO2* rather than *ENO1*. From his analysis of genomic Southern blots, Santiago had postulated that cDNA10 might be a member of a gene family. The fact that there are two enolase genes, which are 95% identical over their coding regions

and 70% identical in their 5'- and 3'-untranslated regions, confirms Santiago's hypothesis. *ENO1* and *ENO2* mRNAs cannot be distinguished using a cDNA10 probe. The half-life measured by Santiago must have been effectively that of the latter. Thus, the possibility that the two mRNAs are turned over at different rates remains to be determined. Whatever the case, however, regulation of *ENO1* and *ENO2* appears to occur predominantly at the level of transcription (McAlister and Holland, 1982).

Differential transcription rates of the enolase genes in cells grown on glucose or gluconeogenic carbon sources appears to be mediated through different upstream elements. In the case of *ENO2*, two UAS elements, located immediately upstream and downstream of position -461 relative to the transcription start site, are required both for expression and glucose induction (Cohen et al, 1986). As for other yeast UAS elements they can function in both orientations relative to the coding region. In the case of *ENO1*, the constitutive enolase gene, a URS (upstream repressor sequence) lies within a 38 bp sequence near position -160. Deletion of this element permits glucose-dependent induction of *ENO1* in a manner quantitatively similar to that of *ENO2* (Cohen et al, 1987). *ENO1* also has sequence elements near to position -445 that are functionally similar to the UAS elements of *ENO2*, and that presumably mediate glucose induction in the absence of the URS (Cohen et al, 1987). Further, DNA binding assays and footprinting demonstrate that the two UAS elements of *ENO2* are indeed the sites of protein/DNA interactions. The two proteins, RAP1 (repressor/activator protein) and ABF1 (autonomously replicating sequences [ARS]-binding factor), bind to overlapping sequences around position -461 of *ENO2*. RAP1 also binds to one of the UAS elements in the *ENO1* gene, while another factor, designated EBF1 (enolase binding factor) binds to sequences just upstream (Brindle et al, 1990). However, the URS of *ENO1* has not been investigated further. Interestingly RAP1 is the same protein as TUF, the *trans*-acting factor that binds to UAS_{rpg} elements in ribosomal protein genes (Section 1.3.4.). Further experiments are required to assess whether the binding of RAP1 and/or ABF1 to *ENO2* mediates glucose induction or whether they are required for transcription *per se*. However, RAP1 dependent transcriptional activation is required for efficient expression, but not glucose regulation, of *ADH1*, another glycolytic gene (Santangel and Tornow, 1990). In cells grown in gluconeogenic conditions a trans-acting factor, the product of the GCR1 gene acts as a positive regulator of both *ENO1* and *ENO2*, as well as of other glycolytic genes (Holland et al, 1987; Santangelo and Tornow, 1990). It is not known whether the product of GCR1, which is an inessential gene, acts directly or indirectly.

Half-lifes of two other mRNAs encoding glycolytic enzymes have been measured: 59.9 +/- 7.8 minutes for the 1600 nt mRNA encoding pyruvate kinase (Santiago, 1986), and variously as 30 minutes; 45 minutes and 70-80 minutes for the 1600nt

mRNA encoding phosphoglycerate kinase (Mellor et al, 1987; Herrick et al, 1990; Chen et al, 1984). The reason for the discrepancies is not clear. Thus, in addition to enolase mRNA, mRNAs encoding at least one, and perhaps two, other glycolytic enzymes fall within the long half-life population.

3.3.1.3. cDNA46

Only one other cDNA that encodes a stable mRNA, of the five studied, has been found to be related to a known gene. Homology between cDNA46 and the mouse gene *MER5* is shown in Fig.3.4.d. *MER5* was cloned by Yamamoto et al (1989) from a cDNA library prepared from erythroblasts derived from the spleen of an anaemic mouse. The authors were interested in the process of erythroid cell differentiation, for which murine erythroleukemia (MEL) cells constitute a model system (Marks and Rifkind, 1978). When MEL cells are treated with a variety of agents, expression of several erythroid-specific genes is induced after a latent period, including several "housekeeping-type" genes (Marks and Rifkind, 1978). In order to prepare an erythroid-specific probe a cDNA library was prepared from induced MEL cells, and mRNA from Ehrlich ascites tumour cells was used as a driver for subtraction hybridisation. The subtracted probe was used to screen the spleen library. The longest cDNA clone they isolated, pMER5, was sequenced and a single open reading frame of 257 codons was found. *MER5* mRNA is notably more abundant in three MEL lines as compared with non-erythroid mouse lines. Thus *MER5* is expressed preferentially in erythroid cells, and on the basis of preliminary evidence, most abundantly in immature erythroblasts.

No TATA-like element has been found upstream of the *MER5* transcription start site, although there are several typical targets for transcription factors within the 150nt proximal to the cap site (Yamamoto et al, 1989). In addition there is a motif reminiscent of one essential for transcription of the human β-globin gene, and that is common in adult β-globin promoters of many species but uncommon in the promoters of other eukaryotic genes (Myers et al, 1986). The function of *MER5* remains unknown, although Yamamoto et al (1989) hypothesise that by virtue of lacking a TATA box it may be a "housekeeping" gene. *ENO2*, *PYK* and *PGK* may also be considered to be "housekeeping" genes, as may *ACT1* (half-life 76.6 +/- 15 minutes; Santiago et al, 1986: half-life approx. 30 minutes; Herrick et al, 1990). Perhaps long, or longish, mRNA half-life is a general characteristic of housekeeping genes. As the similarity observed between the putative amino acid sequence derived from cDNA46 and that derived from *MER5* does not run the full length of the former sequence, it is possible that the similarity is between functional domains of two distinct proteins, and thus that the two proteins are not true counterparts.

3.3.2. cDNAs Representing Short Half-Life mRNAs

All of the short half-life mRNAs for which a function has been identified encode ribosomal proteins.

There are 70 to 80 ribosomal proteins in *S. cerevisiae*. Half-lifes of a few of the corresponding mRNAs have been measured by Kim and Warner (1983b) on the basis of incorporation kinetics. They found all of them to be relatively short. For example, the half-life of the mRNA encoding ribosomal protein L3 (equivalent to mRNA90) was found to be 13.1 minutes, rather more than the 6.6 +/- 0.67 minutes calculated by Santiago (1986), but still low. The discrepancy is most probably due to the different methods, though strain differences and/or growth conditions may have played a part. Kim and Warner measured the abundance of ribosomal protein mRNAs in minimal medium with glucose as a carbon source and at 23°C; in the experiments of Santiago (1986) glucose was also present but the growth medium was rich, and the cells were grown at 30°C. Either or both of these differences may account for the discrepancy. In any case, the fact that mRNA half-lifes vary by more than an order of magnitude in *S. cerevisiae* is no doubt more significant than quibbles about the absolute half-life of any particular transcript.

3.3.2.1. cDNA39

cDNA39 is equivalent to the *CYH2* gene, which in a mutated form confers resistance to cycloheximide, an inhibitor of peptide elongation (McLaughlin, 1974). *CYH2* encodes the ribosomal protein L29 (Fried and Warner, 1982). Both mutant (*cyh2*, resistant to cycloheximide) and wild-type (*CYH2*) forms of the gene have been sequenced by Kaufer et al (1983). They are identical except for a single nucleotide substitution, C to G, at position 428 (Fig. 3.11.a.), which results in replacement of glutamine (*CYH2*) by glutamic acid (*cyh2*) at residue 37 of the protein. Moreover, Stocklein et al (1981) have characterised a second mutant form of the L29 protein, which has a lysine instead of a glutamine at the same position. This presumably arose from a C to A mutation at position 428. cDNA39 is presumably the wild type allele of L29 as it has a C at position 428. Although the complete sequence of cDNA39 was not determined (Fig. 3.11.a.), information was obtained from both ends of the cDNA. The last C residue prior to the poly(A) tail is the most 5' of the three transcription termination sites described by Kaufer et al (1983). These authors also observed three possible transcriptional initiation sites. cDNA39 begins 8 to 19bp short of these (Fig. 3.11.a.).

The *CYH2* gene contains an intron, as do many ribosomal protein genes (see Woolford, 1989). The splice junction occurs within codon 17, near to the 5' end of the gene. Sequences at the 5' and 3' boundaries of the intron are closely related to the consensus donor and acceptor sequences compiled by Mount (1982). Further, the *CYH2* intron contains a consensus branchpoint sequence (Langford et al, 1984) approx. 45 nt upstream of the 3' splice site. Further aspects of splicing in *S. cerevisiae* will be discussed in more detail in Chapters 4 and 5.

CYH2 is a single copy gene (Fried and Warner, 1982) that maps to chromosome VII (Mortimer and Hawthorne , 1966). Fried et al (1985) have demonstrated that *CYH2* is an essential gene: haploid cells in which the only source of L29 is a copy of *CYH2* under the control of the *GAL10* promoter grow normally on galactose but are non-viable on glucose. This was confirmed by Miles et al (1988).

Since mutations in *CYH2* confer resistance to cycloheximide, and cycloheximide interferes with peptide elongation, L29 is presumed to be part of the peptidyltransferase centre.

A putative *Neurospora crassa* counterpart of *CYH2* has been cloned, *cyp-1* (**cy**toplasmic ribosomal **p**rotein gene), and is most similar to the yeast protein over its amino terminal region (Kreader and Heckman, 1987). Belhumeur et al (1987) recently cloned the gene encoding mouse ribosomal protein L27'. The deduced amino acid sequence of L27' is 74% similar, when conservative changes are allowed, with that of the yeast L29 protein. Further, over the first 58 amino acids the two proteins are 86% identical, indicating that this region may be of particular functional importance. Homologies between mammalian and yeast ribosomal proteins have been reported before but they are generally weaker than observed between L29 of *S. cerevisiae* and L27' of mouse (Lin et al, 1983; Itoh et al, 1985). Do these two proteins perform the same function in the cell? Yeast strains in which *cyh2* can be conditionally expressed have been created by substituting the wild-type *CYH2* locus for *cyh2* driven by a *GAL10* promoter. They can grow on galactose but not on glucose, since in the latter case no L29 protein is produced (Fried et al, 1985). Fleming et al (1989) introduced the mouse L27' gene and yeast *CYH2* separately, both on multicopy plasmids, into such a strain. Comparable growth occurred on galactose and glucose. Thus, at least in the presence of abundant mouse L27', the functioning of the ribosome was sufficient to maintain normal growth, although, mouse L27' is not assembled into ribosomes when yeast L29 is present (Fleming et al, 1989). This demonstrates the strong evolutionary conservation between ribosomal proteins of organisms that have diverged considerably.

On the one hand,the mRNA encoded by *CYH2* has been described to decay with a half-life of approx. 18 minutes (Santiago, 1986; Section 3.2.2.5.). On the other, its half-life has been measured to be approx. 43 minutes (Herrick et al, 1990). Although the latter measurement involved the use of a temperature-sensitive mutation in RNA polymerase II, it is important to emphasise that a heat-shock response does not occur under the conditions used by Herrick et al (1990), since the response requires the transcription of of at least some RNA polymerase II genes (Nonet et al, 1988). Since many ribosomal protein mRNAs appear to be turned over rapidly (Kim and Warner, 1983b; Herruer et al, 1988; Santiago, 1986; Herrick et al, 1990), a half-life of 43 minutes for *CYH2* mRNA does appear to be rather long, but cannot be ruled without further experimentation.

3.3.2.2. cDNA90

The gene that encodes ribosomal protein L3 (*TCM1*; equivalent to cDNA90) was one of the first ribosomal protein genes to be cloned, by virtue of the fact that a mutant form of the gene product confers resistance to the antibiotic trichodermin. Trichodermin inhibits eukaryotic protein synthesis by blocking elongation (Vazquez, 1979). *TCM1* is located on chromosome XV (Grant et al, 1976). Both Fried and Warner (1981), and Shultz and Friesen (1983), cloned the mutant allele of the *tcm1* gene and the latter authors sequenced it. L3 is the largest of the ribosomal proteins, having a molecular weight of approx. 43 kD (Shultz and Friesen, 1983). Unlike many ribosomal protein genes, only one copy is present per haploid genome and, as shown by S1 mapping and confirmed by the nucleotide sequence, the gene does not contain an intron (Shultz and Friesen, 1983).

Many ribosomal proteins have common upstream activation sequences denoted UAS_{rpg}, which are present in either one or two copies (Woudt et al, 1987; see Section 1.3.4.). Such sequence elements are not present in the promoter region of *TCM1*. Instead another element, UAS_T, has been found to be involved in the transcription of *TCM1*. Further, a protein denoted TAF (*TCM1* activation factor) binds to UAS_T *in vitro* (Hamil et al, 1988). In support of a hypothesis that TAF is indeed a transcriptional factor for *TCM1*, nucleotide alterations in UAS_T that diminish transcription *in vivo* also diminish TAF binding *in vitro*. TAF can be distinguished from the UAS_{rpg}-binding protein, TUF (Huet et al, 1987), by chromatography and by virtue of the fact that binding of either TAF or TUF to their respective UAS sequences is not competed by the other UAS. Finally the two proteins have different molecular weights. Thus, in *TCM1*, the UAS and the *trans* acting factor that binds to it are different from those used by most other ribosomal protein genes. Hamil et al (1988) also observed, by photochemical cross-linking methods, that there is an 82 kD protein that binds to both UAS_T and UAS_{rpg},

but only in the presence of TAF or TUF respectively. This additional protein may be the *trans*-acting factor that allows coordinate regulation of the ribosomal protein genes.

As stated in Section 3.1.2.3., mRNA90 has a half-life of approx. 7 minutes (Santiago, 1986; Herrick et al, 1990; Table 3.1.). However, Herrick et al (1990) independently estimated the half-life of *TCM1* mRNA (not realising it to be identical to mRNA90) as 11 minutes. The difference in the two measurements is somewhat larger than the +/- 15% accuracy that Herrick et al (1990) claimed. By their criterion, *TCM1* encodes a moderately stable mRNA and cDNA90 an unstable mRNA. Clearly their measurements are not as accurate as they believe.

3.3.2.3. cDNA13

The complete sequence of cDNA13 was determined (Fig. 3.9.a.). At the nucleotide level the cDNA is 72.1% identical to part of the coding region of an *S. pombe* gene encoding ribosomal protein K37; at the amino acid level identity is 70.7% or 82.9% if conserved changes are allowed. Thus it appears likely that cDNA13 encodes a counterpart of the *S. pombe* protein and is thus a newly cloned ribosomal protein gene of *S. cerevisiae.* cDNA13 does not appear to be a full length cDNA; its 5' end is equivalent to codon 66 of the K37 gene. Thus, in order to obtain the full coding sequence, it was necessary to clone and sequence the gene itself. This will be discussed in Chapter 5. Further discussion of *S. pombe* ribosomal protein K37 can be found in Sections 4.1.1. and 5.1. For the present, suffice to say that this is the third identified cDNA, from the set studied by Santiago (1986), that represents an mRNA that is both unstable and encodes a ribosomal protein.

The fact that 3 of the 8 unstable mRNAs encode ribosomal proteins, while the others have yet to be identified, begs the question as to whether any of the latter may also encode ribosomal proteins?

It was with this in mind that the distribution of acidic and basic residues was examined for each of the putative amino acid sequences. Unfortunately, many of the deduced open reading frames are too short for meaningful analysis. All that can be said is that the putative proteins encoded by cDNA9, cDNA11 and cDNA100 appear to be relatively basic and that, with the exception of cDNA11, which had an open reading frame too short to comment on, the codon bias is similar to that of other highly expressed genes of *S. cerevisiae*, including ribosomal protein genes (see Sharp et al, 1986). However, the bias is not as high as those of the genes encoding ribosomal proteins L29 and L3 (cDNA39 and cDNA90). In contrast, the ORFs of cDNA19 all show a very low bias

towards yeast codon usage (N.B. the sequence information obtained from cDNA19 is unfortunately not continuous), although the putative amino acid sequence obtained from the larger of the two subclones (p19L) is more basic than acidic. Again, however, it must be emphasised that all these sequences were derived from only one strand, which makes it difficult to be confident regarding the ORF analysis. Lastly the sizes of all of the short half-life mRNAs are in the same range as those of mRNAs known to encode ribosomal proteins.

It has been hypothesised that ribosomal protein mRNAs of *S. cerevisiae* will have relatively short half-lifes (Kim and Warner, 1983b). This hypothesis is strengthened by the half-life of *RP29* mRNA (note rp29 and L29 are different ribosomal proteins; see Appendix 2) being approx. 11 minutes (Herrick et al, 1990). However, in the same study, the mRNA encoding *RP51A* was stated to have a half-life of approx. 28 minutes, and thus by the criterion of the authors to be a member of the 'stable' group. The same was true of *CYH2* mRNA (see Section 3.3.2.2.). When the estimated half-life of *RP51A* mRNA is plotted on Santiago's graph (Fig. 3.1.), it is outside the line drawn between the values for unstable mRNAs, but is nevertheless closer to this population than to the values for the stable mRNA population.

3.3.3. Comparison with Other *S. cerevisiae* Genes

URA3 encodes orotidine-5'-phosphate decarboxylase, an enzyme that catalyses the last step in the *de novo* synthesis of pyrimidines (Bach et al, 1979). *URA3* mRNA half-life has been measured by four different methods: by the 1,10-phenanthroline method the half-life was found to be 11 minutes (Santiago et al, 1986); on the basis of decay kinetics, 10.5 minutes (Bach et al, 1979); on the basis of incorporation kinetics, 7.3 minutes (Kim and Warner, 1983b); and finally using the temperature-sensitive PolII mutant, 3 minutes (Herrick et al, 1990). Thus, though there are minor disagreements as to the exact half-life of this mRNA, it most certainly is an unstable mRNA.

URA1 encodes dihydroorotic acid dehydrogenase, an enzyme that catalyses a step in the same pathway as does the *URA3* gene product. On the bases of both decay and incorporation kinetics, the half-life of *URA1* has been found to be approx. 15 minutes (Pelsy and Lacroute, 1984). Within the same pathway, *PPR1* mRNA (pyrimidine pathway regulation) with a length of 2900nt, the product of which regulates both *URA1* and *URA3*; has been shown by incorporation kinetics to have the extremely short half-life of approx. 1 minute (Losson et al, 1983). These results support Santiago's hypothesis of two distinct mRNA populations, within each of which length is inversely proportional to stability (Brown, 1989).

Herrick et al (1990) suggested that mRNAs that are degraded rapidly have a higher proportion of rare codons than do stable mRNAs. Table 3.2. shows the frequency of rare codons in the stable and unstable classes of Santiago (1986). However, Herrick et al (1990) were able to analyse complete open reading frames. Further, Herrick et al (1990) defined a rare codon as one that that is used to encode a particular amino acid ≤1.5% of the time. Here, a maximum of 5% was chosen.

From my sequence data, from those of others, and most importantly from the sequences of cDNAs to which a function can be assigned, what conclusions can be drawn with respect to the biological significance of the two populations of mRNAs in *S. cerevisiae* as suggested by Santiago (1986). Among members of the long half-life class are mRNAs that encode products required in large amounts for normal growth, e.g. pyruvate kinase, enolase, phosphoglycerate kinase and actin. However, the same can be said for at least some of the mRNAs within the short half-life class, as ribosomal proteins constitute up to 15% of cellular protein (Mager, 1988), and their presence is clearly required for normal growth.

Therefore, the distinction between the stable and unstable mRNAs is not merely that between "housekeeping" genes and genes that are required under specific growth conditions. Perhaps there is no general rationale for an mRNA belonging to one population or the other as defined by Santiago (1986); instead, it may be preferable to consider each mRNA in isolation with respect to its unique functional role in the cell. Indeed, the results of Herrick et al (1990) suggest that rather than there being two distinct populations, there is a continuum of half-lifes.

Finally the one cDNA (cDNA82) for which no sequence information is available is one of the eight copied from mRNAs with short half-lifes. cDNA82 was not sequenced because of difficulties encountered in subcloning either the *Pst*I fragment containing the entire cDNA or an internal *Hinc*II-*Pst*I fragment. The reasons for this were not investigated.

mRNA	Frame	Codons Analysed	% Rare codons/ codons
Stable:-			
cDNA10 (*ENO2*)	-	437	0
cDNA22	-	-	FND
cDNA46	1	78	15.4
cDNA74	3	51	7.8
cDNA85	1	27	0
Unstable:-			
cDNA9	1	99	13
cDNA11	1	20	15
cDNA13	1	188	0
cDNA19	-	-	FND
cDNA39 (*CYH2*)	-	149	2
cDNA82	-	-	NS
cDNA90 (*TCM1*)	-	387	1
cDNA100	1	41	18

Table 3.2. Rare codons as a percentage of total codons within the open reading frames deduced for cDNA's derived from short and long half-life mRNAs respectively. The ORF used is in each case the one considered to be most likely to encode a protein, as determined either from homology with a known gene, or from codon bias analysis. The relevant frame is indicated for each cDNA. A rare codon is defined as being $\leq$ 5% of a synonymous codon family. FND, frame not determined; NS, no sequence data.

Note Added in Proof

Recently, the putative amino acid sequence derived from cDNA9 has been used again to search the NBRF protein database. Clear homology was observed to ribosomal proteins S12 from rat and mouse (Lin et al, 1987, J. Biol. Chem., **262**, 14343-14351; Aydne et al, 1989, Nucl. Acid. Res., **17**, 6722.). Thus at least four of the eight short half-life mRNAs encode ribosomal proteins.

The nucleotide sequence derived from cDNA11 has been used again to search the GenEMBL database. Clear homology was observed to a region upstream of the *S. cerevisiae* choline transport protein (Nikawa et al, 1990, J. Biol. Chem., **265**, 15996-16003). The 3'-end of the cDNA11 sequence ends 320nt upstream of the coding region of this gene. Thus cDNA11 remains to be identified.

CHAPTER 4

NUTRIENT UPSHIFT, HEAT-SHOCK AND SPLICING

4.1. Introduction

4.1.1. The K37 Ribosomal Protein Gene Family

The amino acid sequence deduced from cDNA13 is 71% identical, and 83% similar when conservative changes are allowed, to the sequence of an *S. pombe* protein, K37 (Nischt et al 1987). Nischt et al (1986) cloned the K37 gene by virtue of its weak cross-hybridisation with the *TCM1* gene of *S. cerevisiae,* which encodes L3, the largest *S. cerevisiae* ribosomal protein. Additional, if rather circumstantial, evidence that K37 is indeed a ribosomal protein was obtained as follows (Nischt et al, 1986). A recombinant plasmid harbouring the K37 gene was linearised and hybridised, under conditions that favour R-loop formation (Woolford and Rosbash, 1979), with *S. pombe* poly(A)$^+$ RNA. RNA released from hybrid molecules was translated in a wheat germ cell-free translation system, and analysed by polyacrylamide gel electrophoresis. 1D-electrophoresis revealed a 31kD product, within the size range characteristic of ribosomal proteins of *S. pombe*, while 2D-electrophoresis, which can be used to separate the rather basic ribosomal proteins from most other cellular proteins (Warner and Gorenstein, 1978), showed that the 31kD product migrates in the same region as do other *S. pombe* ribosomal proteins. However, no specific assignment could be made, since the relevant area of the gel is rather crowded (N.F. Kaufer, personal communication).

Although the K37 gene was cloned by virtue of its homology with *TCM1* of *S. cerevisiae*, it has been shown in this study that cDNA90, rather than cDNA13, encodes L3 (see Sections 3.2.2.6. and 3.3.2.2.). Moreover, L3 has a molecular weight of 43.5kD (Shultz and Friesen, 1983) rather than 31kD, and *TCM1* has only 38.4% nucleotide identity with cDNA13 over the coding region (Fig. 4.1.; note that only part of the cDNA13 coding region is available for analysis).

At this stage it seemed extremely likely that cDNA13 represents a newly discovered ribosomal protein gene of *S. cerevisiae.* However, there are certain general properties of the expression of ribosomal genes that, if also true of the gene corresponding to cDNA13, would provide additional support for such an assignment of function. In this Chapter I describe studies of gene expression in the context of cDNAs representing several short half-life mRNAs. The background to the studies is described below.

