Mitochondrial DNA replication in the sea urchin *Strongylocentrotus purpuratus*

A thesis submitted for the degree of
Doctor of Philosophy
at the
University of Glasgow.

Andrew Geoffrey Mayhook

Institute of Genetics
University of Glasgow
Church Street
GLASGOW
G11 5JS

May 1992

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

This thesis is dedicated to my mum and dad
for all their support and patience throughout.

I think he sits at that strange table
of Eddington's, that is not a table
at all, but nodes and molecules
pushing against molecules
and nodes; and he writes there
in invisible handwriting, the instructions
the genes follow.

from At It, by R.S.Thomas
Oxford Book of Contemporary
Verse 1945-1980, OUP

LIST OF CONTENTS

CHAPTER 5 - *IN VITRO* TRANSCRIPTION OF THE A6/COIII REGION

CHAPTER 6 - CONCLUDING REMARKS

ABBREVIATIONS

1D	-	one-dimensional
2D	-	two-dimensional
12S	-	small subunit mitochondrial rRNA
16S	-	large subunit mitochondrial rRNA
A6/A8/A9	-	subunits 6, 8 and 9 of ATP synthase
AEEF	-	activated enucleate egg fragment
CSB	-	conserved sequence block
COI-III	-	subunits I-III of cytochrome *c* oxidase
cp DNA	-	chloroplast DNA
cyt *b*	-	cytochrome *b*
D-loop	-	displacement loop
D-strand	-	newly-synthesised DNA molecule which forms the displacement loop
EM	-	electron microscopy
EtBr	-	ethidium bromide
H-strand	-	heavy-strand of mitochondrial DNA
HSP	-	heavy-strand promoter
L-strand	-	light-strand of mitochondrial DNA
LSP	-	light-strand promoter
mRNA	-	messenger RNA
MRP	-	RNase mitochondrial RNA processing
mtDNA	-	mitochondrial DNA
ND1-6/4L	-	subunits 1-6 and 4L of NADH dehydrogenase
nDNA	-	nuclear DNA
O_H	-	heavy-strand replication origin
O_L	-	light-strand replication origin
PFGE	-	pulse-field gel electrophoresis
RI	-	replication intermediate
rRNA	-	ribosomal RNA
snRNP	-	small nuclear ribonucleoprotein
tRNA	-	transfer RNA
TAS	-	termination associated sequence

Other abbreviations used in this thesis are as defined in The Merck Index, 11th edition (1989). Editors: S. Budavari *et al.* Merck & Co, Inc, N.J., USA. pp. xiii-xix.

ACKNOWLEDGEMENTS

Firstly I would like to thank all the members of the department, past and present, who have provided advice, encouragement, and friendship during my time here; I am sorry that I do not have the space to mention everyone individually.

I must, of course, thank my supervisor Howy Jacobs for his numerous ideas and his guidance over the course of this project, and for the considerable amount of time and red ink expended in reading and commenting on this thesis. The other members of the lab also deserve a special mention for providing such a pleasant and relaxed atmosphere to work in, and for supplying important solutions at crucial moments, even if they didn't always realise it! My thanks in particular to David and Arvind for showing me the finer points of experimental technique, and to them, and Daniela, Denise, Keif, Leslie, and Rick for numerous coffees, pints, letting me win at squash, and evenings of culture. Thank you also to Sohail, for aliquots of his magic protein fractions; to the media ladies for providing media, glassware *etc*, without which the department would grind to a halt; and to Mary, for providing all sorts of useful advice.

Finally, my thanks to Thomas for many evenings in the pub comparing PhDs, and to him, and Jeremy, Martyn, and Stephen for providing some weekends of sanity away from Glasgow and the department; and to John for letting me use his computer, and teaching me how to draw diagrams.

This project was supported by the Medical Research Council, and a Mrs. T. Memorial Fellowship.