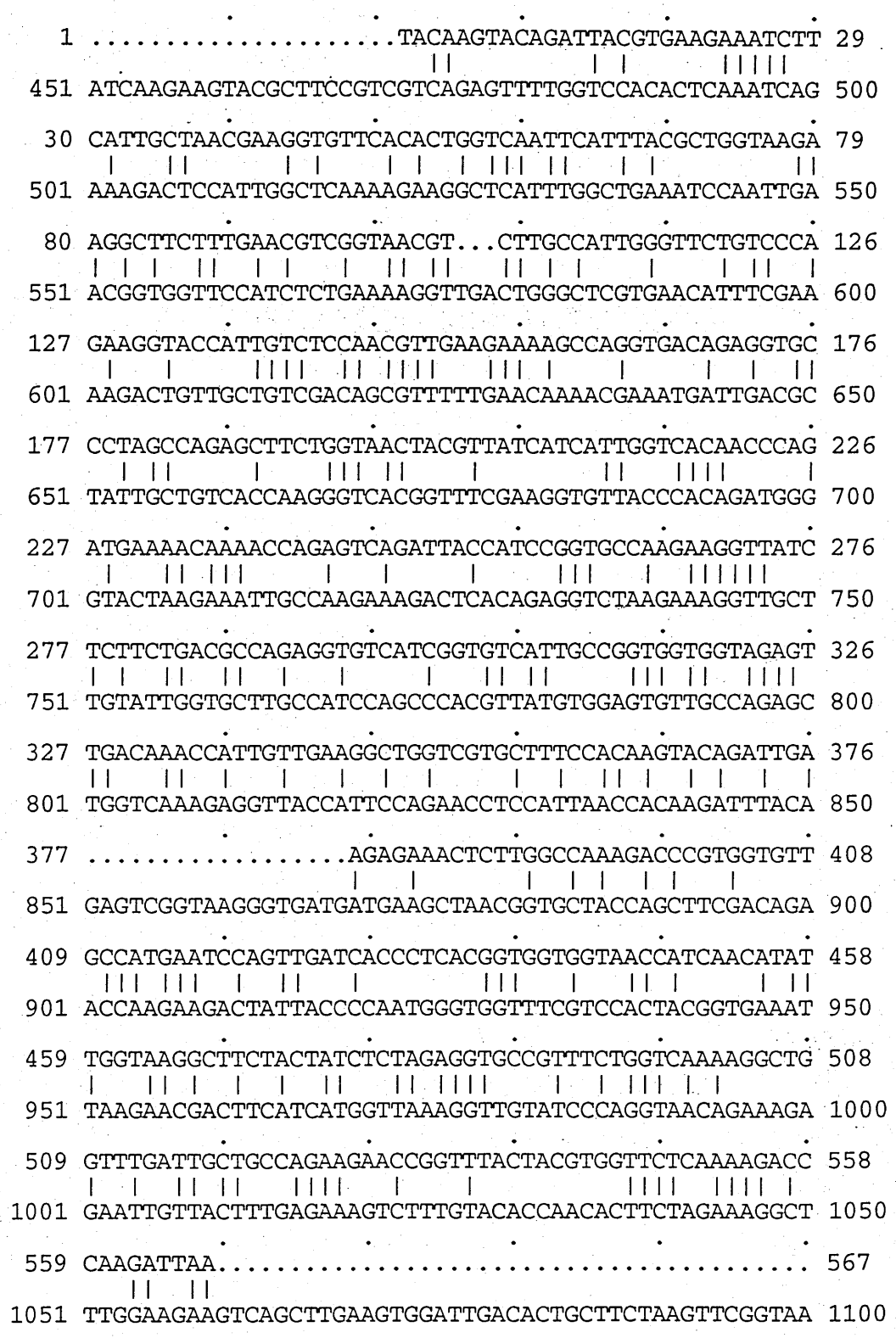


Figure 4.1. Identity between cDNA13 (upper) and *tcm1* (lower; Shultz and Friesen, 1983) at the nucleotide level. Over the region available for anaylysis, the two genes are only 38.4% identical.

4.1.2. Nutrient Upshift

S. cerevisiae is a facultative anaerobe, able to growth oxidatively on ethanol or fermentatively on glucose. The generation time on glucose can be as little as 2 hours, whereas growth on ethanol is roughly three times slower. In turn, ribosome content per cell is proportional to growth rate (Keif and Warner, 1981). A similar relationship has been demonstrated for *E. coli* (Kjeldgaard and Gausing, 1974). Thus, when yeast grown on ethanol are shifted to glucose, there is an immediate increase in the rate of synthesis of ribosomal proteins, such that their abundance rises by 2.5-fold within the first 15 minutes and 4-fold after 1 hour, the latter being a steady-state level characteristic of the new growth medium. rRNA accumulates with similar kinetics. Elevated synthesis of both rRNA and ribosomal proteins begins so rapidly that Keif and Warner (1981) were unable to determine whether one precedes the other. Oddly, cell division is arrested as a consequence of the addition of glucose; only after 60 minutes in the richer medium is the higher growth rate observed. Possibly this is due to a requirement for an increased number of ribosomes per cell before the new growth rate can be established.

Increased levels of ribosomal proteins, after addition of glucose to an ethanol-grown culture, are parallelled by changes in the levels of the respective mRNAs. This reflects changes in the rates of transcription of the respective genes. Take for example, the expression of a hybrid gene formed by fusing a 528 bp domain, derived from sequences upstream of the transcription start site of *RP55A*, to a coding region formed by fusion of a segment of the yeast *CYC1* gene (iso-cytochrome-C) to the *lacZ* gene of *E. coli* (Donovan and Pearson, 1986). The final construct contained no *CYC1* regulatory sequences. The rate of synthesis of β-galactosidase from the gene fusion, after integration into the yeast genome at the *RP55A* locus, increased upon addition of glucose to cells growing in ethanol minimal medium, and with the same kinetics as did the rate of synthesis of several ribosomal proteins. Thus, a glucose response element appears to lie within the 528 bp domain. Further experiments showed that a glucose response element is inseparable from the UAS_{rpg} element found upstream of many yeast ribosomal protein genes, and which is recognised by the *trans*-activation factor TUF (Huet et al, 1985; Section 1.6.). A consensus UAS_{rpg}, introduced by means of a synthetic oligonucleotide into a transcriptionally-inactive reporter gene, confers responsiveness to glucose in addition to restoring transcriptional activity (Herruer et al, 1987).

4.1.3. Heat-Shock

When cells or organisms are exposed to temperatures slightly above the optimum required for their growth, they respond by synthesising a small group of proteins, the heat shock proteins (hsps), which help to protect them from stress. The phenomenon was first observed in *Drosophila* (Ritossa, 1962), and is one of the most highly conserved genetic regulatory systems known. The same set of proteins can be induced by a wide variety of other stresses including exposure to heavy metal ions, ethanol and inhibitors of respiratory metabolism. Thus induction of hsps is a very general response to adverse conditions (see Linquist, 1986).

In the case of *S. cerevisiae*, the nature of the response is influenced by the metabolic state of the cell. The response of cells growing fermentatively at 25°C, for example, is transient at 36°C and sustained at 40°C. The response of respiring cells, on the other hand, is transient at 34°C and sustained at 36°C(Linquist, 1986). In general, the heat-shock response is transient at moderate temperatures, with growth resuming after a temporary pause, while at higher temperatures the effect is sustained and the cells eventually die.

Among 500 individual *S. cerevisiae* proteins analysed with respect to the effect of a mild heat-shock on their synthesis, 80 were found to be transiently induced, more than 300 to be transiently repressed, while the rest were largely unaffected (Miller et al, 1982). Among those that are repressed, the kinetics of repression can vary considerably (Miller et al, 1982). More particularly, heat-shock has a pronounced effect on ribosomal protein synthesis. In the case of cells growing fermentatively, a mild shock causes the rate of synthesis of 40-50 ribosomal proteins to be depressed by 80-90% compared with that of the bulk of cellular protein (Gorenstein and Warner, 1976). Moreover, studies involving *in vitro* translation, followed by 2D-electrophoresis to separate ribosomal proteins from other cellular proteins, indicated that the effect is due to selective depletion of ribosomal protein mRNAs (Warner and Gorenstein. 1977). Repression is only transient, however: after approx. 20 minutes of practically no ribosomal protein synthesis, the cells begin to recover.

Is the depletion of ribosomal protein mRNAs following heat-shock the result of reduced transcription, or of increased mRNA turnover? In fact, both may play a role. Kim and Warner (1983a) measured transcription rates by pulse-labelling cells with [^{3}H]-uracil and quantifying the hybridisation of labelled RNA to cloned ribosomal protein genes. After 10 minutes of a mild heat-shock, the rates of transcription of 5 ribosomal protein genes had decreased to between 10 and 30 % of the rates at 23°C. Following recovery (after 1 hour at the higher temperature), the rates were approx. double those of the 23°C

culture, possibly due to increased efficiency of the polymerase at the higher temperature. The response to heat-shock appears not to be mediated by UAS_{rpg} elements, since *CRY1*, which encodes rp59, and which in mutated form confers resistance to the alkaloid cryptopleurine, can respond to heat-shock even when its UAS_{rpg} elements have been deleted (Larkin et al, 1987). Moreover, introduction of a consensus UAS_{rpg} into a transcriptionally-inactive reporter gene, does not confer a response characteristic of ribosomal protein genes (Herruer et al, 1988). Thus, the promoter element that responds to heat-shock remains to be characterised.

In addition to its effect on transcription, heat-shock seems also to regulate the turnover of ribosomal protein mRNAs. For example, while the half-lifes of S10 and L25 mRNAs do not differ significantly between cells growing continuously at 23°C or at 36°C, both mRNAs decay four-fold faster immediately following a heat-shock. Furthermore, the transcription inhibitor 1,10-phenanthroline blocks this increased turnover, suggesting that a heat-shock-inducible factor is involved (Herruer et al, 1988).

Synthesis of ribosomal RNA does not display a temperature effect, since precursor rRNA continues to be made. Its processing is impeded, however (Warner and Udem, 1972). Processing of rRNA is dependent upon a continuous supply of ribosomal proteins. In their absence unprocessed rRNA is degraded (Warner and Gorenstein, 1977).

4.1.4. Pre-mRNA Splicing

Very few *S. cerevisiae* nuclear genes have introns, possibly due to selective pressure for efficient growth, intensified by millenia of domestication. However, nearly all of the genes that encode ribosomal proteins do have an intron (see Woolford, 1989; Appendix 3). A number of *S. cerevisiae* genes have been identified whose products are involved in the generation of mature mRNAs from their unspliced precursors (pre-mRNAs). Previously known as *RNA* genes (Hartwell et al, 1970), these are now referred to as *PRP* (**p**re-m**R**NA **p**rocessing) genes. Although most of their products are required for spliceosome formation (Lin et al, 1987), an exception is the *PRP2* gene product, which acts after spliceosome assembly and is required for the first step in the splicing reaction (Lin et al, 1987). *PRP2, 3, 4, 8* and *11* have been cloned: all are single-copy genes that are essential for cell viability (Lee et al, 1984; Last et al, 1984; Soltyk et al, 1984; Jackson et al, 1988). Sequencing and/or immunological studies indicate that each encodes a protein (Jackson et al, 1988; Lee et al, 1986; Last and Woolford, 1986; Chang et al, 1988).

Strains bearing temperature-sensitive alleles of *PRP* genes show a dramatic decrease in

ribosomal protein synthesis at the restrictive temperature (Gorenstein and Warner, 1976; Warner and Gorenstein, 1977), as a result of the failure to generate mature ribosomal protein mRNAs. Moreover Northern blot analysis shows that the decrease is generally accompanied by the accumulation of larger RNA species (pre-mRNA; Rosbash et al, 1981).

In the experiments described here, DJy36, an *S. cerevisiae* strain harbouring the temperature-sensitive *prp2-1* mutation, was used to investigate whether any of the genes corresponding to the short half-life mRNAs of Santiago (1986) contain introns.

4.2. Results

4.2.1. Nutrient Upshift

The experimental procedure was essentially that of Herruer et al (1987). DBY746 was grown to stationary phase in EMM supplemented as described in Section 2.6.1., and in the presence of 0.04% glucose. Without the addition of glucose the culture took over 10 days to commence exponential growth. 500µl of the stationary phase culture was used to inoculate 200ml of EMM, as above except for the glucose supplement, and grown at 30°C until an OD_{660nm} of 0.8 was reached. 0.1 volume of 20% glucose was then added, mixed rapidly and incubation was continued at 30°C. Samples (30ml) were taken at 0, 10, 20, 30, 60 and 120 minutes after the carbon-source upshift, mixed immediately with 2 volumes of cold (-70°C) ethanol, and stored at -20°C until required. Total RNA was isolated from each sample (Section 2.8.5.), fractionated on 1.5% agarose/formaldehyde gels (Section 2.11.1), and blotted (Section 2.12.1.). Cell-count on YPG agar was sampled at 1 hour intervals during growth on EMM, and at the same times-points as were chosen for RNA isolation during growth on glucose. Unfortunately the counts varied rather erratically after the nutrient-upshift (data not shown), possibly due to the cells becoming sticky in the presence glucose. When the growth rates of DBY746 in EMM and GMM were measured independently in terms of viable cell counts on YPG agar, however, doubling time in GMM was 80 minutes, while that in EMM was roughly four-fold slower (330 minutes).

In order to assess the loading per lane, a 25 nt oligonucleotide, specific to the *ENO1* gene (Section 2.4.), was used to probe the Northern blots. As discussed in Section 3.3.1., *ENO1* mRNA is expressed constitutively (Holland et al, 1981; Cohen et al 1987). The chosen oligonucleotide is complementary to the 5'-untranslated region of *ENO1* mRNA (residues 2 to 26); when hybridisation is carried out at 42°C and in 40% formamide, it does not hybridise with *ENO2* mRNA (M.J. Holland, personal communication; this study). Fig. 4.2.a. shows a Northern blot probed with both cDNA13 and the *ENO1*-specific oligonucleotide. Even though mRNA loading evidently varies from lane to lane, it is obvious that the abundance of mRNA13 increases on addition of glucose. This result is very similar to results obtained by Herruer et al (1987), using probes specific for ribosomal protein genes L25, S10 and S33.

To investigate the induction of an mRNA known to encode a ribosomal protein, the same experiment was repeated using a cDNA90 probe. cDNA90 encodes ribosomal protein L3, the *TCM1* gene product (Section 3.2.2.6.). The results are shown in

Figure 4.2.a. A Northern blot of RNA isolated at 0 10, 20, 30, 60 and 120 minutes after the addition of glucose (final concentration 2%)to an EMM culture, probed with an *ENO1* oligonucleotide, and with cDNA13.

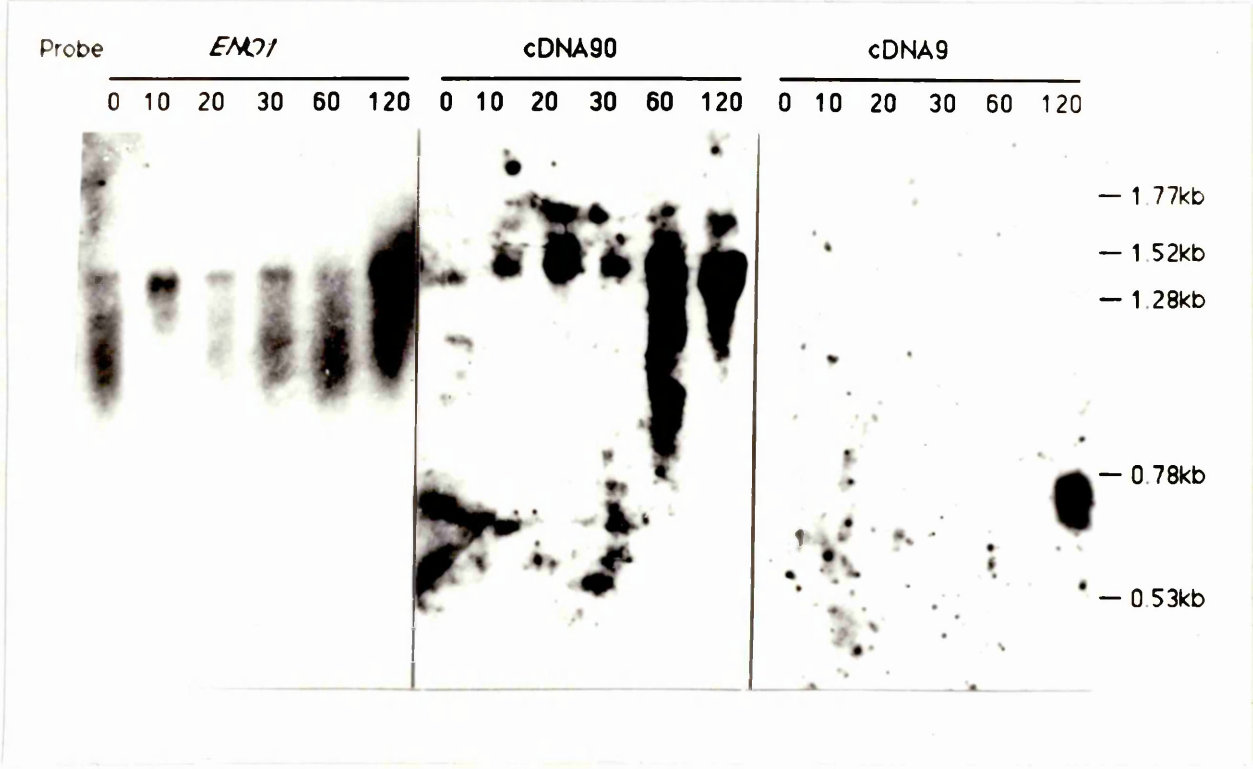


Figure 4.2.b. As Fig. 4.2.a. probed with an *ENO1* oligonucleotide, reprobed with cDNA90 (*TCM1*), and reprobed again with cDNA9.

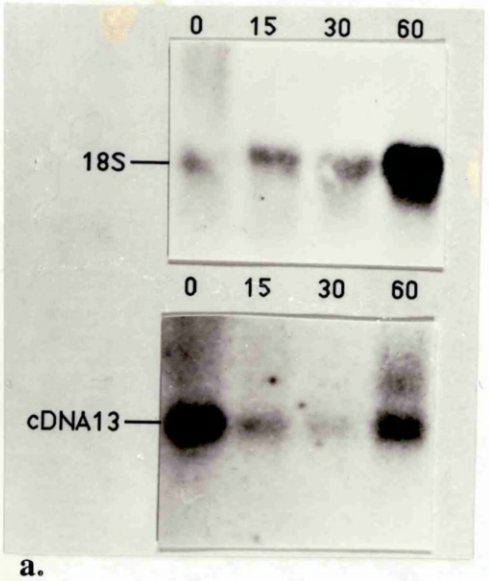
Fig.4.2.b. Again RNA loading is not equal from lane to lane. It also differs from that in Fig. 4.2.a. However, induction of *TCM1* mRNA on addition of glucose is clearly evident and, even though background is rather high, it resembles that of mRNA13. Reprobing the filter with cDNA9 (Fig. 4.2.b.), which represents another short half-life mRNA (Fig.3.1.; Section 3.2.2.1.), indicates that mRNA9 does not begin to be induced until 2 hours after growth on glucose.

4.2.2. Heat-Shock

The experimental procedure was essentially that of Herruer et al (1988). DBY746 was grown at 23°C in GMM plus supplements (Section 2.6.1.), until an OD_{660nm} of 0.8 was reached. An equal volume of fresh medium at a temperature of 49°C was then added, and growth was continued at 37°C. 30ml samples were taken 0, 15, 30 and 60 minutes after the shift, mixed immediately with 2 volumes of cold ethanol (-70°C), and stored at -20°C until required. At the same time as the samples were withdrawn for RNA preparation, a small portion of the culture was labelled for 10 minutes, *in vivo*, with [^{35}S]-methionine in order to confirm that a heat-shock response had occurred (this was carried out by Jane MacKenzie). Protein samples containing equal amounts of radioactivity, and representing each time point, were separated by electrophoresis in a 9% SDS polyacrylamide gel. Autoradiography (not shown) verified that hsps 100, 90 and 70, as well as a number of minor hsps, were present 15 and 30 minutes after heat-shock. Most of the changes in protein synthesis had reverted by 60 minutes, as is expected after a fairly mild heat-shock (Piper et al, 1988). In the knowledge that a heat-shock response had occurred, RNA was prepared from the samples taken at each time point, fractionated on 1.5% agarose/formaldehyde gels, and blotted.

As a loading control the yeast transposon Ty1, which gives rise to a major transcript of 5.7 kb that is expressed constitutively in cells undergoing a heat-shock (Piper et al, 1986), would have been ideal. A blot was probed with pKT15, which contains a Ty1 element within a 9.6 kb *Hin*dIII fragment inserted into pBR322 (P. Piper, personal communication). Unfortunately,the degree of non-specific hybridisation was too high and this approach was abandoned. Instead, hybridisation to 18S rRNA was used to assess loading. Gorenstein and Warner (1976) have shown that synthesis of rRNA continues after a mild heat-shock (from 23°C to 36°C), even though processing of the primary transcript is slowed somewhat.

Fig. 4.3.a. shows that the abundance of mRNA13 declines sharply following heat-shock, and remains low thereafter. The loading representing the 60 minute time point was much higher than the others, and thus the apparent recovery of mRNA13 is



a.

b.

c.

d.

e.

Figure 4.3.a-e. Northern blot of RNA isolated at 0, 15, 30 and 60 minutes after a temperature shift from 23°C to 37°C, probed with the indicated cDNAs, stripped and reprobed with pSP65R (containing an 18S rRNA gene). N.B. RNA representing the 60 minute time point is not present in blot E.

not real. Fig.4.3.b. shows how an mRNA encoding a known ribosomal protein "behaves" under the same conditions. cDNA90 represents *TCM1*, which encodes ribosomal protein L3 (Section 3.2.2.6.; Fig. 4.3.b.). The profiles of mRNA13 and mRNA90 can be seen to be quite similar, although mRNA90 probably recovers more quickly. Three other short half-life mRNAs were studied in the same way (Figs. 4.3.c-e): mRNA9 abundance appears to decline after a heat-shock, but to recover by 30 minutes. mRNA19 behaves similarly. mRNA11, in contrast, is relatively unaffected by heat-shock.

N.B. RNA representing the 60 minute time point in Fig. 4.3.b has run anomalously, as can be seen by the reduced mobility of both 18S rRNA and mRNA90 (c.f. Figs. 4.2.a., c and d). This does not interfere with observations of abundance, however.

4.2.3. Pre-mRNA Splicing

As stated in Section 4.1.4., very few non-ribosomal protein genes of *S. cerevisiae* contain introns, while the presence of an intron in a ribosomal protein gene is more common than its absence. To investigate whether the genes corresponding to various short half-life mRNAs contain introns, their expression was studied in DJy36 (Section 2.2.), which carries the temperature sensitive mutation *prp2-1*. Expression in DJy36 was compared with that in DJy89 (wild type for *PRP2*), which while not the *PRP2* counterpart of DJy36, is nonetheless a good control (Derek Jamieson, personal communication). At the non-permissive temperature, cells bearing *prp2-1* exhibit a dramatic decrease in the abundance of mRNAs derived from genes that contain introns, and an increase in the abundance of their unspliced precursors (Rosbash et al 1981; Teem and Rosbash, 1983).

DJy36 and DJy89 were grown to an OD_{660nm} of 0.4 at 23°C in GMM containing the appropriate supplement (Section 2.6.1.). To half of each culture was added an equal volume of 49°C GMM, and incubation was continued at 37°C. To the other half was added an equal volume of 23°C GMM, and incubation was continued at 23°C. 40 ml samples were taken from these cultures at 0 and 60 minutes. At the latter time point, recovery from the general effects of heat-shock should have occurred. RNA isolated from each sample was fractionated by electrophoresis, and blotted. In order to verify that expression of *prp2-1* had been affected by the temperature shift, one blot was probed with a cloned actin gene. This is one of the few non-ribosomal protein genes of *S. cerevisiae* that contains an intron (Gallwitz and Sures, 1980). As can be seen in Fig. 4.4.a. (lane 5), actin pre-mRNA accumulates in DJy36 at the non-permissive temperature. A small proportion appears either to have been spliced correctly, or to

Figure 4.4.a. Expression of the actin gene in *PRP*$^+$ and *prp2-1* strains. A Northern blot (≈10μg of RNA per lane) was probed with pSPACT9, which contains the yeast actin gene. Lane 1, DJy89 (*PRP*$^+$) 0 min; lane 2, DJy89 (*PRP*$^+$) 60 min at 37°C; lane 3, DJy36 (*prp2-1*) 0 min; lane 4, DJy36 (*prp2-1*) 60 min at 23°C; lane 5, DJy36 (*prp2-1*) 60 min at 37°C.

remain from prior to the temperature shift. In fact, it is surprising that so little mature mRNA remains, since actin mRNA half-life has been measured in one study to be as long as 30 minutes, and in another to be as long as 77 minutes (Herrick et al, 1990; Santiago et al, 1986). Perhaps its half-life decreases following heat-shock, as has been shown for some ribosomal protein mRNAs (Herruer et al, 1988). Assuming equal loading, the abundance of actin pre-mRNA at the restrictive temperature is considerably lower than that of the mature mRNA at the permissive temperature. Such an effect is typical (D. Jamieson, personal communication), and suggests that pre-mRNA is turned over rapidly. Unfortunately, however, accurate quantitation of this effect is not possible in the case of the experiments described here, since loading controls were not carried out.

In the knowledge that DJy36 had accumulated actin pre-mRNA at the restrictive temperature, the effect on the synthesis of mRNAs 9, 11, 13, 19 and 100 (all short half-life mRNAs) was studied. Of most interest is the accumulation of a new RNA species that hybridises with cDNA13 at the restrictive temperature, suggesting the presence of an intron within the corresponding gene (Fig. 4.4.b.). Mature mRNA13 is approx. 980 nt in length, compared with approx. 1300 nt for the putative pre-mRNA. This suggests an intron of approx. 300 nt. On the autoradiograph itself, two higher molecular weight bands that differ insignificantly in intensity between the two lanes are observed at both the restrictive and the permissive temperatures. Unfortunately, due to a photographic artefact, only one of them is visible in the Figure, and then only in one lane. It is possible that these correspond to 25S rRNA (3360 nt) and 18S rRNA (1710 nt; Phillippsen et al 1978b). The same blot, when stripped and reprobed with cDNA100, showed evidence of neither mature nor pre-mRNA at the restrictive temperature (Fig. 4.4.b.). This too may be suggestive of an intron in the corresponding gene, no detectable pre-mRNA perhaps being due to its very low abundance. However, further confirmation would be desirable. Again, two higher molecular weight bands are seen at both the restrictive and the permissive temperatures, perhaps for the same reason as suggested above.

A second blot (an adjacent region of the same gel) was probed first with cDNA19 (Fig. 4.4.b.). Although mRNA19 abundance appears slightly lower at the restrictive than at the permissive temperature, this could well reflect incomplete recovery from heat-shock (Gorenstein and Warner, 1976), or unequal loading. Thus, it seems unlikely that the corresponding gene contains an intron. The blot was stripped and reprobed with cDNA11 (Fig. 4.4.b.). As in the case of cDNA100, neither mature mRNA nor pre-mRNA is detected at the restrictive temperature, again suggesting that the corresponding gene contains an intron. Again, however, in the absence of detectable pre-mRNA, further confirmation would be desirable. The blot was stripped yet again, and probed with cDNA9 (Fig. 4.4.b.). mRNA9 abundance appears considerably lower

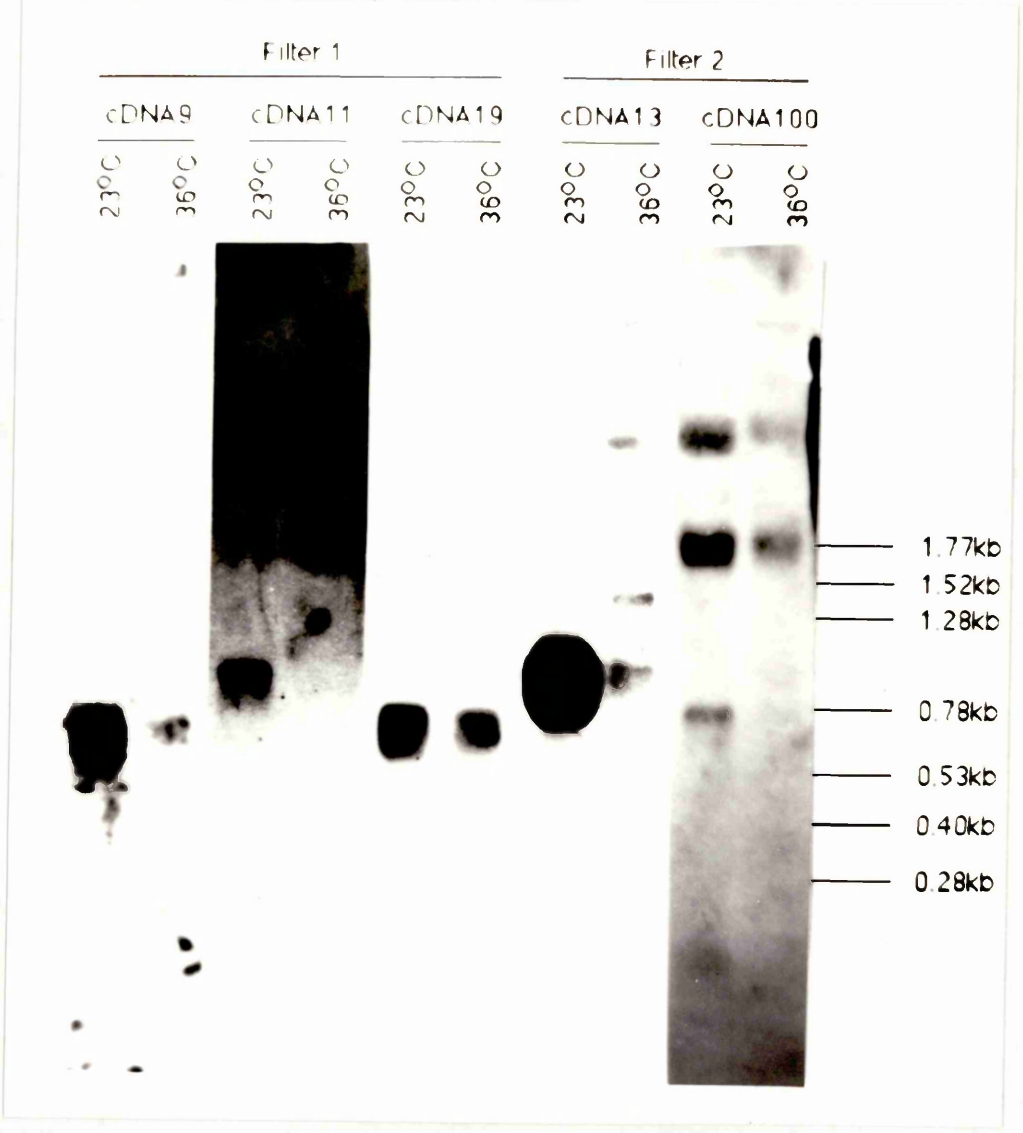


Figure 4.4.b. Effect of a splicing defect on the synthesis of mRNAs 9, 11, 13, 19 and 100. RNA isolated from DJy36 grown at the permissive temperature (23°C), and at the non-permissive temperature (37°C), was fractionated in duplicate and blotted. One blot was probed with cDNA19, followed by cDNA11, and finally cDNA9. The second was probed with cDNA13 followed by cDNA100.

at the restrictive than at the permissive temperature. This may reflect the presence of an intron in the corresponding gene, although no pre-mRNA is evident. Alternatively, it may reflect incomplete recovery from heat-shock.

4.3. Discussion

Nutrient upshift and heat-shock experiments show that mRNA13 abundance is regulated in a fashion similar to that of other mRNAs encoding ribosomal proteins of *S. cerevisiae* (Section 4.1.). In particular, parallel studies of the expression of mRNA90 (*TCM1*) allowed a direct comparison with the expression of a known ribosomal protein gene. It is clear that mRNA13 and mRNA90 are subject to similar control. Further, studies of mRNA13 expression in a temperature sensitive splicing mutant (*prp2-1*), indicate the presence of an intron in the corresponding gene. As discussed in Section 4.1.5., it is rare for an *S. cerevisiae* gene encoding a non-ribosomal protein to contain an intron. These data, together with the short half-life of mRNA13 and, most significantly, the extent of homology observed at both the nucleotide and amino acid levels with K37 of *S. pombe*, support the conclusion that mRNA13 indeed encodes a ribosomal protein.

Even in a wild-type strain, when genes encoding the ribosomal proteins L32 or L29 are present as multiple copies, accumulation of unspliced mRNA occurs (Dabeva et al, 1986). However, not all intron-containing genes, even ribosomal protein genes, behave in the same way (see Warner, 1989), suggesting that the effect is gene-specific rather than being due to a general overloading of the splicing machinery. Indeed, in the case of L32, it seems that the effect is due to excess L32 protein blocking the first step of the splicing reaction, i.e. cleavage at the 5' splice site and formation of a lariat structure (Dabeva et al, 1986). It remains to be seen whether this autoregulation is modulated directly or indirectly by L32. L29, in contrast, appears not to be involved in the regulation of splicing, and the cause of the effect is unknown (Dabeva et al, 1986). It would be of considerable interest to know whether mRNA13 levels are also regulated, in one way or another, at the level of splicing.

mRNA9 recovers rapidly from the general effects of heat-shock (Fig. 4.3.c.), yet in DJy36 it is severely depleted after 1 hour of growth at the restrictive temperature (Fig. 4.4.b.). It seems unlikely that this is due to some secondary effect of the splicing defect, since at least 2 hours at the restrictive temperature are required for there to be significant impairment of translation and tranlsation fidelity (Hartwell et al, 1970). Thus it is tempting to speculate that the gene corresponding to cDNA9 contains an intron, and that pre-mRNA is not detected either because it is turned over rapidly at the restrictive temperature, or because the intron is so short that pre-mRNA cannot be distinguished from the mature mRNA. The latter seems unlikely, however, since although the shortest introns known in *S. cerevisiae* are the 54 nt and 52 nt introns of the *MATα1* gene (Miller, 1984), the majority of *S. cerevisiae* introns are between 200 nt and 500 nt (see

Woolford, 1989). Although mRNA9 is induced 120 minutes after a glucose upshift (from previously undetectable levels; Fig.4.2.b.), this is much slower than expected for an mRNA encoding a ribosomal protein. Similarly, the fast recovery of mRNA9 after heat-shock is not typical of ribosomal protein mRNAs. This being said, however, the abundance of at least one ribosomal protein mRNA, L25, does return to normal levels after 30 minutes of heat-shock (Herruer et al, 1988).

In the case of mRNAs 11 and 100, neither mature mRNA, nor pre-mRNA, was observed in DJy36 grown at the restrictive temperature. This is highly suggestive that both of the corresponding genes contain introns and, since few non-ribosomal protein genes of *S. cerevisiae* contain introns (see Woolford, 1990), that they encode ribosomal proteins. Failure to detect pre-mRNAs may simply reflect their low abundance, due to high turnover. The short half-lifes of mRNAs 11 and 100, together with the rather basic amino acid composition of their putative products, are not at variance with a suggestion that they encode ribosomal proteins. On the other hand, mRNA11 does not appear to be affected by heat-shock. This might be taken as evidence that it does not encode a ribosomal protein. Although the blot was of poor quality, and the experiment should therefore be repeated, the observation is in agreement with the data of Santiago (1986; see Section 1.1.). Unfortunately the effect of heat-shock on the expression of mRNA100 was not investigated.

In contrast with the above, the gene corresponding to cDNA19 appears not to contain an intron. Even this, however, need not exclude the possibility of it encoding a ribosomal protein, since the genes encoding ribosomal proteins L3 and S33 are intronless (Shultz and Friesen, 1983; Leer et al, 1983; Appendix 3). However, mRNA19 is also unlike most ribosomal protein mRNAs in that it recovers relatively quickly from heat-shock.

Data relevant to the question of whether any of the short half-life mRNAs discussed here indeed encode *S. cerevisiae* ribosomal proteins are summarised in Table 4.1.

cDNA	9	11	13	19	39	82	90	100
Corresponding ribosomal protein: *S. cerevisiae*					*CYH2*		*TCM1*	
Corresponding ribosomal protein: other organism(s)	S12 * (R, M)		K37 (S) V1 (Di) L2 (E)		L27'(M) *cyp-1* (N)		L1 (T, A)	
Basic composition?	YES	?	YES	YES	YES	N.D.	YES	?
Glucose-inducible?	YES	N.D.	YES	N.D.	N.D.	N.D.	YES	N.D.
Reduced mRNA after Heat-shock?	YES	NO	YES	YES	N.D.	N.D.	YES	N.D.
Spliced mRNA?	YES	YES	YES	NO	YES	N.D.	NO	YES
Conclusion	++	+	++	+/-	++		++	+

Table 4.1. Summary of information concerning the cDNAs derived from short half-life mRNAs. Abbreviations:- A, *Arabidopsis thaliana*; Di, *Dictyostelium discoideum;* E, *Escherichia coli*; M, mouse; R, rat; S, *Schizosaccharomyces pombe*; T, *Tetrahymena thermophila*; N.D, not done; ?, data not conclusive; ++, encode ribosomal proteins; + and +/-, varying degrees of posibility; *, See note added in proof at end of Chapter 3.

CHAPTER 5

CLONING OF A GENE CORRESPONDING TO CDNA13

5.1. Introduction

cDNA13 appears to represent a previously undescribed ribosomal protein gene of *S. cerevisiae*. Comparison with the sequence of the K37 gene of *S. pombe* (Nischt et al, 1987) suggests that approx. 200bp of coding information is missing at the 5' end, together with any 5'-untranslated sequences. It would clearly be desirable to obtain this missing information, as well as to obtain information regarding potential transcription control sequences. Moreover, K37 is not the only *S. pombe* gene with which comparisons can be made. Kaufer and his colleagues have recently cloned and sequenced two other *S. pombe* genes (K5 and KD4) that are 99% identical with K37 in terms of predicted amino acid sequence (Fig. 5.1., Table 5.1.; Teletski and Kaufer, 1989; Gatermann et al, 1989).

In this Chapter I describe the cloning and characterisation of genomic DNA sequences related to cDNA13.

```
K37   MGRVIRAQRK SGGIFQAHTR LRKGAAQLRT LDFAERHGYI RGVVQKIIHD      50
KD4   MGRVIRAQRK SGGIFQAHTR LRKGAAQLRT LDFAERHGYI RGVVQKIIHD
K5    MGRVIRAQRK SGGIFQAHTR LRKGAAQLRT LDFAERHGYI RGVVQKIIHD

      PGRGAPLAKV AFRNPYHYRT DVETFVATEG MYTGQFVYCG KNAALTVGNV     100
      PGRGAPLAKV AFRNPYHYRT DVETFVATEG MYTGQFVYCG KNAALTVGNV
      PGRGAPLAKV AFRNPYHYRT DVETFVATEG MYTGQFVYCG KNAALTVGNV

      LPVGEMPEGT IISNVEEKAG DRGALGRSSG NYVIIVGHDV DTGKTRVKLP     150
      LPVGEMPEGT IISNVEEKAG DRGALGRSSG NYVIIVGHDV DTGKTRVKLP
      LPVGEMPEGT IISNVEEKAG DRGALGRSSG NYVI[S]VGHDV DTGKTRVKLP

      SGAKKVVPSS ARGVVGIVAG GGRIDN[K]PLLK AGRAFHKYRV KRNCWPRTRG     200
      SGAKKVV[L]SS ARGVVGIVAG GGRIDKPLLK AGRAFHKYRV KRNCWPRTRG
      SGAKKVVPSS ARGVVGIVAG GGRIDKPLLK AGRAFHKYRV KRNCWPRTRG

      VAMNPVDHPH GGGNHQHVGH STTVPRQSAP GQKVGLIAAR RTGLLRGAAA     250
      VAMNPVDHPH GGGNHQHVGH STTVPRQSAP GQKVGLIAAR RTGLLRGAAA
      VAMNPVDHPH GGGNHQHVGH STTVPRQSAP GQKVGLIAAR RTGLLRGAAA

      VEN                                                        254
      VEN
      VEN
```

Figure 5.1. Comparison of the predicted amino acid sequences of ribosomal proteins K37, KD4 and K5 of *S. pombe*. Boxed residues indicate differences. Data taken from Gatermann et al (1989).

	K37		KD4	
	Protein	Nucleotide	Protein	Nucletide
KD4	99.2	92.7	-	-
K5	99.2	92.3	99.2	90.8

Table 5.1 Percentage identity between the three members of the K-family of *S. pombe*, at the nucleotide and amino acid levels. Data taken from Gatermann et al (1989).

5.2 Results

5.2.1. Construction of a Yeast Genomic DNA Library

Genomic DNA isolated from DBY746 (Section 2.8.4.) was used to make a library in EMBL3, a bacteriophage lambda replacement vector. Because of the compatibility of their cohesive ends with those generated by cleavage of the vector DNA with *Bam*HI, yeast DNA fragments were generated by partial cleavage with *Mbo*I. Optimum conditions for generating fragments of approx. 15 kb were determined by a series of trial incubations: 500ng of DNA and 0.8 units of *Mbo*I in a total volume of 80µl for 0, 5, 10, 20, 40, and 80 minutes. The reactions were stopped by the addition of EDTA to 20mM and heating to 65°C for 5 minutes, and were analysed by electrophoresis in a 0.3% agarose gel (Fig.5.2.a.). The 10 minute sample (Lane 4) has the most DNA in the desired range. The reaction was then scaled up as follows: three 40µg samples of DBY746 DNA were cleaved with *Mbo*I, one as determined above, one with half the amount of enzyme and one with twice the amount of enzyme. A small proportion of each was analysed by electrophoresis to verify that the scale-up had been successful (data not shown), following which the three samples were pooled, extracted once with phenol, once with phenol/chloroform, ethanol precipitated and resuspended in 200µl of TE. Although this DNA was rich in fragments in the 15-20 kb range, some fragments larger than 22 kb and some smaller than 9 kb were still present (the max. and min. sizes respectively that can be incorporated into EMBL3). If not removed these would have given rise to recombinant molecules too large to be packaged or, more seriously, to recombinants with multiple inserts. Therefore, fragments in the desired size range were physically separated from the rest by velocity centrifugation through a sucrose gradient.

46ml 10-40% sucrose gradients were poured in two Beckmann SW28 tubes and 100µl of partially digested DNA was loaded onto each. After centrifugation for 18 hours at 26,000 rpm (20°C), 1ml fractions were collected from the bottom of each gradient. 10µl of every other fraction was diluted 5-fold with H_2O and analysed by electrophoresis in a 0.3% agarose gel (data not shown). Fractions containing DNA in the correct size range were precipitated overnight with 2.5 volumes of ethanol in Beckmann SW50 tubes. The DNA was recovered by centrifugation at 45,000 rpm for 1 hour, and was resuspended in 50µl of TE. 5µl samples were analysed by electrophoresis in a 0.3% agarose gel (Fig. 5.2.b.). Fractions 16, 17 and 18 contain DNA of the desired size.

EMBL3 DNA contains sites for *Bam*HI, *Sal*I and *Eco*RI between the arms and the non-essential "stuffer" region. The vector DNA, cleaved with *Eco*RI and *Bam*HI, and

Figure 5.2.a. Partial digests of DBY746 DNA. Lane 1 is intact bacteriophage lambda DNA (≈ 50kb); lanes 2-7 are DBY746 DNA after 0, 5, 10, 20, 40 and 80 minutes of cleavage with *Mbo*I.

Figure 5.2.b. Size fractionated DBY746 DNA.

treated with alkaline phosphatase, was obtained from Promega. The phosphatase treatment prevents the stuffer from competing with insert DNA for ligation to the vector arms. 800ng of size-selected yeast DNA (fractions 16, 17 and 18; Fig. 5.2..) was ligated to 1.6μg of EMBL3 DNA in a total volume of 10μl at 14°C for 16 hours. Recombinant DNA molecules were packaged into phage particles *in vitro* using Gigapack™II Plus (Stratagene), and a proportion was plated on two *E. coli* strains, NM621 and P2PLK-17 (Section 2.1.). Reconstituted vector DNA molecules can grow on the former strain but not the latter, a phage P2 lysogen. Since the titre was the same on both strains, the library consisted almost entirely of recombinants ($3x10^5$ pfu in total). Assuming that the *S. cerevisiae* haploid genome is approx. $1.4x10^7$ bp (Lauer et al, 1977), approx. 320 yeast genome equivalents were present. The library was amplified by growth as plaques on NM621. The titre of the amplified library was approx. 10^{10} pfu/ml.

5.2.2. Screening the Library with cDNA13

In order to isolate the gene(s) represented by cDNA13, approx. 50,000 phage from the amplified library were plated on 5 Petri dishes (150 x 150mm). Plaque-lifts probed with random-primed cDNA13 revealed 20 "positive" signals. Plaques from the corresponding regions were re-plated at 100 pfu per 90mm Petri dish and re-screened: 16 individual positive clones were obtained, although only six were analysed (JMG1-6).

5.2.3. Southern-Blot Analysis of Recombinant DNAs

DNAs prepared from JMG1-6 were cleaved with various enzymes to determine whether the same region of the yeast genome is represented in each. Digests that demonstrate the six clones to be related are shown in Fig. 5.3., together with digests of genomic DNA. The pattern of hybridisation to genomic DNA is more complex than that to JMG1-6 DNAs suggesting that there may be a second yeast locus that shares homology with cDNA13.

5.2.4. Two Yeast Loci Correspond to cDNA13

Further to investigate the nature of the extra bands seen in yeast genomic DNA, a blot of an OFAGE (orthogonal field alternation gel electrophoresis) gel was probed with cDNA13. The blot was kindly provided by Priyal Die Zoysa (University of Strathclyde). A photograph of the gel (Fig. 5.4.a.) shows individual chromosomes from

Figure 5.3.a. Agarose gel of DBY746 DNA and JMG1-6 DNAs cleaved with *Xba*I (1), *Xba*I & *Taq*I (2), and *Taq*I (3). Cleavage with*Xba*I has not gone to completion.

Figure 5.3.b. A blot of the above gel probed with cDNA13. The left panel represents a 5 day exposure, the right panel an overnight exposure, both with intensifying screens at -70^{o}C.

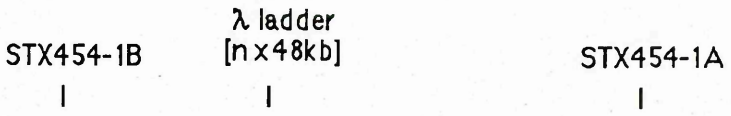


Figure 5.4.a. OFAGE gel showing individual yeast chromosomes.

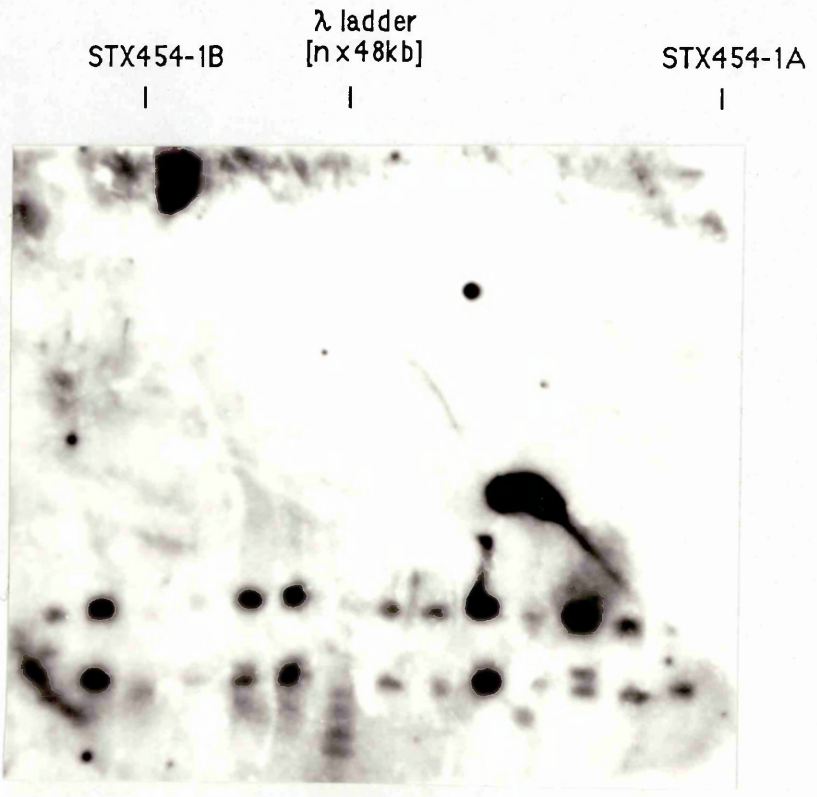


Figure 5.4.b. Blot of above gel probed with cDNA13.

a number of different industrial and laboratory strains of *S. cerevisiae*. In particular, lane 3 shows the chromosomes of strain STX454-IB while lane 14 show the chromosomes of strain STX454-1A. These are both "lab" strains, each carries carrying chromosomal rearrangements that facilitate chromosome identification. By using both STX454-1A and STX454-1B, localisation of a gene to any of the sixteen yeast chromosomes should be possible. Unfortunately lanes 3 and 14 are rather indistinct on the gel. Nonetheless, the blot shows that cDNA13 hybridises with not one but two yeast chromosomes, even though no specific assignments can be made (Fig. 5.4.). This confirms the Southern blot analysis and shows, in addition, that the two loci are unlinked.