SUMMARY

In animal development, the normal regulation of mitochondrial DNA (mtDNA) replication, which ensures a doubling of the amount of mtDNA per cell division, is altered. In oogenesis there is extensive mtDNA synthesis, but no nuclear DNA replication, whereas the situation is reversed in embryogenesis. The aim of this project was to provide basic information about the molecular mechanism of mtDNA replication in the sea urchin *Strongylocentrotus purpuratus*, so that the question of what regulates mtDNA replication can then be addressed. The two-dimensional (2D) agarose gel electrophoresis system developed by Brewer and Fangman (1987) was used to analyse the structures of the replicating mtDNA molecules. This gel system exploits the fact that non-linear DNA molecules are retarded under gel electrophoresis compared to linear DNA molecules. The DNA molecules are subjected to restriction enzyme digestion before electrophoresis, and curves of different, characteristic shapes are generated by replication intermediates (RIs) of different structures. Different restriction enzyme digestions allow the pattern of replication fork movement to be related to the positions of the restriction sites within the genome, which enables the position of the replication origin to be located in the DNA molecules.

The gel electrophoretic analyses provided a complex, but internally consistent set of results, giving a picture of the replication process in *S. purpuratus* mtDNA molecules. The analyses revealed that the replication origin for leading-strand synthesis was located within the non-coding region of the genome, and that replication was unidirectional, by a strand-displacement mechanism, and occurred towards the 12S rRNA gene. The 2D gel experiments also revealed that pause sites for leading-strand synthesis were located at several sites in the mitochondrial genome. The most prominent of these pause sites occurred close to a prominent lagging-strand replication origin, in the region of the genome near the boundary between the genes for subunit 6 of ATP synthase (A6), and subunit III of cytochrome *c* oxidase (COIII).

RNase protection experiments, performed using probes covering the non-coding region of the mtDNA, mapped the 5' end of the nascent DNA strands to nt 1150 $\pm$ 10, and suggested that the 3' end of lagging-strand molecules also mapped to the same location, implying a pause site for lagging-strand synthesis at the leading-strand origin. The ends of the DNA molecules corresponding to the leading-strand replication pause site, and the origin of lagging-strand

synthesis in the A6/COIII region of the genome were, however, below the level of detection of the RNase protection experiments. The ends of the transcripts for the A6 and COIII genes were mapped, which revealed one major and one minor 3' end for the A6 RNA, and a single discrete 5' end for the COIII RNA. From the sizes of the molecules identified by the RNase protection experiments, it appeared that the ends of the 2 transcripts were very close together.

An *in vitro* transcription system was used to investigate whether there is a link between transcriptional control and DNA replication in sea urchin mitochondria. A possible link between the 2 processes was suggested by the location of some of the replication pause sites close to the 3' ends of genes: sites where transcription termination could be involved in the production of transcripts. The *in vitro* transcription system used bacteriophage RNA polymerase, because the enzyme had not been isolated from *S. purpuratus* mitochondria at the time the experiments were performed. Bacteriophages have single subunit RNA polymerases, which show considerable sequence similarity at the amino acid ($\alpha\alpha$) level to the mtRNA polymerase isolated from yeast (Masters *et al*, 1987).

Transcription run-off assays were carried out using the region of the mtDNA containing the A6/COIII gene boundary, because this portion of the genome contains both a replication pause site and a lagging-strand replication origin, as well as being a possible site for transcription termination. When transcription of the mtDNA was carried out in the same direction as RNA synthesis *in vivo*, 6 sites for transcription termination were detected. However, no sites were detected when transcription through the mtDNA occurred in the opposite direction. The termination sites did not correspond with the 3' end of the A6 transcript detected *in vivo* by the RNase protection experiments. This would, therefore, imply that if these sites were acting to terminate transcription *in vivo*, then further processing of the transcript would have to occur to generate the mature mRNA. Although 2 DNA-binding proteins which act close to the A6/COIII gene boundary have been identified in mitochondrial protein extracts (S.A.Qureshi & H.T.Jacobs, unpublished data), the addition of these protein extracts (prepared by SAQ) to the *in vitro* transcription reactions did not appear to affect the pattern of transcription termination.

On the basis of the gel electrophoretic data, I propose a model for mtDNA

replication in *S. purpuratus* (Mayhook *et al*, 1992). The origin of leading-strand replication is located in the non-coding region of the genome at nt 1150 ± 10. DNA synthesis is initially unidirectional, by strand displacement (as proposed for vertebrate mtDNA; Clayton, 1982), towards the 12S rRNA gene. The gel electrophoretic data are consistent with the electron microscopy data of Matsumoto *et al* (1974), which imply that multiple origins exist for lagging-strand synthesis. A prominent lagging-strand origin was detected near the A6/COIII gene boundary, and its occurrence in the same region of the genome as a pause site for leading-strand synthesis suggests that the pausing of leading-strand replication may have a role in the initiation of lagging-strand synthesis.