5.2.5. Restriction Mapping of JMG2 DNA and Subcloning

JMG2 was chosen for further analysis. A restriction map of JMG2 DNA is shown in Fig. 5.5. Two *Kpn*I fragments (880 bp and 4.8 kb) that hybridise with cDNA13 were subcloned into Bluescript SK$^+$ cleaved with *Kpn*I, and named JMB1 and JMB2 respectively (Fig.5.5).

5.2.6. DNA Sequence Analysis

The sequencing strategy was as shown in Fig. 5.5. Double-stranded rather than single-stranded sequencing was used. The sequence obtained is shown in Fig. 5.6. (see legend for details). This sequence was compared with the sequence of cDNA13 by means of the GCG programme, "GAP" (Fig. 5.7.). GAP uses the algorithm of Needleman and Wunsch (1970). The identity is 89.9% over the complete cDNA13 sequence (i.e. the portion of the coding region that is present in the cDNA together with the 3'-untranslated region), though when only the coding regions are compared the identity is 96.8%. The latter difference is due to 18 nucleotide substitutions, all of which are conservative since the derived amino acid sequences are 100% identical (Fig. 5.8.). Thus, as pointed out earlier, there must be two genes related to cDNA13 in the *S. cerevisiae* genome. However, it is still too early to say whether they encode identical products, since cDNA13 is not a full length cDNA. There may be differences in the N-terminal regions of the two proteins. It is also conceivable that only the gene represented by cDNA13 is transcribed.

Evidence was presented in Section 4.2.3. to suggest that mRNA13 is generated by splicing of a precursor. There is clear evidence in the JMG2 sequence of an intron characteristic of those found in other yeast genes. The putative 5'-splice site,

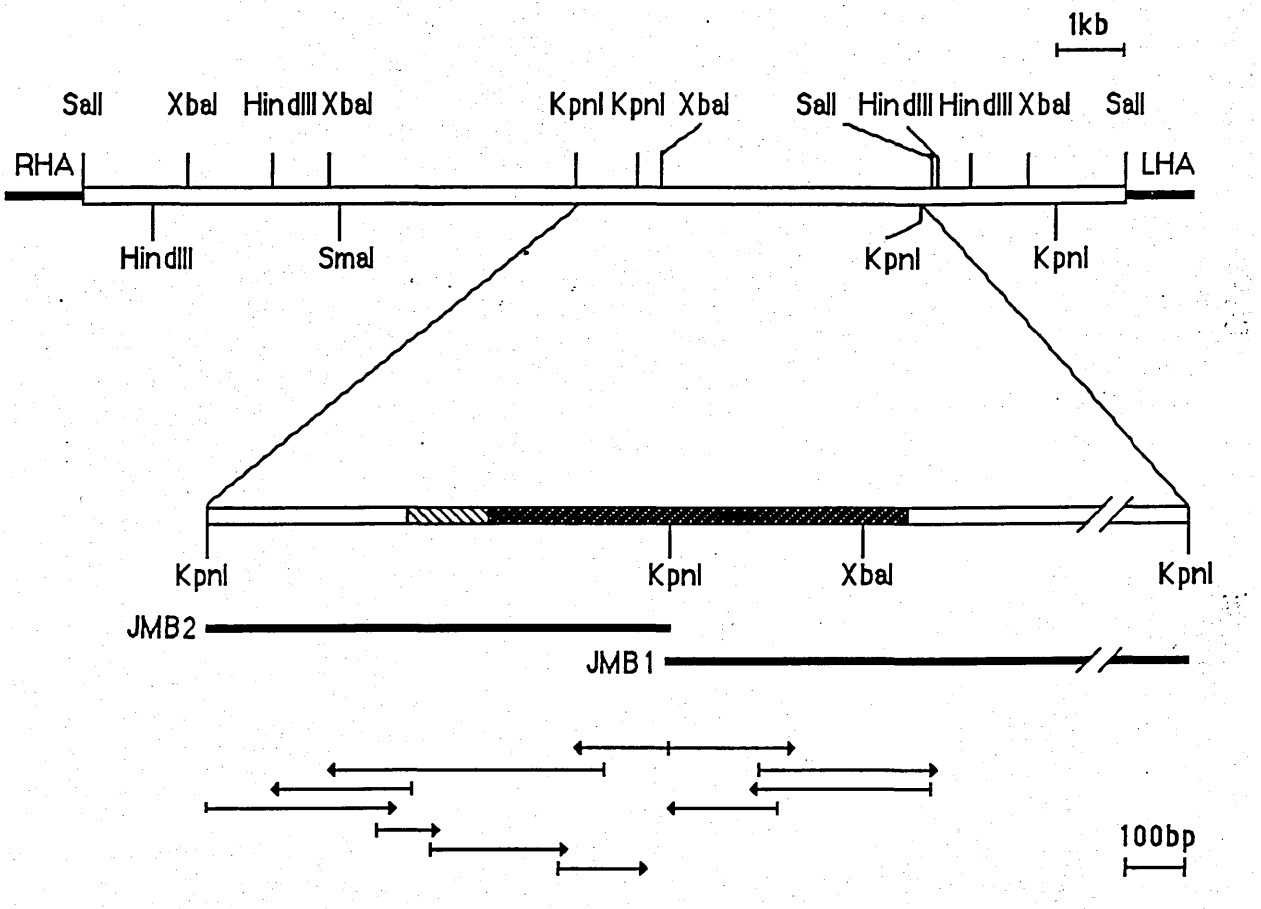


Figure 5.5. Above: Restriction map of JMG2 DNA. Note that since *Xba*I consistently cleaved only partially, thus there may be more sites for this enzyme. The estimated length of the insert is 16.5kb. Below: Subclones JMB1 and JMB2, together with the sequencing strategy. ▬ denotes approx. loctation of coding sequences: and ▨ denotes approx. location of intron. The intron is at the 5' end of the gene.

```
KpnI
GGTACCGTAGGCAGAGGCCTATCCTCATTCTCCCACTGTGTAAACTAGGCGTTCTGCCAGCG    62

GCTAAGCGTTCTCCGCGTATTGTTTCACCGTTGGGCCTACGGTCGATGGTTTGGCATTGC     122

TCTGCACGATCAGATCAATCTGGAATAGTATGGGAATTGCCAAACGCAGCACCAGTGCAC     182

ATGGATGTGAACCTATGGATTACTACGCAATGCACTATCGCATAGGCATAATCGATTGAA     242

CCCTGCCCCTCCCCTCCTTCAATATCATTACCTCGTTTTTTTTATTATTGGGTCATTTTT     302
   1     2        3                                    ↓
TAGCGACTAATAACCATTTCTGAGCAAACTAAGAAACCATTAGATCAATAAGCAATGGGT     362
                                                    MetG(ly)

ATGTCAGCACCGTGATTTAACAACCATGTGATTGTCGTAATGGCGCAATGTCATCACACC     422
a
CTTTTACAGTCTTTTTATTCCTTTTGTTGAAAATGTATCGTGTATACTAACAGAACTGTG     482
                                                 b
TTTTTTCTCATTATTTTTTATTCAGGTAGAGTTATTCGTAACCAAAGAAAGGGTGCTGGT     542
                        ↓
                      c (G)lyArgValIleArgAsnGlnArgLysGlyAlaGly

TCTATCTTTACCTCCCACACCAGATTGAGACAAGGTGCTGCCAAGTTGAGAACTTTGGAT     602
SerIlePheThrSerHisThrArgLeuArgGlnGlyAlaAlaLysLeuArgThrLeuAsp

TATGCTGAACGTCACGGTTACATTCGTGGTATCGTTAAGCAAATTGTCCACGACTCCGGT     662
TyrAlaGluArgHisGlyTyrIleArgGlyIleValLysGlnIleValHisAspSerGly

AGAGGTGCCCCATTGGCCAAGGTTGTCTTCCGTGACCCATACAAGTACAGATTGCGTGAA     722
ArgGlyAlaProLeuAlaLysValValPheArgAspProTyrLysTyrArgLeuArgGlu

GAAATCTTTATTGCTAACGAAGGTGTCCACACTGGTCAATTCATTTACGCCGGTAAGAAG     782
GluIlePheIleAlaAsnGluGlyValHisThrGlyGlnPheIleTyrAlaGlyLysLys
                                                    KpnI
GCTTCTTTGAACGTCGGTAACGTCTTGCCATTGGGTTCTGTCCCAGAAGGTACCATTGTC     842
AlaSerLeuAsnValGlyAsnValLeuProLeuGlySerValProGluGlyThrIleVal

TCCAACGTTGAAGAAAAGCCAGGTGACAGAGGTGCCCTAGCCAGAGCTTCCGGTAACTAC     902
SerAsnValGluGluLysProGlyAspArgGlyAlaLeuAlaArgAlaSerGlyAsnTyr

GTTATTATCATCGGTCACAACCCAGACGAAAACAAGACTAGAGTCAGATTACCATCCGGT     962
ValIleIleIleGlyHisAsnProAspGluAsnLysThrArgValArgLeuProSerGly

GCCAAGAAGGTTATCTCTTCTGATGCCAGAGGTGTCATCGGTGTCATTGCCGGTGGTGGT    1022
AlaLysLysValIleSerSerAspAlaArgGlyValIleGlyValIleAlaGlyGlyGly

AGAGTTGACAAACCATTGTTGAAGGCTGGTCGTGCTTTCCACAAGTACAGATTGAAGAGA    1082
ArgValAspLysProLeuLeuLysAlaGlyArgAlaPheHisLysTyrArgLeuLysArg

AACTCTTGGCCAAAGACCCGTGGTGTTGCCATGAATCCAGTTGATCACCCTCACGGTGGT    1142
AsnSerTrpProLysThrArgGlyValAlaMetAsnProValAspHisProHisGlyGly
```

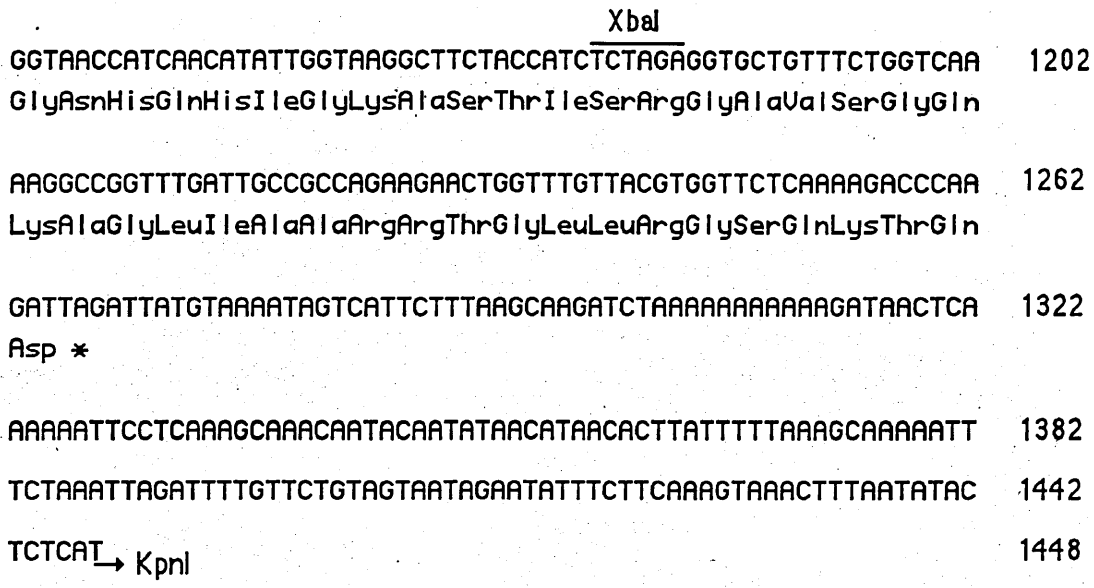


Figure. 5.6. DNA and predicted amino acid sequences of the YL6a gene. Boxed regions in the 5' flanking DNA indicate sequences related to the consensus for UAS_{rpg}s (see text). Regions underlined in bold (a, b and c) are identical to consensus sequences for yeast 5' splice sites, branchpoints and 3' splice sites respectively. Arrows above the 5' and 3' splice sites indicate where splicing would occur. The branchpoint A residue is indicated (see text). Bold overlining indicates the putative TATA box and dashed overlining (1, 2, and 3) shows homology with yeast transcription initiation sequences. Sequence element 3 is related to two different transcription initiation consensus sequences. The putative ATG start codon at position 357 is just upstream of the 5' splice junction; an in-frame stop codon 6 bp upsteam is underlined. *Kpn*I sites used for subcloning as well as an *Xba*I site are shown. The TAG stop codon is indicated by an asterisk. Underlining within the coding sequence indicates the extent of the homology at the amino acid level with the YL6 sequence obtained by Otaka et al (1984). Their sequence lacks the initial methionine residue.

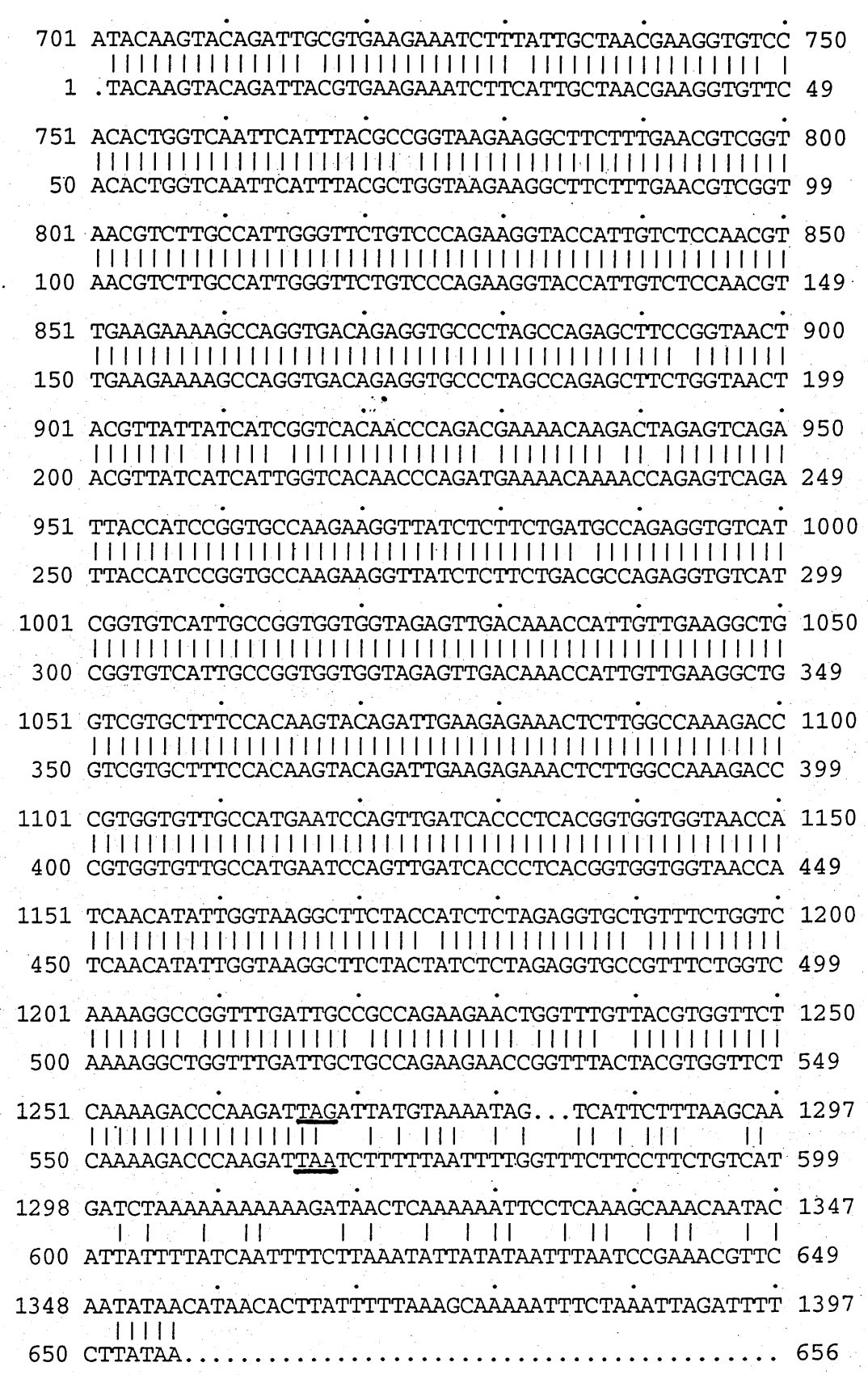


Figure 5.7. DNA sequence of *YL6a* (upper) compared with the sequence of cDNA13 (lower). Underlined nucleotides indicate stop codons. The identity between the two sequences is 96.8% over the coding region, falling to 89.9% when 3'-untranslated sequences are included in the comparison.

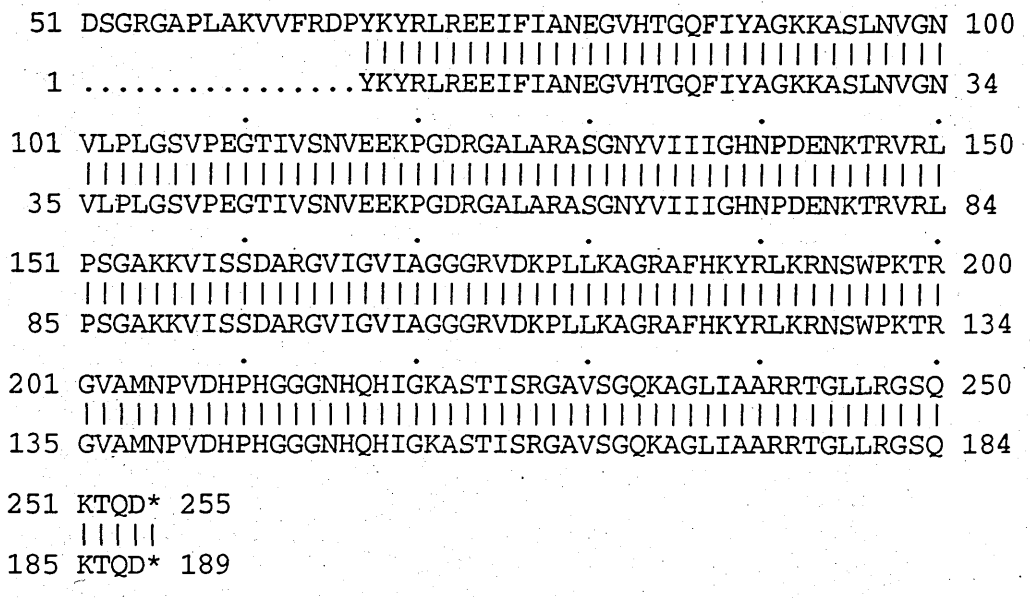


Figure 5.8. Homology between the putative amino acid sequences of ribosomal proteins YL6a (genomic copy; upper) and YL6b (cDNA13; lower), showing that they are 100% identical over the region of sequence available from cDNA13. * represent stop codons.

branchpoint sequence and 3'-splice site (Fig. 5.6.) all match the consensus sequences exactly (Krainer and Maniatis, 1988). The 3'-splice consensus sequence for *S. cerevisiae* is (T/A)NYAG where Y represents a pyrimidine (Teem et al, 1984). The JMG2 sequence is TTCAG. The putative branchpoint is separated by 36 nucleotides from the 3'-splice site, within the range observed for other yeast genes (22-58nt when the *MATα1* and *TUB3* introns are excluded; Krainer and Maniatis, 1988). By analogy with other *S. cerevisiae* genes the A residue indicated in the branchpoint sequence (Fig.5.6.) is where branching would occur. Arrows above the 5'- and 3'-splice sites indicate where splicing and ligation would occur (Krainer and Maniatis, 1988). The length of the putative intron is 147 nt. This is somewhat smaller than the 300 nt anticipated from Northern blots (Section 4.2.3. and Fig. 4.3.b.). The disparity may be due to errors in analysis of the blots, or may indicate that the gene actually corresponding to cDNA13 has a longer intron, and that it is the transcript of this gene that is the predominant species seen on Northern blots.

Given that the above analysis of the position of the intron is correct, it is possible to deduce an amino acid sequence corresponding to the JMG2 gene. Reading backwards through the intron we see that it separates the first and second residues of a glycine codon, immediately upstream of which is a methionine codon. That this is the initiation codon is suggested by the fact there is an frame termination codon 6 nt upstream. Further, when the putative amino acid sequence is compared with that of K37, this methionine is equivalent to the first amino acid residue of the *S. pombe* protein (Fig.5. 13.). Such close proximity of an intron to an initiation codon is typical of *S. cerevisiae* (Woolford, 1989). The next ATG upstream of the putative start codon is 145 bp distant (position 212; Fig. 5.6.).

The amino acid sequence deduced from the JMG2 gene was compared by eye with N-terminal sequences that have been obtained from a number of *S. cerevisiae* ribosomal proteins (Otaka et al, 1984). The first forty amino acid residues of ribosomal protein YL6 (L5) have been sequenced, and are identical to the putative N-terminal sequence of the JMG2 protein (Fig.5.6.), except for the terminal methionine residue. Thus, it seems likely that either the JMG2 gene, or the gene corresponding to cDNA13, if not both, encodes ribosomal protein YL6. To simplify further description, I will tentatively name them *YL6a* and *YL6b*, the former being present in JMG2 and the latter being the as yet uncloned gene represented by cDNA13.

A putative TATA box (TATTAT) is present at position 285 (-72 relative to the putative translation initiation codon). A sequence related to the consensus TATT(T/A)(T/A) is present between -42 and -97 relative to the initiation codon in 7 out of 8 characterised

ribosomal protein genes of *S. cerevisiae*, and in particular the sequence TATTAT is present at position -63 in the gene encoding ribosomal protein L16 (Teem et al, 1984).

The two boxed regions in Fig. 5.6 are candidates for UAS_{rpg} elements. Differences from the UAS_{rpg} consensus sequence are shown in Fig. 5.9. Except for the T at position 272, all of these differences are seen in UAS_{rpg} elements from other yeast ribosomal protein genes (for a compilation of UAS_{rpg}s see Woudt et al, 1987). Even so, there is some doubt that both could be functional. In particular, the gene-proximal element, at -79 relative to the putative start codon, must therefore be less than 100 bp upstream of the transcription start site, the minimum that appears to be required for activation (Woudt et al, 1987). The gene distal element, on the other hand, is 172 bp upstream of the putative start codon. Since the 5'-untranslated region is probably only 30-53 nt (see below), it is reasonable to assume that the gene distal element could be functional.

Sequences of the form RRYRR (R = purine; Y = pyrimidine) can act as transcription start sites in *S. cerevisiae* (Guarente, 1987). Three such sequences are present downstream of the TATTAT element; AGCGA (304), AATAA (311) and AGCAA (325). The last of these overlaps with another sequence that may act as a transcription initiation site, CAAA (327), related to the consensus YAAR (Burke et al, 1983). Transcripts initiated at these sites would have 5'-untranslated regions of 53, 46, 32 and 30 nucleotides respectively, assuming that the proposed methionine codon is indeed the initiation codon. The latter conclusion is strengthened by the observations of Kozak (1984), who has shown that in 95% of eukaryotic mRNAs the first ATG is the initiation codon.