CHAPTER

ONE

INTRODUCTION

1.1 MITOCHONDRIA - STRUCTURE, FUNCTION, AND EVOLUTION.

In 1898, Benda coined the term mitochondrion to describe the intracellular bodies observed by light microscopy; structures which were first seen in the 1840s, but whose ubiquitous occurrence was not recognised until 1890 by Altmann (reviewed by Ernster & Schatz, 1981). The word mitochondrion is derived from the Greek mitos, meaning thread, and chondros, meaning granule.

Mitochondria are rod-shaped structures 0.5-1.0 um in diameter, bounded by a double membrane, the inner one of which is folded to form a series of structures called cristae. The number of mitochondria vary in different cell types, from none in mature, mammalian erythrocytes, to 500-1000 per mammalian liver cell, and approximately 10^5 per mouse egg cell (Piko & Matsumoto, 1976).

Mitochondria are the site of aerobic respiration in the cell, where sugars and fatty acids are metabolised to generate energy in the form of ATP (Stryer, 1988). Sugars (in the form of pyruvate), and fatty acids are imported into the mitochondrial matrix, where they are converted to acetyl coenzyme A (acetyl CoA), which is then oxidised in the citric acid (TCA) cycle. The products of the TCA cycle ($NADH/H^+$ and $FADH_2$) are reoxidised by the transfer of electrons to the electron carriers present in the inner mitochondrial membrane (Fig. 1.1). Movement of electrons along this electron transport chain is accompanied by the translocation of H^+ ions from the matrix to the inter-membrane space (IMS). ATP synthesis occurs as a result of the movement of these H^+ ions through the ATP synthase protein complex, from the IMS to the matrix, down an electrochemical proton gradient (Fig. 1.1).

Since the discovery of DNA in mitochondria in 1963, a great deal of research has been carried out into the structure (section 1.2), function (section 1.2), replication (sections 1.3 & 1.4), and transcription (section 1.5) of this separate genome within the cell.