The mean length of an *S. cerevisiae* 5'-untranslated region is 52 nt (taken from a comparison of 131 yeast genes), ranging from 11nt for the *MATα1* gene to 591nt for the *GCN4* gene. Most (70%) are in the range 20-60 nt, and are generally rich in adenine nucleotides (Cigan and Donahue, 1987). However, it has been noted that mRNAs encoding ribosomal proteins have relatively short 5'-untranslated regions, averaging 25 nt and ranging from 15 to 46 nt (Cigan and Donahue, 1987). Short 5'-untranslated regions have also been observed in two of the *S. pombe* counterparts of YL6a. The transcription initiation sites for K5 are at -19 and -30 relative to the initiation codon. The -19 site appears to be used preferentially (Gatermann et al, 1989). Transcription initiation sites for K37 have been mapped to positions -18 and -29 (Nischt et al, 1987).

Two surveys have been carried with the aim of characterising the sequences surrounding *S. cerevisiae* initiation codons. Hamilton et al, (1987) compared 96 yeast genes, while Cigan and Donahue (1987) compared 131. The consensus sequences derived from these

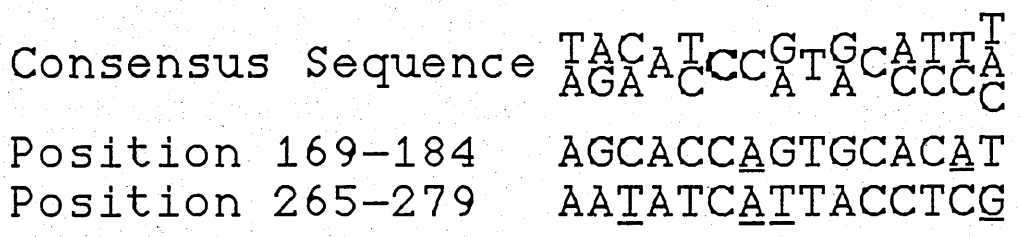


Figure 5.9. Upper: Consensus sequence for *S. cerevisiae* UAS_{rpg} elements (Woudt et al, 1987). An invariant C residue is depicted in bold. Lower: Sequences occuring in the 5' region of YL6 that are similar to the consensus sequence. Underlined residues are those that diverge from the consesus.

studies differ slightly from one another, but both suggest that the most critical residue is at -3 relative to the initiation codon, this being an A in 75% - 81% of genes analysed. Figure 5.10. shows that the region surrounding the putative initiation codon of *YL6a* compares well with the consensus sequences at every position except -3, where a G residue is present. G residues at the -3 position are next most common to A residues (9-13% of yeast genes). Indeed the same three nucleotides (GCA) that are present in *YL6a* immediately upstream of the putative start codon are also found in two other yeast genes, *CAN1* and *CBP1* (see Cigan and Donahue, 1987). Table 5.3. shows a different type of comparison. Here, the frequency of each nucleotide within the 25 bp upstream of the putative translation initiation site of *YL6a*, is seen to be similar to the frequency of each nucleotide within the average yeast gene (Cigan and Donahue, 1987).

cDNA13 probes give rise to what is apparently a single band on Northern blots (Fig. 4.1.a.). This may indicate, a) that only *YL6b* (cDNA13) is transcribed, b) that both genes are transcribed but that the transcripts are the same size, as is the case for the two enolase genes (Holland et al, 1981), or c) that, even though the transcripts are of different sizes, one is present at levels too low to be easily detected on the blots. Given that YL6a has all the hallmarks of a *bona fide* yeast gene, it seems not unreasonable to assume for the moment that it is indeed transcribed. However, further studies will be necessary, including the cloning of YL6b, Northern blotting with gene-specific oligonucleotide probes, 5' end mapping, and perhaps the isolation of other cDNAs, in order to resolve this issue.

The sequence of *YL6a* (Fig. 5.6.) suggests a protein of 254 amino acids, including the N-terminal methionine residue. The approx. molecular weight would be 27.4 kD. The sequence is relatively basic, containing 26 arginine, 19 lysine and 10 histidine residues in contrast to 8 glutamic acid and 9 aspartic acid residues. This is shown graphically in Fig. 5. 11. The basic residues are more or less evenly scattered throughout the protein, except between residues 71 and 122 where there are only 6 basic residues, while 7 of the 17 acidic residues are in this region. The equivalent region of each of the ribosomal proteins belonging to the K-family of *S. pombe* has a similar bias (Gatermann et al, 1989).

5.2.7. Why Has Only One of the Genes Been Cloned?

YL6a was isolated from an amplified rather than a primary library. Possibly YL6b lies within a DNA segment that is not stably propagated in *E. coli*, perhaps due to peculiarities of its sequence or structure. Certainly, *YL6a* was considerably better represented in the library than *YL6b*. This being the case, a differential screening

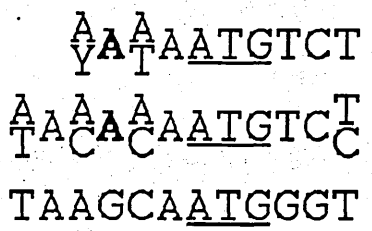


Figure 5.10. Consensus sequences derived from a comparison of sequences surrounding the start codons of a number of different yeast genes. The A residue depicted in bold is the most rigidly conserved residue. Cigan and Donahue (1987) compared 131 yeast genes, while Hamilton et al (1987) compared 96 yeast genes. See text for details.

Nucleotide	Average Yeast Gene (%)	*YL6a* (%)
A	46	52
T	25	20
C	21	16
G	12	12

Table 5.3. The frequency of each nucleotide within the 25 bp upstream of the putative translation initiation site of *YL6a*, compared with the frequency of each nucleotide within the average yeast gene (Cigan and Donahue, 1987).

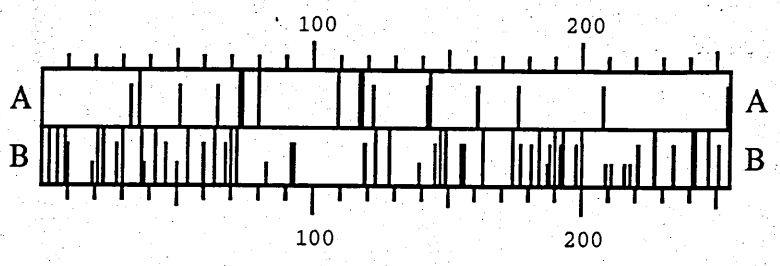


Figure 5.11. Distribution of acidic (A) and basic (B) residues within ribosomal protein YL6a. The acidic residues are glutamic acid (full bar) and aspartic acid (intermediate bar); the basic residues are histidine (small bar), lysine (intermediate bar) and arginine (full bar). Numbers above and below the diagram indicate amino acid positions relative to the putative start codon.

strategy might be a less time consuming way of cloning *YL6b*. The basis of such a strategy (Fig. 5.12.) is that, although *YL6a* and *YL6b* are very closely related, they are likely to be embedded within regions with no significant sequence relationship. Most duplicated ribosomal protein genes are at least unlinked (see Warner, 1989). Thus, the three probes shown in Fig. 5.12.b. should be able to distinguish between clones containing the two genes. In the first instance the probes were used to screen the 16 clones (JMG1-16) originally isolated by virtue of their hybridisation with cDNA13, and from which JMG1-6 had been chosen for further study (Section 5.2.2.). 15 hybridised strongly with cDNA13 and strongly with Probe 2. 6 hybridised strongly with Probe 1. One clone (JMG16) hybridised very weakly with cDNA13, but not with either of the other probes, even though the plaque was of normal size. Since all plaques that hybridised significantly with cDNA13 hybridised also with one or both of the other probes, it seems likely that all contain *YL6a*, although it cannot be ruled out that the initial assumption regarding the unrelatedness of flanking DNA sequences was wrong. Further studies have not been carried out.

5.2.8. Similarities Between YL6a and Ribosomal Proteins From Other Organisms

As discussed above (Section 5.1.), YL6a and YL6b are related to a ribosomal protein family of *S. pombe*, encoded by three independent genes (Nischt et al, 1987; Teletski and Kaufer, 1989; Gatermann et al, 1989). Once the complete amino acid sequence of YL6a was established, a further search of the data base revealed that it also shares significant homology with ribosomal protein V1 of *D. discoideum* (Singleton, 1989). Further, both K37 of *S. pombe* (Teletski and Kaufer, 1989) and protein VI of *D. discoideum* (Singleton, 1989) are rather similar to ribosomal protein L2 of *E. coli*, which is encoded by a gene within the S10 operon (Zurawski and Zurawski, 1985). These similarities are illustrated in Fig. 5.13. and Table 5.4. In Fig. 5.13., all sequences except that of L2 of *E. coli* begin at the (putative) N-terminus and extend to the C-terminus. In L2 there are 29 extra residues at the N-terminus. The degree of relatedness of each of the proteins reflects the degree of evolutionary relatedness of the respective organisms. Examination of the distribution of acidic and basic amino acid residues in V1 and L2 shows both to be very basic proteins, as expected. However, neither protein appear to have the relatively acidic domain seen in YL6 and in the K-proteins of *S. pombe* (Section 5.2.7.; Gatermann et al, 1989).

Figure 5.12.a. A 1% agarose gel. Lanes 1, 4, 5 and 8; λ DNA cleaved with *Hin*dIII: lanes 2 and 6; cDNA 13 cleaved with *Pst*I: lane 3; JMG2 DNA cleaved with *Kpn*I: lane 7, JMG2 DNA cleaved with *Kpn*I and *Sma*I. Arrow heads indicate fragments that were used as probes. Their position in JMG2 DNA is shown in Fig. 12.b.

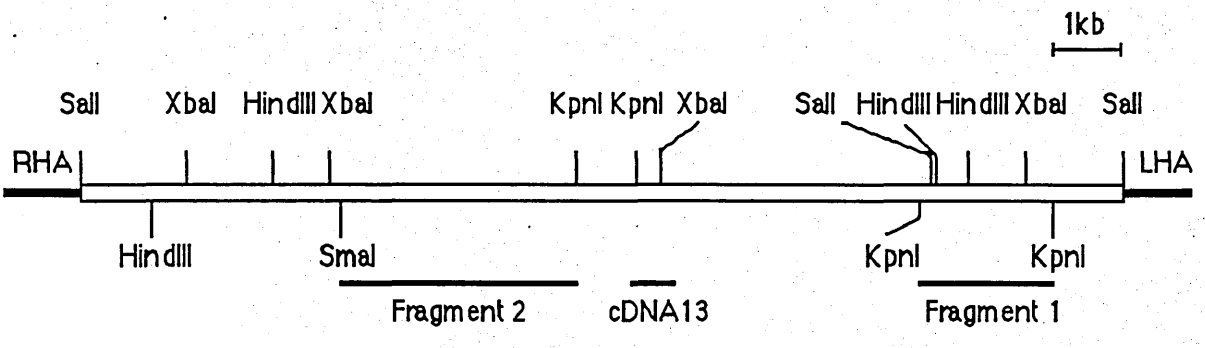


Figure 5.12.b. Restriction map of JMG2 DNA. The regions corresponding to cDNA13, and to the two fragments used as probes, are shown beneath the map. RHA = right hand arm and LHA = left hand arm of EMBL3 DNA.

Figure 5.12.c. A blot of the gel shown in Fig. 5.12.a. probed with Fragments 1 and 2.

Figure 5.12.d. Blots of JMG1-16 plaques probed with cDNA13, Fragment 1 and Fragment 2. 15 clones hybridise with cDNA13; the same 15 clones hybridise with fragment 2; 6 clones hybridise with fragment 1 . The diagram below indicates the position of each recombinant phage; note that some are present in duplicate.

	1	2	3	4	5
1	3	2	2	1	1
2	7	7	6	5	4
3	10	9	9	8	8
4	13	12	11	11	10
5			16	15	14

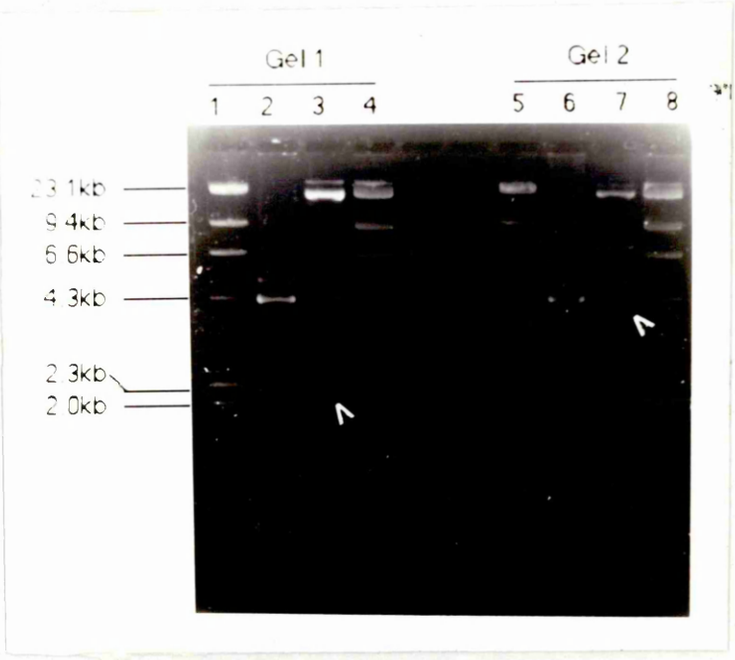


Figure 5.12.a.

Figure 5.12.c.

Figure 5.12.d.

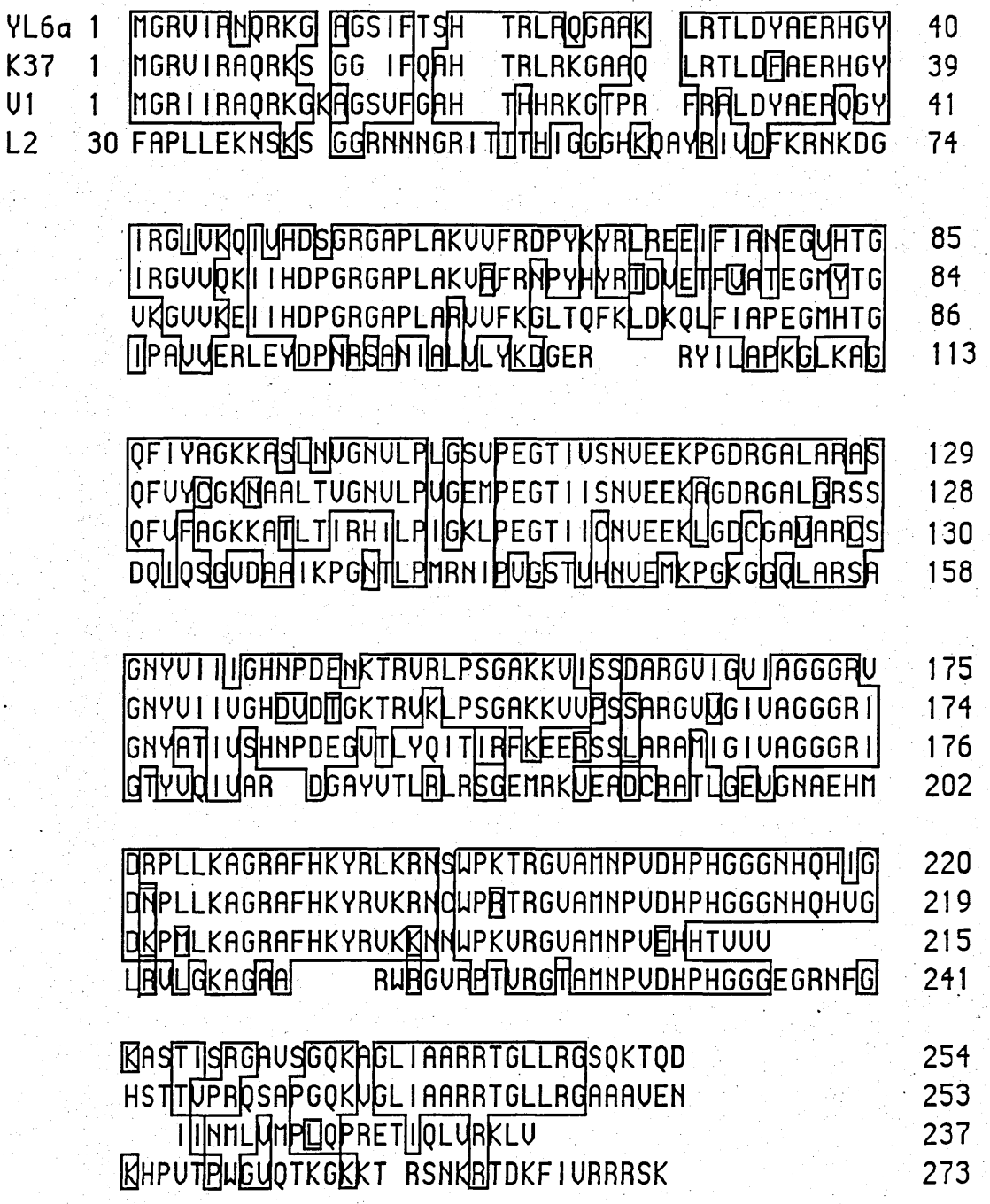


Figure 5.13. Comparison of the amino acid sequences of ribosomal proteins YL6a, K37 of *S.pombe* (Nischt et al, 1987), V1 of *D. discoideum* (Singleton, 1989) and L2 of *E.coli* (Zurawski and Zurawski, 1985). Boxed areas show identities.

	K37		V1		L2	
	Identity	Similarity	Identity	Similarity	Identity	Similarity
YL6a	72.2	83.7	56.2	75.3	32.6	54.4
K37	-	-	56.7	74.9	33.5	55.0
V1	-	-	-	-	29.7	58.0

Table 5.4. Percentage identity and similarity (when conservative changes are allowed) between ribosomal proteins YL6a of *S. cerevisiae*, K37 of *S. pombe* (Nischt et al, 1987), V1 of *D. discoideum* (Singleton, 1989) and L2 of *E. coli* (Zurawski and Zurawski, 1985).

5.3. Discussion

There are at least two genes in *S. cerevisiae* related to cDNA13. The one that has been cloned and sequenced (*YL6a*) would appear to encode a protein, the putative N-terminal sequence of which is identical to that of ribosomal protein YL6, a component of the large subunit (Otaka et al, 1984). YL6a would be a rather basic protein (Fig. 5.11.), of 254 amino acid residues in length (approx. 27.4 kD). Even though the deduced amino acid sequences are the same, the nucleotide sequences of *YL6a* and cDNA13 are only 96.8% identical over their coding regions, and diverge even more in their 3'-untranslated regions. Thus cDNA13 must represent the second gene (*YL6b*). cDNA13 is not complete at its 5' end, and is missing sequences that would encode the N-terminal domain of the protein. Thus we cannot yet be sure that *YL6a* and *YL6b* encode identical proteins. On the other hand most ribosomal protein genes of *S. cerevisiae* are present in duplicate (see Warner, 1989).

One important question remaining to be asked is whether both genes are functional. The gene corresponding to cDNA13 (*YL6b*) does appear to be functional, to the extent that it is transcribed. At present we do not know whether *YL6a* is transcribed. However, the fact that pseudogenes have never been found in *S. cerevisiae* (Fink, 1987), together with the evidence within the *YL6a* sequence of promoter-like elements, strongly suggests that *YL6a* is a functional gene. In addition, all three members of the K-family of *S. pombe* ribosomal protein genes are transcribed (although here it remains possible that there are functional differences between the three proteins; Gatermann et al, 1989).

The question of whether both copies of a ribosomal gene pair are functionally equivalent was addressed by Abovich and Rosbash (1984), who disrupted individual copies of the *S. cerevisiae RP51* genes. Deletion of both genes led to spore inviability, indicating that rp51 is an essential ribosomal protein (Abovich et al, 1985). However, they found that single mutants lacking only one copy were viable, albeit with a growth-rate slower than wild-type. Further, although *RP51b* contributes only 30-40% of *RP51* mRNA and protein, neither the gene nor its product are intrinsically inadequate, as extra copies of the gene restore the growth rate to wild-type levels in a strain deleted for *RP51a*.

Similarly, ribosomal protein L16, also an essential component of the ribosome (Rotenberg et al 1988), is encoded by two unlinked genes, both of which are transcribed. Single *L16a* or *L16b* mutants are viable but grow more slowly than wild-type. *L16b* normally gives rise to approx. twice the amount of mRNA and protein as does *L16a*. However, when an extra copy of *L16a* is introduced into an *L16b* mutant, wild-type growth rate is restored, indicating that the two proteins are functionally interchangeable (Rotenberg et al, 1988). Strains that have lower than a critical level of

either rp51 or L16 have insufficient numbers of 40S or 60S ribosomal subunits respectively (rp51 is a component of the small subunit and L16 a component of the large subunit). Decreased amounts of one ribosomal subunit have little effect on the other (Abovich et al, 1985; Rotenberg et al, 1988).

The above examples suggest that there has been evolution of the promoter strength of the duplicated ribosomal protein genes, such that the combined strength is equivalent to that of the promoter of a ribosomal protein gene present as a single copy (Section 1.5.). Perhaps duplication is a safety measure to ensure the production of these essential proteins, although while a number of essential housekeeping proteins, such as histones, ubiquitin and tubulin, share the property of being encoded by duplicated genes, many do not.

It should also be mentioned that not all ribosomal protein genes are essential for cell viability. A ribosomal protein gene that is fused to one of the genes encoding ubiquitin (*UBI3*) appears to be inessential, although its absence leads to very slow cell growth, probably resulting from the inefficient processing of 20S to 18S RNA, and hence in substantial deficiency in 40S ribosomal subunits (Finley et al, 1989). Another example is the ribosomal protein L30, which is encoded by two genes, both of which are functional but utilised to different extents. Cells are still viable when both copies are disrupted, doubling time being increased by only 30% (Baronas-Lowell and Warner, 1990). However, stalling of translation initiation complexes occurs, as well as a delay in the processing of rRNA. Inessential ribosomal proteins have also been described in *E. coli*. Here, at least 13 of the 52 ribosomal proteins are inessential, and a strain lacking as many as four ribosomal protein genes is viable, albeit barely (Dabbs et al, 1986).

The sequence of YL6a was analysed for evidence of upstream activation sites, TATA-like elements, transcription initiation sites, and sequences typical of those surrounding ATG start codons. Evidence for all such elements was found, suggesting that the gene is indeed expressed. In the absence of further experimentation, however, it is not possible to be sure that any individual element is functional.

The sequence of *YL6a* confirms the presence of an intron, first suggested by the experiments described in Section 4.2.3. Unlike in higher eukaryotes and *S. pombe* (Mount, 1982; Padgett et al, 1986; Russell and Nurse, 1986), *S. cerevisiae* 5' and 3' splice sites and branchpoint sequences generally deviate little from the respective consensus sequences (Langford et al, 1984; Leer et al, 1984; Teem et al, 1984). *YL6a* is no exception in this respect (Fig. 5.6.; reviewed in Woolford, 1989). However, the size of the intron (147 nt) is less than is generally observed for other *S. cerevisiae* ribosomal

protein genes (typically between 300 to 500 nt; see Woolford et al, 1989). Further, the length of the intron estimated by Northern blot analysis was approx. 300 nt. The discrepancy may be due to inaccuracies in measurement, or to *YL6b* both containing a longer intron and giving rise to the predominant transcript. Differences in the lengths of introns between other duplicated ribosomal protein genes have been observed (see Woolford, 1989). Interestingly, there is no evidence of introns in any of the three *S. pombe* counterparts of YL6 (Gatermann et al, 1989).

From the sequence of *YL6a* it would appear that the intron is located after the first nucleotide of codon 2. The position of the intron in *YL6b* remains to be determined. Although in most cases of *S. cerevisiae* gene pairs containing introns, the position of the intron is the same in both genes, there is at least one case where the introns are located at different positions (Molenaar et al, 1984). All but one (Miller, 1984) intron-containing gene of *S. cerevisiae* contains only a single intron, and in most cases it is located within 10 codons of the translation start site (see Woolford, 1989).

Given that so few *S. cerevisiae* genes have introns, why do they appear to have been preferentially retained in ribosomal protein genes. At least one ribosomal protein (L32) can inhibit the splicing of its own pre-mRNA (see Section 4.3. ; Dabeva et al, 1986). Since ribosomal proteins assemble into ribosomes in the nucleus, which is also where splicing occurs, such autoregulation may be the functional equivalent of the autoregulation of ribosomal protein translation which occurs in *E. coli* (Nomura et al, 1984; Section 1.6.). Thus, if autoregulation of splicing of ribosomal protein pre-mRNAs were the rule, it would not be surprising that their introns had been preferentially retained. However, such autoregulation has not yet been observed in cases other than that of L32 (see Warner, 1989). Autoregulation of the splicing of ribosomal protein pre-mRNAs may occur in other eukaryotes. For example, *Xenopus* oocytes injected with excess copies of the gene for ribosomal protein L1 accumulate transcripts in which the second intron is aberrantly spliced, and the third intron is not spliced at all (Bozzoni et al, 1984).

5.3.1. Suggestions for Future Work

Clones containing *YL6b* were significantly under-represented in the genomic DNA library. Should the cloning of *YL6b* by the differential screening method described in Section 5.2.7. prove problematic, another approach would be to screen the library with an oligonucleotide specific to the 3'-end of *YL6b* (cDNA13). Cloning of YL6b would allow direct comparison of the two members of the gene pair.

With regard to questions of function, it would first be desirable to determine whether both genes are expressed. This could be done by using gene-specific oligonucleotides to probe Northern blots. Further studies would include 5' end mapping, promoter analysis, and gene disruption.

The *S. cerevisiae* ribosome is composed of 4 rRNA molecules and 70 to 80 ribosomal proteins. To transcribe the respective genes requires all three nuclear RNA polymerases. Post-transcriptional events include processing of pre-rRNA, splicing of pre-mRNA, transport of mature mRNA to the cytoplasm, regulation of mRNA stability, translation, and transport of the ribosomal proteins into the nucleus. There, assembly of the ribosome takes place, followed by its export back to the cytoplasm. To understand completely the synthesis, assembly, structure and function of the ribosome will require the cloning and characterisation of all of the genes encoding the various ribosomal components. The studies described here represent a small contribution towards this end.

APPENDICES

APPENDIX 1

Abbreviations

Chemicals

BSA	-	bovine serum albumin
DEPC	-	diethylepyrocarbonate
DTT	-	dithiothreitol
EDTA	-	ethylenediaminetetra-acetic acid (disodium salt)
EtBr	-	ethidium bromide
SDS	-	sodium dodecylsulphate
SSC	-	standard saline citrate
Tris	-	tris (hydroxymethyl) amino ethane
TEMED	-	N,N,N',N'-tetramethylethylenediamine

Nucleic Acids

dATP	-	deoxyadenosine triphosphate
dCTP	-	deoxycytidine triphosphate
dGTP	-	deoxyguanosine triphosphate
dTTP	-	deoxythymidine triphosphate
DNA	-	deoxyribonucleic acid
cDNA	-	complementary DNA
RNA	-	ribonucleic acid
mRNA	-	messenger RNA

Miscellaneous

bp	-	base pair(s)
℃	-	degrees centigrade
Ci	-	Curie
cpm	-	counts per minute
kb	-	kilo base(s)
kD	-	kilo dalton
log	-	logarithm/logarithmic
M	-	molar (moles per litre)
MW	-	molecular weight
nt	-	nucleotides
OFAGE	-	orthogonal field alternation gel electrophoresis
pfu	-	plaque forming units
pH	-	negative $\log_{10}$ $[H^+]$
rpm	-	revolutions per minute
UV	-	ultra-violet
v/v	-	volume per volume

APPENDIX 2

40S Ribosomal Proteins			60S Ribosomal Proteins		
A[1]	B[2]	C[3]	A[1]	B[2]	C[3]
S0	YS2	-	L1a	YL3	rp3?
S1	YS1	-	L1b	-	-
S2	YS8	rp14	L1c	YL31	rp4?
S3	YS3	rp13	L2	YL2	rp2
S4	YS5	rp12	L3	YL1	rp1
S6	YS7	-	L4	YL5	rp6
S7	YS6	rp5	L5	YL6	rp8
S10	YS4	rp9	L6	YL8	rp11
S11	?	rp40	L7	YL9	-
S12	YS10	rp30	L8	YL11	rp25
S13	YS11	rp21	L9	YL7	-
S14	YS9	rp19	L10	YL12	rp16
S16a	YS16	rp55	L12	YL13	-
S17	YS17?	-	L13	YL10	rp15
S18	YS12	rp41	L15	YL23	-
S19	YS13	-	L16	YL22	rp39
S20	YS21	rp42	L17	YL16	rp18?
S21	YS18?	rp52	L17a	YL32	-
S22	YS19?	-	L18	YL18	-
S24	YS22	rp50	L19	YL19	-
S25	-	-	L20	YL17	-
S26	YS25	-	L21	YL15	rp22
S26a	YS26	-	L22	YL20	rp33?
S27	YS20	rp61	L23	YL14	rp33?
S27a	YS15	-	L25	YL25	-
S28	YS14	rp37	L26	YL28	-
S31	YS23	rp45	L29	YL24	rp62?
S32	YS28	-	L30	YL21	rp29
S33	YS27	-	L31	YL29	-
S36	YS29	-	L32	YL38	rp73
S37	YS24	-	L33	YL33	-
			L34	YL36	-
			L35	YL30	-
			L36	YL34	-
			L37	YL37	rp47
			L38	YL26	-
			L39	YL39	-
			L40	-	-
			L41	YL27	-
			L42	YL42	-
			L43	YL35	-
			L44/45	YL44	-
			L46	YL40	-
			L47	YL41	-

Correlation of different yeast ribosomal protein nomenclatures.

[1] Bollen et al (1981); Michel et al (1983).
[2] Otaka and Osawa (1981).
[3] Warner and Gorenstein (1978).

APPENDIX 3

Ribosomal Protein	Predicted Size of Protein	Copy No. of Gene	No. of UAS_{rpg} elements	No. of Introns	Counterparts in other organisms
L2	362	2	1	0	L1 (X, D)
L3 (*TCM1*)	387	1	0	0	L1 (T, A)
L5 (YL6)	254	2	1 ?	1	K-family (S); V1 (Di) [c]
L16	174	2	2	0	L5 (E); L21 (T); V18 (Di) [d]
L17a	137	2	1	1	?
L25	137	1	2	1	L23 (E)[a]
L29 (*CYH2*)	148	1	2	1	L29 (N); L27' (M)
L30	155	2	1?	1	? [b]
L32	105	1	2	1	L30 (M, R)
L34	112	2	2	1	L31 (R)
L41	103				L36a (H, R)
L43	88				L37 (R)
L46	51	2	1	1	L39 (R); L36 (S)
L?(*UBI1*,*UBI2*)	52	2	2	1	-
rp28	185	2	2	1	L18 (R); L14 (X)
rp29	155	2	2	1	?
A0	312	1			PO (H)
A1	106	1			-
L44 (A2)	106	1	2	0	-
L44'	106	1	2	0	-
L45	110	1	2	0	-
S10	236	2	2	1	S6 (H, M, R, S)
S16A	143	2	2	1	S12 (Hal)
S24	129	2	2	0	S16 (Hal)
S26	87				S21 (R); S28 (S)
S31	108	2	2	1	?
S33	67	2	0	0	?
S37 (*UBI3*)	76	1	1	1	S27a (R)
rp51	136	2	2	1	S17 (H, R, Ha, Ch)
rp59 (*CRY1*)	137	2	1	1	S14 (H, Ha, D)

Characteristics of *S. cerevisiae* ribosomal proteins and the genes that encode them. The genes that encode L41, L43 and S26 have not been cloned. The predicted sizes of the proteins are in the most part deduced from the DNA sequences. Abbreviations: A, *Arabidopsis thaliana*; Ch, chicken; D, *Drosophila melanogaster*; Di, *D. discoideum;* E, *E. coli*; H, human; Hal, *Halobacterium morismortui*; M, mouse; R, rat; S, *S.pombe*; T, *Tetrahymena thermophila*; X, *X. laevis.*

[a] Not unequivocal from sequence comparison but clear from RNA-binding studies (El-Baradi et al, 1985).

The data are taken from Warner (1989) and:-

[b] Baronas-Lowell and Warner (1990)

[c] This study.

[d] Singleton et al (1989).

APPENDIX 4

Measurement of mRNA Half-Life

Several methods have been used for the estimation of mRNA half-life. However, each has its associated draw-backs. The methods are discussed below, mainly in the context of *S. cerevisiae*, and beginning with those in which inhibition of transcription is followed by the assessment of mRNA abundance at various intervals by means of hybridisation with gene-specific probes. Inhibition of transcription can be achieved either by the use of drugs, or by making use of temperature-sensitive mutations that disable RNA polymerase II.

Various drugs can be used to inhibit transcription in *S. cerevisiae*. These include lomofungin, daunomycin, ethidium bromide, thiolutin and 1,10-phenanthroline, all of which have been used in studies of mRNA half-life (Kuo et al, 1973; Tonnesen and Friesen, 1973; Herrick et al, 1990; Santiago, 1986). In particular, 1,10-phenanthroline, a zinc chelator, was used by Santiago (1986) in his studies of the cDNAs that form the subject matter of this thesis. However, 1,10-phenanthroline is not as specific as might be desirable, since its effect is not restricted to RNA polymerase II, the polymerase that transcribes protein-coding genes. In fact, all three RNA polymerases are inhibited by 1,10-phenanthroline, since all three are zinc-metalloenzymes (Auld et al, 1976; Lattke and Weser, 1976; Wandzilak and Benson, 1977). Further, either by a knock-on effect of the inhibition of RNA synthesis, or perhaps more directly, 1,10-phenanthroline inhibits protein synthesis. Whether or not an mRNA is being translated is known to be an important determinant of the stability of at least some mRNAs. Moreover, inhibition of protein synthesis will reduce the levels of any protein factors involved in mRNA turnover. Of necessity, therefore, the concentration of 1,10-phenanthroline used in mRNA half-life experiments is a compromise that minimises inhibition of protein synthesis while maintaining a significant effect on transcription. Despite all of these reservations, however, the half-lifes of *URA3* and *TCM1* mRNAs estimated by use of the phenanthroline method are not significantly different from the half-lifes estimated by use of the non-disruptive incorporation kinetics method (Santiago et al, 1986; Losson and Lacroute, 1979; Kim and Warner, 1983).

Transcription can also be inhibited by shifting *S. cerevisiae*, bearing a temperature-sensitive mutation (*rpb-1*) in the large subunit of RNA polymerase II, from the permissive to the restrictive temperature (Nonet et al, 1987; Herrick et al, 1990). This mutation has no effect on the other two RNA polymerases, and the potential complication of the heat-shock is minimised since transcription by RNA polymerase II is

essential in order for a heat-shock response to occur. Nonetheless, as in the case of the methods involving transcription inhibitors, there will inevitably be a knock-on effect on protein synthesis, and in any case dilution into hot medium must perturb the physiology of the cells due to the general temperature dependence of enzyme activity. Thus although there may be fewer complications with this method than with the methods involving transcription inhibitors, the results may not be directly comparable. Perhaps this explains the significant differences in half-life estimated for *CYH2* mRNA and mRNA74 (Herrick et al, 1990; Santiago et al, 1986; see Sections 1.2.3. & 3.3.2.1.).

Transcription inhibitors have also been used in the context of other organisms. For example, rifampicin is used extensively in studies of *E. coli* mRNA half-life. Similarly actinomycin D, an intercalating agent, is often used in studies of higher eukaryotes. Again, however, there will always be knock-on effects. Clear indications of such effects are the stabilisation of mammalian *c-fos* mRNA following actinomycin D treatment, and the stabilisation of *S. cerevisiae* ribosomal protein mRNAs after heat-shock (Shyu et al, 1989; Herruer et al, 1988; see Sections 1.2.2. & 4.1.3.).

Alternatives to the above methods involve the labelling of mRNAs *in vivo*. Labelled mRNA extracted at various time points is then used to probe blots of gene specific DNA fragments. Two different approaches can be used.

One approach is known as "decay kinetics". It has been used in studies of a range of organisms, including *S. cerevisiae*. Here, RNA is labelled *in vivo* to steady-state, following which the culture is transferred to fresh medium to initiated the chase. The half-life of any given mRNA is determined by following the decay with time of its activity. There are several disadvantages of this approach, however. First, growth of the culture is interrupted during transfer to fresh medium. Secondly, in the case of rare mRNAs, it can be difficult to obtain probes of sufficient activity. Thirdly, radiolabelled precursors are not instantaneously flushed from the cells. Finally, radio-labelled degradation products can be re-incorporated into newly synthesised mRNA. Mathematical corrections for re-incorporation can be used, but are complicated by the fact that different mRNAs are synthesised at different rates.

The second approach, known as "incorporation kinetics", suffers from less drawbacks. In particular cell growth is not interrupted. Here, radiolabelled precursors are added to exponentially-growing cells, and RNA is isolated at various time points thereafter. In the absence of mRNA decay, one would expect the activity of any specific mRNA to increase as a linear function of time. Thus, the departure from linearity serves as an

indicator of mRNA half-life (Greenberg, 1972). Even more so than in the case of the previous method, however, it can be difficult to obtain probes of sufficient activity. In all other respects this is the method of choice for estimation of mRNA half-life.

APPENDIX 5

Amino acid	Codon	Frequency	Amino acid	Codon	Frequency
Gly	GGG	0.0075	Thr	ACG	0.015
	GGA	0.005		ACA	0.055
	GGU	0.95		ACU	0.485
	GGC	0.0375		ACC	0.445
Glu	GAG	0.085	Trp	UGG	1.00
	GAA	0.915			
			Stop	UGA	0.08
Asp	GAU	0.42		UAG	0.053
	GAC	0.58		UAA	0.87
Val	GUG	0.0325	Cys	UGU	0.89
	GUA	0.01		UGC	0.11
	GUU	0.545			
	GUC	0.4125	Tyr	UAU	0.13
				UAC	0.87
Ala	GCG	0.01			
	GCA	0.03	Leu	UUG	0.75
	GCU	0.68		UUA	0.13
	GCC	0.2825		CUG	0.022
				CUA	0.07
Arg	AGG	0.0233		CUU	0.022
	AGA	0.8667		CUC	0.003
	CGG	0.00			
	CGA	0.00	Phe	UUU	0.21
	CGU	0.106		UUC	0.79
	CGC	0.0016			
			Gln	CAG	0.055
Ser	AGU	0.283		CAA	0.945
	AGC	0.0266			
	UCG	0.015	His	CAU	0.026
	UCA	0.038		CAC	0.74
	UCU	0.53			
	UCC	0.36	Pro	CCG	0.0075
				CCA	0.86
Lys	AAG	0.81		CCU	0.125
	AAA	0.19		CCC	0.0075
Asn	AAU	0.14	Ile	AUA	0.02
	AAC	0.86		AUU	0.45
				AUC	0.53
Met	AUG	1.00			

Codon bias table for highly expressed genes of *S. cerevisiae*. The values are modifications of those of Sharp et al (1986; Table 2). See Section 3.1.1. for explanation.

REFERENCES

Abovich, N., Gritz, L., Tung, L., & Rosbash, M. (1985). Effect of *RP51* gene dosage alterations on ribosome synthesis in *Saccharomyces cerevisiae*. Mol. Cell. Biol., **5**, 3429-3435.

Abovich, N., & Rosbash, M. (1984). Two genes for ribosomal protein 51 of *Saccharomyces cerevisiae* complement and contribute to the ribosomes. Mol. Cell. Biol., **4**, 1871-1879.

Aghib, D. F., Bishop, J. M., Ottolenghi, S., Guerrasio, A., Serra, A., & Saglio, G. (1990). A 3' truncation of *MYC* caused by chromosomal translocation in a human T-cell leukemia increases mRNA stability. Oncogene, **5**, 707-711.

Attardi, G., & Amaldi, F. (1970). Structure and synthesis of ribosomal RNA. Ann. Rev. Biochem., **39**, 183-226.

Auld, D. S., Atsuya, I., Campino, C., & Valenzula, P. (1976). Yeast RNA polymerase I: A eukaryotic zinc metalloenzyme. Biochem. Biophys. Res. Comm., **69**, 548-554.

Bach, M.-L., Lacroute, F., & Botstein, D. (1979). Evidence for transcriptional regulation of orotidine-5'-phosphate decarboxylase in yeast by hybridization of mRNA to the yeast structural gene cloned in *E. coli*. Proc. Natl. Acad. Sci. USA., **76**, 386-390.

Baronas-Lowell, D. M., & Warner, J. R. (1990). Ribosomal protein L30 is dispensable in the yeast *Saccharomyces cerevisiae*. Mol. Cell. Biol., **10**, 5235-5243.

Baughman, G., & Nomura, M. (1983). Localization of the target site for translational regulation of the L11 operon and direct evidence for translational coupling in *Escherichia coli*. Cell, **34**, 979-988.

Belasco, J. G., Beatty, J. T., Adams, C. W., von Gabain, A., & Cohen, S. N. (1985). Differential expression of photosynthesis genes in *R. capsulatus* results from segmental differences in stability within the polycistronic *rxcA* transcript. Cell, **40**, 171-181.

Belasco, J. G., & Higgins, C. F. (1988). Mechanisms of mRNA decay in bacteria: A prospective. Gene, **72**, 15-23.

Belhumeur, P., Paterno, G. D., Boiler, G., Claverie, J-M., & Skup, D. (1987). Isolation and characterisation of a murine cDNA clone highly homologous to the yeast L29 ribosomal protein gene. Nucl. Acid. Res., **15**, 1019-1029.

Bennetzen, J. L., & Hall, B. D. (1982). Codon selection in yeast. J. Biol. Chem., **257**, 3026-3031.

Bernstein, P., & Ross, J. (1989). Poly(A), poly(A) binding protein and the regulation of mRNA stability. TIBS., **14**, 373-377.

Bettany, A. J. B., Moore, P., Cafferkey, R., Bell, L. D., Goodey, A. R., Carter, B. L. A., & Brown, A. J. P. (1989). 5'-Secondary structure formation, in contrast to a short string of non-preferred codons, inhibits the translation of the pyruvate kinase mRNA in yeast. Yeast, **5**, 187-198.

Birnboim, H. C., & Doly, J. (1979). Rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res., **7**, 1513-1523.

Blobel, G. (1973). A protein of molecular weight 78,000 bound to the polyadenylated region of eukaryotic messenger RNAs. Proc. Natl. Acad. Sci. USA., **70**, 924-928.

Blume, J. E., & Shapiro, D. J. (1989). Ribosome loading, but not protein synthesis, is required for estrogen stabilization of *Xenopus laevis* vitellogenin mRNA. Nucl. Acid. Res., **17**, 9003-9014.

Bollen, G. H. P. M., Mager, W. H., & Planta, R. J. (1981). High resolution mini-two-dimensional gel electrophoresis of yeast ribosomal proteins. Mol. Biol. Rep., **8**, 37-44.

Bozzoni, I., Fragapane, P., Annesi, F., Pierandrei-Amaldi, P., Amaldi, F., & Beccari, E. (1984). Expression of two *Xenopus laevis* ribosomal protein genes in injected frog oocytes: A specific splicing block interferes with the L1 RNA maturation. J. Mol. Biol., **180**, 987-1005.

Brand, R. C., Klootwijk, J., Sibom, C. P., & Planta, R. J. (1979). Pseudouridylation of yeast ribosomal precursor RNA. Nucl. Acid. Res., **7**, 121-134.

Brewer, G., & Ross, J. (1988). Poly(A) shortening and degradation of the 3' A+U-rich sequences of human *c-myc* mRNA in a cell-free system. Mol. Cell. Biol., **8**, 1697-1708.

Brindle, P. K., Holland, J. P., Willett, C. E., Innis, M. A., & Holland, M. J. (1990). Multiple factors bind the upstream activation sites of the yeast enolase genes *ENO1* and *ENO2*: ABF1 protein, like repressor activator protein RAP1, binds *cis*-acting sequences which modulate repression or activation of transcription. Mol. Cell. Biol., **10**, 4872-4885.

Brock, M. L., & Shapiro, D. J. (1983). Estrogen stabilises vitellogenin mRNA against cytoplasmic degradation. Cell, **34**, 207-214.

Brow, D. A., & Geiduschek, E. P. (1987). Modulation of yeast 5S rRNA synthesis *in vitro* by ribosomal protein YL3. A possible regulatory loop. J. Biol. Chem., **262**, 13953-13958.

Brown, A. J. P. (1989). Messenger RNA stability in yeast. Yeast, **5**, 239-257.

Brown, A. J. P., Purvis, I. J., Santiago, T. C., Bettany, A. J. E., Loughlin, L., & Moore, J. (1988). Messenger RNA degradation in *Saccharomyces cerevisiae*. Gene, **72**, 151-160.

Burke, R. L., Tekamp-Olsan, P., & Najarian, R. (1983). The isolation, characterization, and sequence of the pyruvate kinase gene of *Saccharomyces cerevisiae*. J. Biol. Chem., **258**, 2193-2201.

Casey, J. L., Hentze, M. W., Koeller, D. M., Caughman, R. D., & Harford, J. B. (1988). Iron-responsive elements: Regulatory RNA sequences that control mRNA levels and translation. Science, **240**, 924-928.

Chang, T-H., Clark, M. W., Lustig, A. J., Cusick, M. E., & Abelson, J. (1988). RNA11 protein is associated with yeast spliceosome and is localized in the periphery of the cell nucleus. Mol. Cell. Biol., **8**, 2379-2393.

Chen, C. Y. A., Beatty, J. T., Cohen, S. N., & Belasco, J. G. (1988). An intercistronic stem-loop structure functions as an mRNA decay terminator necessary but insufficient for *puf* mRNA stability. Cell, **52**, 609-619.

Chen, C. Y., Oppermann, H., & Hitzeman, R. A. (1984). Homologous versus heterologous gene expression in the yeast *Saccharomyces cerevisiae*. Nucl. Acids. Res., **12**, 8951-8970.

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., & Rutter, W. J. (1979). Isolation of biologically active RNA from sources enriched in ribonuclease. Biochem., **24**, 5294-5299.

Cigan, A., & Donahue, T. (1987). Sequence and structural features associated with translational initiator regions in yeast - a review. Gene, **59**, 1-18.

Cleveland, D. W. (1987). The multitubulin hypothesis revisited: What have we learned? J. Cell. Biol., **104**, 381-383.

Cleveland, D. W., & Havercroft, J. C. (1983). Is apparent autoregulatory control of tubulin synthesis nontranscriptionally regulated? J. Cell.Biol., **97**, 919-924.

Cohen, R., Holland, J. P., Yokoi, T., & Holland, M. J. (1986). Identification of a regulatory region that mediates glucose-dependent induction of the *Saccharomyces cerevisiae* enolase gene *ENO2*. Mol. Cell. Biol., **6**, 2287-2297.

Cohen, R., Yokoi, T., Holland, J. P., Pepper, A. E., & Holland, M. J. (1987). Transcription of the constitutively expressed yeast enolase gene *ENO1* is mediated by positive and negative *cis*-acting regulatory sequences. Mol. Cell. Biol., **7**, 2753-2761.

Dabbs, E. R. (1986). Mutant studies on the prokaryotic ribosome. *In* B. Hardesty and G. Kramer (Ed.), Structure, function and genetics of ribosomes. (pp733-748). Springer-Verlag, New York.

Dabeva, M. D., Post-Beittenmiller, M. A., & Warner, J. R. (1986). Autogenous regulation of splicing of the transcript of a yeast ribosomal protein gene. Proc. Natl. Acad. Sci. USA., **83**, 5854-5857.

Dawid, I. B., & Wellauer, P. K. (1976). A reinvestigation of 5'-3' polarity in 40S ribosomal precursor of *Xenopus laevis*. Cell, **8**, 443-448.

Devereux, J., Haeberli, P., & Smithies, O. (1984). A comprehensive set of sequence analysis for the VAX. Nucl. Acid. Res., **12**, 387-395.

Donovan, D. M., & Pearson, N. J. (1986). Transcriptional regulation of ribosomal proteins during a nutritional upshift in *Saccharomyces cerevisiae*. Mol. Cell. Biol., **6**, 2429-2435.

Donovan, W. P., & Kushner, S. R. (1986). Polynucleotide phosphorylase and ribonuclease II are required for cell viability and mRNA turnover in *Escherichia coli* K12. Proc. Natl. Acad. Sci. USA., **83**, 120-124.

Draper, D. E. (1989). How do proteins recognize specific RNA sites? New clues from autogenously regulated ribosomal proteins. TIBS., **14**, 335-338.

El-Baradi, T.T.A.L., Raue, H. A., de Regt, V. C. H. F., & Planta, R. J. (1984). Stepwise dissociation of yeast 60S ribosomal subunits by LiCl and identification of L25 as primary 26S rRNA binding protein. Eur. J. Biochem., **144**, 393-400.