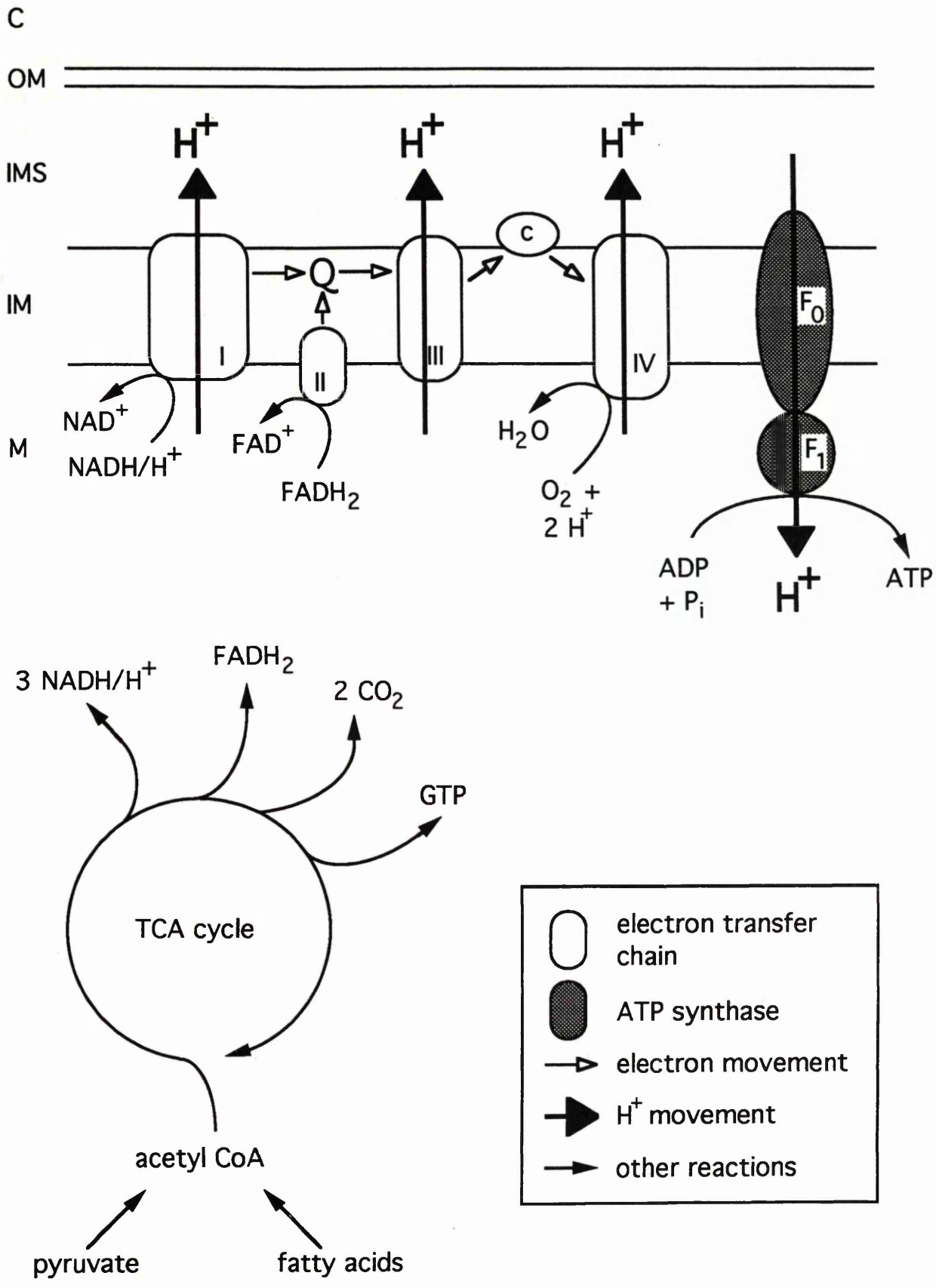


FIGURE 1.1: Outline of the metabolic reactions involved in the generation of ATP from sugars (pyruvate) and fatty acids within the mitochondrion. The protein complexes of the electron transport chain are labelled I-IV: I - NADH dehydrogenase; II - succinate dehydrogenase; III - *b* -*c* $_1$complex; IV - cytochrome *c* oxidase. Other electron carriers are: Q - ubiquinone (coenzyme Q); c - cytochrome *c* . The abbreviations showing the different cellular/mitochondrial compartments are : C - cytoplasm; OM - outer mitochondrial membrane; IMS - inter-membrane space; IM - inner mitochondrial membrane; M - matrix. Adapted from Alberts *et al*, 1989.

The generally accepted theory to explain the evolutionary origin of mitochondria, and their genomes (the endosymbiotic hypothesis; reviewed by Gray, 1989) is that mitochondria developed from the symbiotic association(s) between a proto-eukaryotic cell, and a free-living, aerobic bacterium . This theory is based largely on comparisons between the rRNA sequences from organellar, prokaryotic, and nuclear genomes, which showed that the organellar sequences were prokaryotic (eubacterial) in character, and clearly different from the nuclear rRNA sequences (Yang *et al*, 1985). The transfer of genes from the endosymbiont to the nuclear genome is proposed to account for the much reduced coding capacity of the mitochondrial genome, compared to that of a prokaryotic organism.

Accepting that this transfer of genetic material has occurred, it is a puzzle as to why mitochondrial DNA has been maintained as a "separate" genome, since it only encodes a small subset of the proteins present in the electron transport and ATP synthase complexes (section 1.2). In addition, the components needed to maintain the mitochondrial genome (DNA and RNA polymerases, most (if not all) ribsomal proteins, and transcription factors) are encoded by the nucleus.

One possibility is that some of the proteins encoded by the mitochondrial genome have to be produced within the mitochondria, because their import into the organelle is impossible, or they are toxic if expressed in the cytoplasm. This cannot apply to some genes, such as subunit 9 of ATP synthase (A9), because they are encoded by the mitochondrial genomes of some organisms (*S. cerevisiae*), and the nuclear genomes of others (*N. crassa*, animals). It could, however, apply to those genes present in all mitochondrial genomes so far studied (*e.g.* cytochrome *b*, subunit 6 of ATP synthase (A6), and subunit I of cytochrome *c* oxidase (COI). This hypothesis could be tested by attempting to target the protein from a modified version of the mitochondrial gene present in the nucleus, to the mitochondrion.