El-Baradi, T. T. A. L., de Regt, V. C. H. F., Einerhand, S. W. C., Teixido, J., Planta, R. J., Ballesta, J. P. G., & Raue, H. A. (1987). Ribosomal proteins EL11 from *Escherichia coli* and L15 from *Saccharomyces cerevisiae* bind to the same site in both yeast 26S and mouse 28S rRNA. J. Mol. Biol., **197**, 909-917.

El-Baradi, T. T. A. L., Raue, H. A., de Regt, V. C. H. F., Verbree, E. C., & Planta, R. J. (1985). Yeast ribosomal protein L25 binds to an evolutionary conserved site on yeast 26S and *E. coli* 23S rRNA. EMBO J., **4**, 2101-2107.

Fabian, G. R., & Hopper, A. K. (1987). *RRP1*, a *Saccharomyces cerevisiae* gene affecting rRNA processing and production of mature ribosomal subunits. J. Bact., **169**, 1571-1578.

Fahrner, K., Yarger, J., & Hereford, L. M. (1980). Yeast histone mRNA is polyadenylated. Nucl. Acid. Res., **8**, 5725-5737.

Fink, G. R. (1987). Pseudogenes in yeast? Cell, **49**, 5-6.

Finley, D. Bartel, B. Varshavsky, A. (1989). The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitin facilitates ribosome biogenesis. Nature, **338**, 394-401.

Fleming, G., Belhumeur, P., Skup, D., & Fried, H. M. (1989). Functional substitution of mouse ribosomal protein L27' for yeast ribosomal protein L29 in yeast ribosomes. Proc. Natl. Acad. Sci. USA., **86**, 217-221.

Fried, H. M., Nam, H. G., Loechel, S., & Teem, J. (1985). Characterization of yeast strains which conditionally expressed variants of ribosomal proteins *tcm1* and *cyh2*. Mol. Cell. Biol., **5**, 99-108.

Fried, H. M., & Warner, J. R. (1981). Cloning of yeast gene for trichodermin resistance and ribosomal protein L3. Proc Nat. Acad. Sci. USA., **78**, 238-242.

Fried, H. M., & Warner, J. R. (1982). Molecular cloning and analysis of yeast gene for cycloheximide resistance and ribosomal protein L29. Nucl. Acid. Res., **10**, 3133-3148.

Frischauf, A.-M., Lehrach, H., Poustka, A., & Murray, N. (1983). Lambda replacement vectors carrying polylinker sequences. J. Mol. Biol., **170**, 827-842.

Gallwitz, D., & Sures, I. (1980). Structure of a split yeast gene: Complete nucleotide sequence of the actin gene in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA., **77**, 2546-2550.

Gatermann, K. B., Teletski, C., Gross, T., & Kaufer, N. F. (1989). A ribosomal protein gene family from *Schizosaccharomyces pombe* consisting of three active members. Curr. Genet., **16**, 361-367.

Gay, D. A., Yen, T. J., Lau, J. T. Y., & Cleveland, d. W. (1987). Sequences that confer β-tubulin autoregulation through modulated mRNA stability reside within exon 1 of a β-tubulin mRNA. Cell, **50**, 671-679.

Geiduschek, E. P., & Tocchini-Valentini, G. P. (1988). Transcription by RNA polymerase III. Annu. Rev. Biochem., **57**, 873-914.

Georgiev, O. I., Nikolaev, N., Hadjiolov, A. A., Skryabin, K. G., Zakharyev, V. M., & Bayev. A. A. (1981). The structure of the yeast ribosomal RNA genes. Complete sequence of the 25S rRNA gene from *Saccharomyces cerevisiae*. Nucl. Acid. Res., **9**, 6953-6958.

Ginsburg, D., & Steitz, J. A. (1975). The 30S ribosomal precursor RNA from *Escherichia coli*. A primary transcript containing 23S, 16S and 5S sequences. J. Biol. Chem., **250**, 5647-5654.

Gorenstein, C., & Warner, J. R. (1976). Coordinate regulation of the synthesis of eukaryotic ribosomal proteins. Proc. Nat. Acad. Sci. USA., **73**, 1547-1551.

Gough, J. A., & Murray, N. E. (1983). Sequence diversity among related genes for recognition of specific targets in DNA molecules. J. Mol. Biol., **166**, 1-19.

Grant, P. G., Schindler, D., & Davis, J. E. (1976). Mapping of trichodermin resistance in *Saccharomyces cerevisiae*: a genetic locus for a component of the 60S ribosomal subunit. Genetics, **83**, 667-673.

Graves, R. A., Pandey, N. B., Chodchoy, N., & Marzluff, W. F. (1987). Translation is required for regulation of histone mRNA degradation. Cell, **48**, 615-626.

Greenberg, J. R. (1972). High stability of messenger RNA in growing cultured cells. Nature, **240**, 102-104.

Gribskov, M., Devereux, J., & Burgess, R. R. (1984). The codon preference plot: graphic analysis of protein coding sequences and prediction of gene expression. Nucl. Acid. Res., **12**, 539-549.

Gritz, L., Abovich, N., Teem, J. L., & Rosbash, M. (1985). Posttranscriptional regulation and assembly into ribosomes of a *Saccharomyces cerevisiae* ribosomal protein-β-galactosidase fusion. Mol. Cell. Biol., **5**, 3436-3442.

Guarente, L. (1987). Regulatory proteins in yeast. Ann. Rev. Genet., **21**, 425-452.

Hackett, P. B., & Sauerbier, W. (1975). The transcriptional organisation of the ribosomal RNA genes in mouse L cells. J. Mol. Biol., **91**, 235-256.

Hamil, K. G., Nam, H. G., & Fried, H. M. (1988). Constitutive transcription of yeast ribosomal protein gene *TCM1* is promoted by uncommon *cis*- and *trans*-acting elements. Mol. Cell. Biol., **8**, 4328-4341.

Hamilton, R., Watanabe, C., & de Boer, H. (1987). Compilation and comparison of the sequence context around the AUG start codons in *Saccharomyces cerevisiae* mRNAs. Nucl. Acids Res., **15**, 3581-3593.

Hardy, S. J. S. (1975). The stoicheiometry of ribosomal proteins of *Escherichia coli*. Mol. Gen. Genet., **140**, 253-274.

Hartwell, L. H., McLaughlin, C. S., & Warner, J. R. (1970). Identification of ten genes that control ribosome formation in yeast. Mol. Gen. Genet., **109**, 42-56.

Hentze, M. W., Caughman, S. W., Rouault, T. A., Barriocanal, J. G., Dancis, A., Harford, J. B., & Klausner, R. D. (1987). Identification of the iron-responsive element for the translational regulation of human ferritin mRNA. Science, **238**, 1570-1573.

Hereford, L. M., Osley, M. A., Ludwig II, J. R., & McLaughlin, C. S. (1981). Cell-cycle regulation of yeast histone mRNA. Cell, **24**, 367-375.

Herrick, D., & Jacobson, A. (1988). Use of chimeric mRNAs to study the structural determinants of mRNA stability in yeast. Yeast, **4**, S488.

Herrick, D., Parker, R., & Jacobson, A. (1990). Identification and comparison of stable and unstable mRNAs in *Saccharomyces cerevisiae*. Mol. Cell. Biol., **10**, 2269-2284.

Herruer, M. H., Mager, W. H., Raue, H. A., Vrekon, P., Wilms, E., & Planta, R. J. (1988). Mild temperature shock effects transcription of yeast ribosomal protein genes as well as the stability of their mRNAs. Nucl. Acids. Res., **16**, 7917-7929.

Herruer, M. H., Mager, W. H., Woudt, L. P., Nieuwint, R. T. M., Wassanaar, G. M., Groeneveld, P., & Planta, R. J. (1987). Transcriptional control of yeast ribosomal protein synthesis during carbon source upshift. Nucl. Acid. Res., **15**, 10133-10144.

Hinnesbusch, A. G. (1988). Mechanisms of gene regulation in the general control of amino acid biosynthesis in *Saccharomyces cerevisiae*. Microbiol. Rev., **52**, 248-273.

Hodgson, C. P., & Fisk, R. Z. (1987). Hybridization probe size control: Optimized "oligolabelling". Nucl. Acid. Res., **15**, 6295-6295.

Hoekema, A., Kastelein, R. A., Vasser, M., & DeBoer, H. A. (1987). Codon replacement in the *PGK1* gene of *Saccharomyces cerevisiae*: experimental approach to study the role of biased codon usage in gene expression. Mol. Cell. Biol., **7**, 2914-2924.

Holland, M. J., Holland, J. P., Thill, G. P., & Jackson, K. A. (1981). The primary structure of two yeast enolase genes: Homology between the 5' noncoding flanking regions of yeast enolase and glyceraldehyde-3-phosphate dehydrogenase genes. J. Biol. Chem., **256**, 1385-1395.

Holland, M. J., Yokoi, T., Holland, J. P., Myarnbo, K., & Innis, M. A. (1987). The *GCR1* gene encodes a positive transcriptional regulator of the enolase and glyceraldehyde-3-phosphate dehydrogenase gene families in *Saccharomyces cerevisiae*. Mol. Cell. Biol., **7**, 813-820.

Holmes, D. S., & Quigley, M. (1981). A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem., **114**, 193-197.

Hu, W., & Messing, J. (1982). The making of strand-specific M13 probes. Gene, **17**, 271-277.

Huber, P. W., & Wool, I. G. (1986a). Use of cytotoxic nuclease α-sarcin to identify the binding site on eukaryotic 5S ribosomal ribonucleic acid for ribosomal protein L5. J. Biol. Chem., **261**, 3002-3005.

Huber, P. W., & Wool, I. G. (1986b). Identification of the binding site on 5S rRNA for the transcription factor IIIA: Proposed structure of a common binding site on 5S rRNA on the gene. Proc. Natl. Acad. Sci. USA., **83**, 1593-1597.

Huet, J., Cottrelle, P., Cool, M., Vignais, M.-L., Thiele, D., Marck, C., Buhler, J. M., Sentenac, A., & Fromageot, P. (1985). A general upstream binding factor for genes of the yeast translational apparatus. EMBO J., **4**, 3539-3547.

Huet, J., & Sentenac, A. (1987). TUF, the yeast DNA-binding factor specific for UAS_{rpg} upstream activating sequences: Identification of the protein and its DNA-binding domain. Proc. Natl. Acad. Sci. USA., **84**, 3648-3652.

Ikemura, T. (1981). Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. J. Mol. Biol., **151**, 389-409.

Ikemura, T. (1982). Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. J. Mol. Biol., **158**, 573-597.

Innis, M. A., Myambo, K. B., Gelfand, D. H., & Brow, M. A. D. (1988). DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction amplified DNA. Proc. Natl. Acad. Sci. USA., **85**, 9436-9440.

Itoh, T., Otaka, E., & Matsui, K. A. (1985). Primary structures of ribosomal protein YS25 from *Saccharomyces cerevisiae* and its counterparts from *Schizosaccharomyces pombe* and rat liver. Biochem., **24**, 7418-7423.

Jackson, S. P., Lossky, M., & Beggs, J. D. (1988). Cloning of the *RNA8* gene of *Saccharomyces cerevisiae*, detection of the RNA8 protein, and demonstration that it is essential for nuclear pre-mRNA splicing. Mol. Cell. Biol., **8**, 1067-1075.

Jemtland, R., Maehlum, E., Gabrielsen, O. S., & Oyen, T. B. (1986). Regular distribution of length heterogeneities within non-transcribed spacer regions of cloned and genomic rDNA of *Saccharomyces cerevisiae*. Nucl. Acid. Res., **14**, 5145-5158.

Johnston, G. C., & Singer, R. A. (1978). RNA synthesis and control of cell division in the yeast *S. cerevisiae*. Cell, **14**, 951-958.

Jones, T. R., & Cole, M. D. (1987). Rapid cytoplasmic turnover of *c-myc* mRNA: Requirement of the 3' untranslated sequences. Mol. Cell. Biol., **7**, 4513-4521.

Kaufer, N. F., Fried, H. M., Schwindinger, W. F., Jasin, M., & Warner, J. R. (1983). Cycloheximide resistance in yeast: the gene and its protein. Nucl. Acid. Res., **11**, 3123-3135.

Kearney, K., & Nomura, M. (1987). Secondary structure of the autoregulatory mRNA binding site of ribosomal protein L1. Mol. Gen. Genet., **210**, 60-68.

Keif, D. R., & Warner, J. R. (1981). Coordinate control of synthesis of ribosomal ribonucleic acid and ribosomal proteins during nutritional shift-up in *Saccharomyces cerevisiae*. Mol. Cell. Biol., **1**, 1007-1015.

Kempers-Veenstra, A. E., Oliemans, J., Offenberg, H., Dekker, A. F., Piper, P. W., Planta, R. J., & Klootwijk, J. (1986). 3'-End formation of transcripts from the yeast rRNA operon. EMBO J., **5**, 2703-2710.

Kim, C. H., & Warner, J. R. (1983a). Mild temperature shock alters the transcription of a discrete class of *Saccharomyces cerevisiae* genes. Mol. Cell. Biol, **3**, 457-465.

Kim, C. H., & Warner, J. R. (1983b). The mRNA for ribosomal proteins in yeast. J.Mol. Biol., **165**, 79-89.

Kjeldgaard, N. Q., & Gausing, K. (1974). Regulation of biosynthesis of ribosomes. *In* M. Nomura, A. Tissieres, & P. Lengyel (Ed.), Ribosomes. (pp. 369-392). Cold Spring Harbour.

Klekamp, M. S., & Weil, P. A. (1982). Specific transcription of homologous class III genes in yeast-soluble cell-free extracts. J. Biol. Chem., **257**, 8432-8441.

Klootwijk, J., & Planta, R. J. (1973). Modified sequences in yeast ribosomal RNA. Mol. Biol. Rept., **1**, 187-191.

Klug, G., Adams, C. W., Belasco, J., Doerge, B., & Cohen, S. N. (1987). Biological consequences of segmental alterations in mRNA stability: Effects of deletion of the intercistronic hairpin loop region of the *Rhodobacter capsulatus puf* operon. EMBO J., **6**, 3515-3520.

Koch, H., & Friesen, J., D. (1979). Individual messenger RNA half lives in *Saccharomyces cerevisiae*. Mol. Gen. Genet., **170**, 129-135.

Koeller, D. M., Casey, J. L., Hentze, M. W., Gerhardt, E. M., Chan, L.-N. L., Klausner, R. D., & Harford, J. B. (1989). A cytosolic protein bind to structural elements within the iron regulatory region of the transferrin receptor mRNA. Proc. Natl. Acad. Sci. USA., **86**, 3574-3578.

Kozak, M. (1984). Compilation and analysis of sequences upstream from a translational start site in eukaryotic mRNAs. Nucl. Acid. Res., **12**, 857-872.

Krainer, A. R., & Maniatis, T. (1988). RNA splicing. *In* B. D. Hames, & D. M. Glover (Ed.), Transcription and Splicing. (pp. 131-206). Oxford, Washington DC: IRL Press.

Kramer, R. A., Philippsen, P., & Davis, R. D. (1978). Divergent transcription in the yeast ribosomal RNA coding region as shown by hybridization to separated strands and sequence analysis of cloned DNA. J. Mol. Biol., **123**, 405-416.

Kreader, C. A., & Heckman, J. E. (1987). Isolation and characterization of a *Neurospora crassa* ribosomal protein gene homologous to *CYH2* of yeast. Nucl. Acid. Res., **15**, 9027-9042.

Kruiswijk, T., Planta, R. J., & Krop, J. M. (1978). The course of assembly of ribosomal subunits in yeast. Biochim. Biophys. Acta., **517**, 378-389.

Kuo, S. C., Cano, F. R., & Lampden, J. O. (1973). Lomofungin, an inhibitor of RNA synthesis in yeast protoplasts: its effect on enzyme formation. Antimicrobial Agents and Chemotherapy, **3**, 716-722.

Langford, C. J., Klinz, F.-J., Donath, C., & Gallwitz, D. (1984). Point mutations identify the conserved, intron-contained TACTAAC box as an essential splicing signal sequence in yeast. Cell, **36**, 645-653.

Larkin, J. C., Thompson, J. R., & Woolford Jr, J. L. (1987). Structure and expression of the *Saccharomyces cerevisiae CRY1* gene: A highly conserved ribosomal protein gene. Mol. Cell. Biol., **7**, 1764-1775.

Last, R. L., Stavenhagen, J. B., & Woolford Jr, J. L. (1984). Isolation and characterization of the *RNA2*, *RNA3*, and *RNA11* genes of *Saccharomyces cerevisiae*. Mol. Cell. Biol., **4**, 2396-2405.

Last, R. L., & Woolford Jr, J. L. (1986). Identification and nuclear localization of yeast pre-messenger RNA processing components: *RNA2* and *RNA3* proteins. J. Cell. Biol., **103**, 2103-2112.

Lattke, H., & Weser, U. (1976). Yeast RNA-polymerase B: A zinc protein. FEBS Lett., **65**, 288-292.

Lauer, G. D., Roberts, T. M., & Klotz, L. C. (1977). Determination of the nuclear DNA content of *Saccharomyces cerevisiae* and implications for the organization of DNA in yeast chromosomes. J. Mol. Biol., **114**, 507-526.

Lautenberger, J. A., & Chen, Z-Q. (1987). A method for the preparation of high molecular weight yeast DNA. Gene. Anal. Techn., **4**, 87-88.

Lee, M. G., Lane, D. P., & Beggs, J. D. (1986). Identification of the RNA2 protein of *Saccharomyces cerevisiae*. Yeast, **2**, 59-67.

Lee, M. G., Young, R. A., & Beggs, J. D. (1984). Cloning of the *RNA2* gene of *Saccharomyces cerevisiae*. EMBO J., **3**, 2825-2830.

Leer, R. J., van Raamsdonk-Duin, C., Schoppink, P. J., Cornelissen, M. T. E., & Cohen, L. H. (1983). Yeast ribosomal protein S33 is encoded by an unsplit gene. Nucl. Acid. Res., **11**, 7759-7768.

Leer, R. J., van Raamsdonk-Duin, M. M. C., Hagendoorn, J. M., Mager, W. H., & Planta, R. J. (1984). Structural comparison of yeast ribosomal protein genes. Nucl. Acid. Res., **12**, 6685-6700.

Leer, R. J., van Raamsdonk-Duin, M. M. C., Mager, W. H., & Planta, R. J. (1985). Conserved sequences upstream of yeast ribosomal protein genes. Curr. Genet., **9**, 273-277.

Leibold, D., & Munro, H. W. (1988). Cytoplasmic protein binds *in vitro* to a highly conserved sequence in the 5' untranslated region of ferritin heavy- and light-subunit mRNAs. Proc. Natl. Acad. Sci. USA., **85**, 2171-2175.

Liau, M. C., & Hurlbert, R. B. (1975). The topographical order of 18S and 20S ribosomal ribonucleic acids within the 45S precursor molecule. J. Mol. Biol., **98**, 321-332.

Lin, A., McNally, J., & Wool, I. G. (1983). The primary structure of rat liver ribosomal protein L37. Homology with yeast and bacterial ribosomal proteins. J. Biol. Chem., **258**, 10664-10671.

Lin, R-J., Lustig, A. J., & Abelson, J. (1987). Splicing of yeast nuclear pre-mRNA *in vitro* requires functional 40S spliceosomes and several extrinsic factors. Genes Dev., **1**, 7-18.

Lindquist, S. (1981). Regulation of protein synthesis during heat shock. Nature, **293**, 311-314.

Lindquist, S. (1984). Heat shock - a comparison of *Drosophila* and yeast. J. Embryology and Experimental Morphology, **83** (supp), 147-161.

Lindquist, S. (1986). The heat shock response. Ann. Rev. Biochem., **55**, 1151-1191.

Littlefield, J. W., Keller, E. B., Gros, J., & Zamecnik, P. C. (1955). Studies on cytoplasmic ribonucleoprotein particles from the liver of rat. J. Biol. Chem., **217**, 111-123.

Losson, R., Fuchs, R. P. P., & Lacroute, F. (1983). *In vivo* transcription of a eukaryotic regulatory gene. EMBO J., **2**, 2179-2184.

Losson, R., & Lacroute, F. (1979). Interference of nonsense mutations with eukaryotic messenger RNA stability. Proc. Natl. Acad. Sci. USA., **76**, 5134-5137.

Luria, S. E., Delbruck, M., & Anderson, T.F. (1943). Electron microscope studies of bacterial viruses. J. Bacteriol., **46**, 57-67.

Lycan, D. E., Osley, M. A., & Hereford, L. M. (1987). Role of transcriptional and posttranscriptional regulation in expression of histone genes in *Saccharomyces cerevisiae*. Mol. Cell. Biol., **7**, 614-621.

Mager, W. H. (1988). Control of ribosomal protein gene expression. Biochim. Biophys. Acta., **949**, 1-15.

Maicas, E., Pluthero, F. G., & Friesen, J., D,. (1988). The accumulation of three yeast ribosomal proteins under conditions of excess mRNA is determined primarily by fast protein decay. Mol. Cell. Biol., **8**, 169-175.

Mankin, A. S., Skryabin, K. G., & Rubstov, P. M. (1986). Identification of ten additional nucleotides in the primary structure of yeast 18S rRNA. Gene, **44**, 143-145.

Marks, P., & Rifkind, P. (1978). Erythroleukemia differentiation. Ann. Rev. Biochem., **47**, 419-448.

Maxam, A. M., Tizard, R., Skryabin, K. G., & Gilbert, W. (1977). Promoter region for yeast 5S ribosomal RNA genes in the yeast *Saccharomyces cerevisiae*. Nature., **267**, 643-645.

McAlistair, L., & Holland, M. J. (1982). Targeted deletion of a yeast enolase structural gene: Identification and isolation of yeast enolase isozyme. J. Biol. Chem., **257**, 7181-7188.

McCarthy, J. E. G., & Gualerzi, C. (1990). Translational control of prokaryotic gene expression. TIG., **6**, 78-85.

McLaughlin, C. S. (1974). Yeast ribosomes: Genetics. *In* M. Nomura, A. Tissieres, & P. Lengyel (Ed.), Ribosomes. (pp. 815-827). New York: Cold Spring Harbour Press.

McQuillen, K., Roberts, R. B., & Britten, R. J. (1959). Synthesis of nascent protein by ribosomes in *Escherichia coli*. Proc. Natl. Acad. Sci. USA., **45**, 1437-1447.

Meijlink, F. T., Curran, T., Miller, A. D., & Verma, I. M. (1985). Removal of a 67-base-pair sequence in the noncoding region of proto-oncogene *fos* converts it to a transforming gene. Proc. Natl. Acad. Sci. USA., **82**, 4987-4991.

Mellor, J., Dobson, M. J., Kingsman, A. J., & Kingsman, S. M. (1987). A trascriptional activator is located in the coding region of the yeast *PGK* gene. Nucl. Acid. Res., **15**, 6243-6259.

Messing, J. (1979). A multipurpose cloning system based on single-stranded DNA bacteriophage M13. Recomb. DNA Tech. Bull., **2**, 43.

Messing, J. (1983). New M13 vectors for cloning. Methods Enzymol., **101**, 20-78.

Michel, S., Traut, R. R., & Lee, J. C. (1983). Yeast ribosomal proteins: electrophoretic analysis in four two-dimensional gel systems - correlation of nomenclature. Mol. Gen. Genet., **191**, 251-256.

Miles, D. J., Donovan, D. M., & Pearson, N. J. (1988). Construction and characterization of a haploid strain of *Saccharomyces cerevisiae* that completely lacks all genomic *CYH2* sequences. Curr. Genet., **14**, 325-329.

Miller, A. B., MacKay, V. L., & Nasmyth, K. A. (1985). Identification and comparison of two sequence elements that confer cell-type specific transcription in yeast. Nature, **314**, 598-603.

Miller, A. M. (1984). The yeast *MAT*α*1* gene contains two intros. EMBO J., **3**, 1061-1065.

Miller, M. J., Xuong, N.-H., & Geiduschek, E. P. (1982). Quantitative analysis of the heat shock response of *Saccharomyces cerevisiae*. J. Bacteriol., **151**, 311-327.

Mitsui, K., & Tsurugi, K. (1988a). cDNA and deduced amino acid sequence of 38 kDA-type acidic ribosomal protein A0 from *Saccharomyces cerevisiae*. Nucl. Acid. Res., **16**, 3573.

Mitsui, K., & Tsurugi, K. (1988b). cDNA and deduced amino acid sequence of acidic ribosomal protein A1 from *Saccharomyces cerevisiae*. Nucl. Acid. Res., **16**, 3574.