Another idea is that at a certain stage of cellular evolution, the transfer of genetic information from the mitochondrion to the nucleus was blocked. However, Thorness and Fox (1990) have shown that in S. cerevisiae the URA3 gene, transformed into the mitochondrion on a plasmid capable of both nuclear and mitochondrial replication, can be transferred to the nucleus of a uracil auxotroph (ura3-52) under selection on uracil deficient medium. Recently, a third idea has been put forward, suggesting that an active mechanism, akin to the hok system of the E. coli plasmid R1, existed to ensure the maintenance of mtDNA over evolutionary time (Jacobs, 1991).

1.2 MITOCHONDRIAL GENOME STRUCTURE AND ORGANISATION.

1.2.1 CODING INFORMATION.

The majority of mitochondrial genomes so far studied have been found to be circular, double-stranded DNA molecules, although in a few organisms the mtDNA is linear. The latter include Chlamydomonas reinhardtii (Boer et al, 1985), and Tetrahymena thermophila (Morin & Cech, 1986). Mitochondrial genomes are also similar to one another with respect to the genes they contain, in that the gene products are involved either in mitochondrial protein synthesis, or are subunits of the ATP synthase, or one of the electron transport complexes. The exceptions to this generalisation are the reverse transcriptase-like genes found in some plant mitochondrial genomes (Schuster & Brennicke, 1987), and the genes encoding maturases present in fungal and plant mitochondrial genomes.

The isolation and characterisation of mtDNA from a host of different organisms has led to the identification of a number of protein coding genes, which are common to the majority of mitochondrial genomes. These are the genes for subunits I, II, and III of cytochrome c oxidase (complex IV), cytochrome b, and subunits 6 and 8 of ATP synthase. However, the absence of the A8 gene in the mitochondrial genomes of nematodes

(Wolstenholme et al, 1987), and the genes for COII and COIII in the mtDNA of Chlamydomonas reinhardtii (Gray & Boer*, 1988), means that only the genes for COI, cyt b, and A6 are encoded by the mtDNA of all organisms so far studied.

The majority of the protein synthesis apparatus in the mitochondrion is encoded by the nucleus, including the RNA polymerase, all transcription factors, and the majority (in some species all) of the ribosomal proteins. All mitochondrial genomes do, however, contain at least 2 rRNA genes (3 in higher plants), although these do vary considerably in size between animals, fungi, plants, and kinetoplastid protozoa. Kinetoplastid protozoa have 9S and 12S rRNA molecules, which are the smallest known rRNAs (Eperon et al, 1983). Looking at the mitochondrial rRNA molecules in animals, plants and fungi, animals have the smallest rRNAs, and plants the largest (18S and 26S; Levings & Brown, 1989), with fungal rRNA molecules being of intermediate size. Higher plants also have a third, much smaller, rRNA gene (5S), which is closely linked to the 18S gene in the genome.

The numbers of tRNA genes present in the mitochondrial genomes of different organisms varies, with those of animals and fungi having more than 20, which in each case are sufficient for their protein synthesis requirements. It is not clear whether this is the case for all higher plant mitochondrial genomes, although it has been reported that they contain around 25 tRNAs. In the bean (Phaseolus vulgaris), 4 mitochondrial tRNAs have been shown to be encoded by nuclear genes, and therefore have to be imported into the mitochondria (Marechal-Drouard et al, 1988). Transformation of one of these bean tRNA genes into potato has shown that import of the tRNA into the mitochondrion can occur in vivo (Small et al, 1992). No tRNA genes have been detected in the mitochondrial genomes of the kinetoplastid protozoa, and purified mitochondrial tRNAs hybridise to nuclear, not mitochondrial DNA, so it appears that all their tRNAs are imported (Simpson et al, 1989; Hancock & Hajduk,

* ref listed on p219.

1990).

The mitochondrial genomes of some protozoa, higher plants, and S. cerevisiae also contain one, or more ribosomal protein genes. These proteins all form part of the small ribosomal subunit (SSU), and are absent from animal mtDNA. Three SSU ribosomal protein genes (rps4, 13 & 14) have been detected in plant mtDNA (*Bland et al, 1986; *Wahleithner & Wolstenholme, 1988), although there appears to be some variation in the genes present in different species. The rps13 gene, for example, has been detected in the mtDNA of tobacco, wheat and maize, but not that of bean and pea (*Bland et al, 1986).