Mitsui, K., & Tsurugi, K. (1988c). cDNA and deduced amino acid sequence of acidic ribosomal protein A2 from *Saccharomyces cerevisiae*. Nucl. Acid. Res., **16**, 3575.

Molenaar, C. M. T., Woudt, L. P., Jansen, A. E. M., Mager, W. H., Planta, R. J., Donovan, D. M., & Pearson, N. J. (1984). Structure and organisation of two linked ribosomal protein genes in yeast. Nucl. Acid. Res., **12**, 7345-7359.

Moreland, R. B., Nam, H. G., Hereford, L. M., & Fried, H. M. (1985). Identification of a nuclear localization signal of a yeast ribosomal protein. Proc. Natl. Acad. Sci. USA., 82, 6561-6565.

Mortimer, R. K., & Hawthorne, D. C. (1966). Genetic mapping in *Saccharomyces*. Genetics, **53**, 165-173.

Mount, S. M. (1982). A catalogue of splice junction sequences. Nucl. Acid. Res., **10**, 459-472.

Mullner, E. W., & Kuhn, L. C. (1988). A stem-loop in the 3' untranslated region mediates iron-dependent regulation of transferrin receptor mRNA stability in the cytoplasm. Cell, **53**, 815-825.

Mullner, E. W., Neuport, B., & Kuhn, L. C. (1989). A specific mRNA binding factor regulates the iron-dependent stability of cytoplasmic transferrin receptor mRNA. Cell, **58**, 373-382.

Myers, R. M., Tilly, K., & Maniatis, T. (1986). Fine structure genetic analysis of a β-globin promoter. Science, **227**, 613-618.

Needleman, S. B., & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol., **48**, 443-453.

Newbury, S. F., Smith, N. H., & Higgins, C. F. (1987b). Differential mRNA stability controls relative gene expression within a polycistronic operon. Cell, **51**, 1131-1143.

Newbury, S. F., Smith, N. H., Robinson, E. C., Hiles, I. D., & Higgins, C. F. (1987a). Stabilization of translationally active mRNA by prokaryotic REP sequences. Cell, **48**, 297-310.

Nischt, R., Gross, T., Gatermann, K., Swida, U., & Kaufer, N. F. (1987). Sequence and regulatory responses of a ribosomal protein gene from the fission yeast. *Schizosaccharomyces pombe*. Nucl. Acids Res., **15**, 1477-1482.

Nischt, R., Thuroff, E., & Kaufer, N. F. (1986). Molecular cloning of a ribosomal protein gene from the fission yeast *Schizosaccharomyces pombe*. Curr. Genet., **10**, 365-370.

Nomura, M., Gourse, R., & Baughman, G. (1984). Regulation of the synthesis of ribosomes and ribosomal components. Annu. Rev. Biochem., **53**, 75-117.

Nomura, M., Morgan, E. A., & Jaskunas, S. R. (1977). Genetics of bacterial ribosomes. Ann. Rev. Genet., **11**, 297-347.

Nomura, M., Yates, J. L., Dean, D., & Post, L. E. (1980). Feedback regulation of ribosomal protein gene expression in *Escherichia coli*: Structural homology of ribosomal RNA and ribosomal protein mRNA. Proc. Natl. Acad. Sci. USA., **77**, 7084-7088.

Nonet, M., Scafe, C., Sexton, J., & Young, R. (1987). Eukaryotic RNA polymerase conditional mutant that rapidly ceases mRNA synthesis. Mol. Cell. Biol., **7**, 1602-1611.

Nygard, O., & Nilsson, L. (1990). Translational dynamics Interactions between the translational factors, tRNA and ribosomes during eukaryotic protein synthesis. Eur. J. Biochem., **191**, 1-17.

Osley, M. A., & Hereford, L. M. (1981). Yeast histone genes show dosage compensation. Cell, **24**, 377-384.

Otaka, E., Higo, K.-i., & Itoh, T. (1984). Yeast ribosomal proteins VIII. Isolation of two proteins and sequence characterization of twenty-four proteins from cytoplasmic ribosomes. Mol. Gen. Genet., **195**, 544-546.

Owen, D., & Kuhn, L. C. (1987). Noncoding 3' sequences of the transferrin receptor gene are required for mRNA regulation by iron. EMBO J., **6**, 1287-1293.

Pachter, J. S., Yen, T. J., & Cleveland, D. W. (1987). Autoregulation of tubulin expression is achieved through specific degradation of polysomal mRNAs. Cell, **51**, 283-292.

Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S., & Sharp, P. A. (1986). Splicing of messenger RNA precursors. Annu. Rev. Biochem., **55**, 1119-1150.

Palmiter, R. D., & Carey, N. H. (1974). Rapid inactivation of ovalbumin messenger ribonucleic acid after acute withdrawal of estrogen. Proc. Natl. Acad. Sci. USA., **71**, 351-2361.

Pandey, N. B., & Marzluff, W. F. (1987). The stem-loop structure at the 3' end of histone mRNA is necessary and sufficient for regulation of histone mRNA stability. Mol. Cell. Biol., **7**, 4557-4559.

Pardue, M. L., Brown, D. D., & Birnsteil, M. L. (1973). Location of the genes for 5S ribosomal RNA in *Xenopus laevis*. Chromosoma, **42**, 191-203.

Parker, R., & Jacobson, A. (1990). Translation and a 42-nucleotide segment within the coding region of the mRNA encoded by the *MATα1* gene are involved in promoting rapid mRNA decay in yeast. Proc. Natl. Acad. Sci. USA., **87**, 2780-2784.

Pearson, N. J., Howard, M. F., & Warner, J. R. (1982). Yeast use translational control to compensate for extra copies of a ribosomal protein gene. Cell, **29**, 347-355.

Pearson, N. T., & Haber, J. E. (1980). Changes in regulation of ribosomal protein synthesis during vegetative growth and sporulation of *Saccharomyces cerevisiae* . J. Bacteriol., **143**, 1411-1419.

Pelsy, F., & Lacroute, F. (1984). Effect of ochre nonsense mutations on yeast *URA1* mRNA stability . Curr. Genet., **8**, 277-282.

Peltz, S. W., Brewer, G., Kobs, G., & Ross, J. (1987). Substrate specificity of the exonuclease activity that degrades H4 histone mRNA. J. Biol. Chem., **262**, 9382-9388.

Peltz, S. W., & Ross, J. (1987). Autogenous regulation of histone mRNA decay by histone proteins in a cell-free system. Mol. Cell. Biol., **7**, 4345-4356.

Petes, T. D. (1979). Meiotic mapping of yeast ribosomal deoxyribonucleic acid on chromosome XII. J. Bac., **138**, 185-192.

Phillippsen, P., Kramer, R. A., & Davis, R. W. (1978a). Cloning of the yeast ribosomal DNA repeat units in *Sst*I and *Hin*dIII lambda vectors using genetic and physical size selections. J. Mol. Biol., **123**, 371-386.

Phillippsen, P., Thomas, M. T., Kramer, R. A., & Davis, R. W. (1978b). Unique arrangement of coding sequences for 5S, 5.8S, 18S and 25S ribosomal RNA in *Saccharomyces cerevisiae* as determined by R-loop and hybridization analysis. J. Mol. Biol., **123**, 387-404.

Piper, P. N., Curran, B., Davies, M. W., Lockheart, A., & Reid, G. (1986). Transcription of the phosphoglycerate kinase gene of *Saccharomyces cerevisiae* increases when fermentative cultures are stressed by heat-shock. Eur. J. Biochem., **161**, 525-531.

Piper, P. W., Bellatin, J. A., & Lockheart, A. (1983). Altered maturation of sequences at the 3' terminus of 5S gene transcripts in a *Saccharomyces cerevisiae* mutant that lacks a RNA processing endonuclease. EMBO J., **2**, 353-359.

Piper, P. W., Curran, B., Davies, M. W., Hirst, K., Lockheart, A., Ogden, J. E., Stanway, C. A., Kingsman, A. J., & Kingsman, S. M. (1988). A heat shock element in the phosphoglycerate kinase gene promoter of yeast. Nucl. Acid. Res., **16**, 1333-1348.

Piper, P. W., Curran, B., Davies, W., Hirst, K., & Seward, K. (1987). *Saccharomyces cerevisiae* mRNA populations of different intrinsic stability in unstressed and heat shock cells display almost m^7GpppA : m^7GpppG5'-cap structure ratios. FEBS Lett., **220**, 177-180.

Pittenger, M. F., & Cleveland, D. W. (1985). Retention of autoregulatory control of tubulin synthesis in cyoplasts: demonstration of cytoplasmic mechanism that regulates the level of tubulin expression. J. Cell. Biol., **101**, 1941-1952.

Planta, R. J., Mager, W. H., Leer, R. J., Woudt, L. P., Raue, H. A., & El-Baradi, T. T. A. L. (1986). Structure and expression of ribosomal protein genes in yeast. *In* B. Hardesty and G. Kramer (ed.), Structure, function and genetics of ribosomes. (pp 699-718). Springer-Verlag, New York.

Purvis, I. J., Loughlin, L., Bettany, A. J. E., & Brown, A. J. P. (1987). Translation and stability of an *Escherichia coli* β-galactosidase mRNA expressed under the control of pyruvate kinase sequences in *Saccharomyces cerevisiae*. Nucl. Acid. Res., **15**, 7963-7974.

Rabbitts, P. H., Forster, A., Stinson, M. A., & Rabbitts, T. H. (1985). Truncation of exon 1 from the *c-myc* gene results in prolonged *c-myc* mRNA stability. EMBO J., **4**, 3727-3733.

Remacha, M., Saenz-Robles, M. T., Vilella, M. D., & Ballesta, J. P. G. (1988). Independent genes coding for three acidic proteins of the large ribosomal subunit from *Saccharomyces cerevisiae*. J. Biol. Chem., **263**, 9094-9101.

Rigby, P. W. J., Dieckmann, M., Rhodes, C., & Berg, P. (1977). Labelling DNA to high specific activity *in-vitro* by nick translation with DNA polymerase I. J. Mol. Biol., **113**, 237-257.

Ritossa, F. (1962). A new puffing pattern induced by temperature shock and DNP in *Drosophila*. Experentia., **18**, 571-573.

Rosbash, M., Harris, P. K. W., Woolford Jr, J. L., & Teem, J. L. (1981). The effect of temperature-sensitive RNA mutations on the transcription products from cloned ribosomal protein genes of yeast. Cell, **24**, 679-686.

Ross, J., & Kobs, G. (1986). H4 histone messenger RNA decay in cell-free extracts initiates at or near the 3' terminus and proceeds 3' to 5'. J. Mol. Biol., **188**, 579-593.

Ross, J., Kobs, G., Brewer, G., & Peltz, S. W. (1987). Properties of the exonuclease activity which degrade H4 histone mRNA. J. Biol. Chem., **262**, 9374-9381.

Ross, J., Peltz, S. W., Kobs, G., & Brewer, G. (1986). Histone mRNA degradation *in vivo*: The first detectable step occurs at or near the 3' terminus. Mol. Cell. Biol., **6**, 4362-4371.

Rotenberg, M. O., Moritz, M., & Woolford Jr, J. L. (1988). Depletion of *Saccharomyces cerevisiae* ribosomal protein L16 causes a decrease in 60S ribosomal subunits and formation of half-mer polyribosomes. Genes Dev., **2**, 160-172.

Rotenberg, M. O., & Woolford, J. L. (1986). Tripartite upstream promoter element essential for expression of *Saccharomyces cerevisiae* ribosomal protein genes. Mol. Cell. Biol., **6**, 674-687.

Rouault, T. A., Hentze, M. W., Caughman, S. W., Harford, J. B., & Klausner, R. D. (1988). Binding of a cytosolic protein to the iron-responsive element of human ferritin messenger RNA. Science, **241**, 1207-1210.

Rubin, G. M. (1973). The nucleotide sequence of *Saccharomyces cerevisiae* 5.8S ribosomal ribonucleic acid. J. Biol. Chem., **248**, 3860-3875.

Rubstov, M. M., Musakhanov, M. M., Zakharyev, V. M., Krayev, A. S., Skryabin, K. G., & Bayev, A. A. (1980). The structure of the yeast ribosomal RNA genes. I. The complete nucleotide sequence of the 18S ribosomal RNA gene from from *Saccharomyces cerevisiae*. Nucl. Acid. Res., **8**, 5779-5794.

Russell, P., & Nurse, P. (1986). *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*: A look at yeasts divided. Cell, **45**, 781-782.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). Cold Spring Harbour Molecular Cloning: A laboratory manual (2nd Edition).

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Nat. Acad. Sci. USA., **74**, 5463-5467.

Santangelo, G. M., & Tornow, J. (1990). Efficient transcription of the glycolytic gene *ADH1* and three translational component genes requires the *GCR1* product, which can act through TUF/GRF/RAP binding sites. Mol. Cell. Biol., **10**, 859-862.

Santiago, T. C. (1986). Investigation of messenger RNA stability in the yeast *Saccharomyces cerevisiae*. Ph.D Thesis, University of Glasgow.

Santiago, T. C., Bettany, A. J. E., Purvis, I. J., & Brown, A. J. P. (1987). Messenger RNA stability in *Saccharomyces cerevisiae*: the influence of translation and poly(A) tail length. Nucl. Acid. Res., **15**, 2417-2429.

Santiago, T. C., Purvis, I. J., Bettany, A. J. E., & Brown, A. J. P. (1986). The relationship between mRNA stability and length in *Saccharomyces cerevisiae*. Nucl. Acid. Res., **14**, 8347-8360.

Schwindinger, W. F., & Warner, J. R. (1987). Transcriptional elements of the yeast ribosomal protein gene *CYH2*. J. Biol. Chem., **262**, 5690-5695.

Scripati, C. E., Groner, Y., & Warner, J. R. (1976). Methylated, blocked 5' termini of yeast mRNA. J. Biol. Chem., **251**, 2898-2904.

Shapiro, D. J., Blume, J. E., & Nielson, D. A. (1987). Regulation of messenger RNA stability in eukaryotic cells. BioEssays, **6**, 221-226.

Sharp, P. M., Tuohy, T. M. F., & Mosurski, K. R. (1986). Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. Nucl. Acid. Res., **14**, 5125-5143.

Shaw, G., & Kamen, R. (1986). A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. Cell, **46**, 659-667.

Shultz, L. D., & Friesen, J. D. (1983). Nucleotide sequence of the *tcm1* gene (ribosomal protein L3) of *Saccharomyces cerevisiae*. J. Bacteriol., **155**, 8-14.

Shyu, A.-B., Greenberg, M. E., & Belasco, J. G. (1989). The *c-fos* transcript is targeted for rapid decay by two distinct mRNA degradation pathways. Genes. Dev., **3**, 60-72.

Singleton, C. K. (1989). Nucleotide sequence of V1, a ribosomal protein gene from *Dictyostelium discoideum*. Nucl. Acid. Res., **17**, 7989.

Singleton, C. K., Manning, S. S., & Ken, R. (1989). Primary structure and regulation of vegetative specific genes of *Dictyostelium discoideum*. Nucl. Acid. Res., **17**, 9679-9692.

Sittman, D. B., Graves, R. A., & Marzluff, W. F. (1983). Histone mRNA concentrations are regulated at the level of transcription and mRNA degradation. Proc. Natl. Acad. Sci. USA., **80**, 1849-1853.

Skryabin, K. G., Eldarov, M. A., Larionov, V. L., Bayev, A. A., Klootwijk, J., de Regt, V. C. H. F., Veldman, G. M., Planta, R. J., Georgiev, O. I., & Hadjiolov, A. A. (1984). Structure and function of the nontranscribed spacer regions of yeast rDNA. Nucl. Acid. Res., **12**, 2955-2968.

Sollner-Webb, B., & Tower, J. (1986). Transcription of cloned eukaryotic ribosomal RNA genes. Annu. Rev. Biochem., **55**, 801-830.

Soltyk, A., Tropak, M., & Friesen, J. D. (1984). Isolation and characterization of the *RNA2*$^+$, *RNA4*$^+$, and *RNA11*$^+$ genes of *Saccharomyces cerevisiae*. J. Bacteriol., **160**, 1093-1100.

Sor, F., & Nomura, M. (1987). Cloning and DNA sequence determination of the L11 ribosomal protein operon of *Serratia marcescens* and *Proteus vulgaris*: Translational feedback regulation of the *Escherichia coli* L11 operon by heterologous L1 proteins. Mol. Gen. Genet., **210**, 52-59.

Stern, M. J., Ames, G. F.-L., Smith, N. H., Robinson, E. C., & Higgins, C. F. (1984). Repetitive extragenic palindromic sequences: A major component of the bacterial genome. Cell, **37**, 1015-1026.

Stocklein, W., Pipersberg, W., & Bock, A. (1981). Amino acid replacements in ribosomal protein YL24 of *Saccharomyces cerevisiae* causing resistance to cycloheximide. FEBS Lett., **136**, 265-268.

Swanson, M. E., Yip, M., & Holland, M. J. (1985). Characterization of an RNA polymerase I-dependent promoter within the spacer region of yeast ribosomal cistrons. J. Biol. Chem., **260**, 9905-9915.

Tabor, S., & Richardson, C. C. (1987). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Nat. Acad. Sci. USA., **84**, 4767-4771.

Teem, J. L., Abovich, N., Kaufer, N. F., Schwindinger, W. F., Warner, J. R., Levy, A., Woolford, J., Leer, R. J. v. R.-D., M.M.C., Mager, W. H., Planta, R. J., Schultz, L., Friesen, J. D., Fried, H., & Rosbash, M. (1984). A comparison of yeast ribosomal protein gene DNA sequences. Nucl. Acid. Res., **12**, 8295-8312.

Teem, J. L., & Rosbash, M. (1983). Expression of a β-galactosidase gene containing the ribosomal protein 51 intron is sensitive to the *rna2* mutation of yeast. Proc. Natl. Acad. Sci. USA., **80**, 4403-4407.

Teletski, C., & Kaufer, N. F. (1989). Sequence of the ribosomal protein gene KD4 from *Schizosaccharomyces pombe*. Nucl. Acids Res., **17**, 10118.

Thomas, M. S., & Nomura, M. (1987). Translational regulation of the L11 ribosomal protein operon of *Escherichia coli*: mutations that define the target site for repression by L1. Nucl. Acid. Res., **15**, 3085-3096.

Tonnesen, T. & Friesen, J.D. (1973). Inhibitors of RNA synthesis in *S. cerevisiae*: decay rate of mRNA. J. Bacteriol., **115**, 889-896.

Trapman, J., & Plant, R. J. (1975). Detailed analysis of the ribosomal RNA synthesis in yeast. Biochim. Biophys. Acta., **414**, 115-125.

Udem, S. A., Kaufman, K., & Warner, J. R. (1971). Small ribosomal ribonucleic acid species of *Saccharomyces cerevisiae*. J. Bacteriol., **105**, 101-106.

Udem, S. A., & Warner, J. R. (1972). Synthesis and processing of ribosomal RNA in *Saccharomyces cerevisiae*. J. Mol. Biol., **65**, 227- 242.

Udem, S. A., & Warner, J. R. (1973). The cytoplasmic maturation of a ribosomal precursor ribonucleic acid in yeast. J. Biol. Chem., **248**, 1412-1416.

Underwood, M. R., & Fried, H. M. (1990). Characterization of nuclear localizing sequences derived from yeast ribosomal protein L29. EMBO J., **9**, 91-99.

Vazquez, D. (1979). Inhibitors of protein biosynthesis. (pp. 138-144). Springer Verlag, New York.

Veinot-Drebot, L. M., Singer, R. A., & Johnston, G. C. (1988). Rapid initial cleavage of nascent pre-rRNA transcripts in yeast. J. Mol. Biol., **199**, 107-113.

Wandzilak, T. M., & Benson, R. W. (1977). Yeast RNA polymerase III: A zinc metalloenzyme. Biochem. Biophys. Res. Comm., **76**, 247-252.

Warner, J. R. (1979). Distribution of newly formed ribosomal proteins in HeLa cell fractions. J. Cell. Biol., **80**, 767-772.

Warner, J. R. (1989). Synthesis of ribosomes in *S. cerevisiae*. Microbiol. Rev., **56**, 256-271.

Warner, J. R., & Gorenstein, C. (1977). The synthesis of eukaryotic ribosomal proteins *in vitro*. Cell, **11**, 201-212.

Warner, J. R., & Gorenstein, C. (1978). The ribosomal proteins of *Saccharomyces cerevisiae*. *In* D. M. Prescott (Ed.), Methods in Cell Biology XX. (pp. 45-60). New York, San Francisco, London: Academic Press.

Warner, J. R., Mitra, G., Schwindinger, W. F., Studeny, M., & Fried, H. M. (1985). *Saccharomyces cerevisiae* coordinates accumulation of yeast ribosomal proteins by modulating mRNA splicing, translational initiation, and protein turnover. Mol. Cell. Biol., **5**, 1512-1521.

Warner, J. R., & Udem, S. A. (1972). Temperature sensitive mutations affecting ribosome synthesis in *Saccharomyces cerevisiae*. J. Mol. Biol., **65**, 243-257.

Weber, H. J. (1972). stoicheiometry measurements of 30S and 5S ribosomal proteins from *Escherichia coli*. Mol. Gen. Genet., **119**, 233-248.

Whittaker, P. A., Campbell, A. J. B., Southern, E. M., & Murray, N. M. (1988). Enhanced recovery and restriction mapping of DNA fragments cloned in a lambda vector. Nucl. Acid. Res., **7**, 1513-1523.

Wilbur, W. J., & Lipman, D. J. (1983). Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. USA., **80**, 726-730.

Wilson, T., & Triesman, R. (1988). Removal of the poly(A) and consequent degradation of *c-fos* mRNA facilitated by 3' AU-rich sequences. Nature, **336**, 396-399.

Wimber, D. E., Duffey, P. A., Steffensen, D. M., & Prensky, W. (1974). Localization of the 5S RNA genes in *Zea mays* by RNA-DNA hybridization *in situ*. Chromosoma, **47**, 353-359.

Wimber, D. E., & Steffensen, D. M. (1970). Localization of 5S RNA genes on Drosophila chromosomes by hybridization. Science., **170**, 639-641.

Wood, K. O., & Lee, J. C. (1976). Integration of synthetic globin genes into an *E. coli* plasmid. Nucl. Acid. Res., **3**, 1961-1971.

Woolford Jr, J. L. (1989). Nuclear pre-mRNA splicing in yeast. Yeast, **5**, 439-457.

Woolford Jr, J. L., & Rosbash, M. (1979). The use of R-looping for structural gene identification and mRNA purification. Nucl. Acid. Res., **6**, 2483-2497.

Woudt, L. P., Mager, W. H., Niewint, R. T. M., Wassenaar, G. M., van der Kuyl, A. C., Murre, J. J., Hoekman, M. F. M., Brockhoff, P. G. M., & Planta, R. J. (1987). Analysis of upstream activation sites of yeast ribosomal protein genes. Nucl. Acid. Res., **15**, 6037-6048.

Woudt, L. P., Smit, A. B., Mager, W. H., & Planta, R. J. (1986).Conserved sequence elements upstream of the gene encoding yeast ribosomal protein L25 are involved in transcription activation. EMBO J., **5**, 1037-1040.

Yamamoto, T., Matsui, Y., Natori, S., & Obinata, M. (1989). Cloning of a housekeeping-type gene (*MER5*) preferentially expressed in murine erythroleukemia cells. Gene, **80**, 337-343.

Yen, T. J., Gay, D. A., Pachter, J. S., & Cleveland, D. W. (1988). Autoregulated changes in stability of polyribosomal-bound β-tubulin mRNAs are specified by the first 13 untranslated nucleotides. Mol. Cell. Biol., **8**, 1224-1235.

Yonath, A., Leonard, K. R., & Wittman, H. G. (1987). A tunnel in the large subunit revealed by three-dimensional image reconstruction. Science, **236**, 813-816.

Zimmermann, R. D. (1980). Interactions among protein and RNA components of the ribosome. *In* G. Chamblis, C. G.R., K. Davies, L. Kahan, & M. Nomura (Ed.), Ribosomes, Structure, function and Genetics. (pp. 135-169). Baltimore: University Park Press.

Zurawski, G., & Zurawski, S. M. (1985). Structure of the *Escherichia coli* S10 ribosomal protein operon. Nucl. Acid. Res., **13**, 4521-4526.