The genes encoding the proteins which form the NADH dehydrogenase electron transport complex (complex I) show a large degree of diversity in their genomic location, which has no clear correlation with species classification. The complex is made up of about 26 protein subunits, up to 7 of which may be encoded in the mtDNA. The genes for 7 of the subunits of complex I have been found in all the animal mitochondrial genomes so far sequenced. In contrast, fungal mtDNA shows a wide variation, with no NADH dehydrogenase genes detected in the S. cerevisiae mitochondrial genome (Grivell, 1989), but 8 (the ND4L gene is duplicated) detected in that of P. anserina (Cummings et al, 1990). Some NADH dehydrogenase subunit genes have been identified in both plant and kinetoplastid protozoan mitochondrial genomes, with the genes for subunits 1 and 5 identified in 3 plant species, and in the kinetoplastid protozoan L. tarentolae, which also has a mitochondrially-encoded subunit 4. It is possible that genes for other subunits of complex I are also present since unidentified ORFs are present in the mtDNA of all of these species. The gene for ND8, a subunit of complex I not located in animal mtDNA, has recently been identified in the mitochondrial genome of another kinetoplastid protozoan, Trypanosoma brucei (Souza et al, 1992).

In addition to the A6 and A8 genes, which are found in all mitochondrial genomes so far studied except nematodes, which lack A8 (Wolstenholme et al, 1987), 2 other subunits of the

* ref listed on p219.

ATP synthase complex are also mitochondrially encoded in some organisms. Plant mitochondrial genomes encode the gene for the α subunit of the F_1 portion of the complex, and the A9 subunit of the F_0 portion The A9 gene is also located in the mitochondrial genomes of some fungi, such as Aspergillus nidulans, but in the nuclear genomes of others (e.g. S. cerevisiae and N. crassa).

Other proteins, such as maturases and reverse transcriptase-like proteins, are also encoded by the mtDNA of particular groups of organisms, and these are discussed in more detail below (section 1.2.2).

1.2.2 SIZE AND GENOME ORGANISATION.

Animal mtDNA ranges in size from 15-20 kb, and contains very little non-coding DNA. No introns have been found in any animal mitochondrial genome yet studied, and the intergenic regions are minimal, with some genes overlapping slightly. In vertebrate mtDNA there are 2 major regions of non-coding DNA, covering each of the 2 replication origins. The region covering the leading-strand replication origin (O_H) ranges in size from 879 bp in mouse mtDNA (Bibb et al, 1981) to 2134 bp in Xenopus mitochondrial genomes (Roe et al, 1985), and also contains the transcriptional promoters for each strand (section 1.5). The region covering the lagging-strand replication origin (O_L) is much smaller, being approximately 40 bp in length, although it is absent in chicken and quail mtDNA (Desjardins & Morais, 1990).

In Drosophila mitochondrial genomes, there is an extensive A+T-rich region covering the replication origins for both strands. This non-coding region shows considerable size variation between species, from 0.8 kb in D. yakuba, to 5.4 kb in D. melanogaster (Goddard & Wolstenholme, 1978; 1980).

Fungal mitochondrial genomes show a much greater variation in size compared to those of animals. They range from 19 kb in

Torulopsis glabrata and Schizosaccharomyces pombe, to 102 kb in some strains of Podospora anserina (Cummings et al, 1990). The range of genome sizes in fungi reflects differences in the amount of non-coding DNA present in the genomes. This occurs as introns and long intergenic sequences; in P. anserina race A, the 36 introns account for 60 kb of the 102 kb genome (Cummings et al, 1990). Two classes of introns have been identified in fungal and plant mtDNA: group I and group II (Tzagoloff & Myers, 1986). This classification is based on their secondary structure, and the presence of conserved sequence motifs (Michel & Dujon, 1983), and there are also differences between the splicing mechanisms of the 2 groups (Jacquier, 1990).

In S. cerevisiae, the introns of the cytochrome b (COB) gene vary in number between strains, and some introns contain ORFs which encode maturases and endonucleases. Maturases are generally involved in the splicing out of the introns which encode them, although the maturase encoded by intron 4 of the COB gene also splices out intron 4 of the COX1 gene (Cech, 1991). Circular and linear double-stranded plasmids have also been found in some fungal and plant mitochondria. In Podospora the amplification of small circular DNA molecules derived from the mitochondrial genome is associated with fungal senescence (Cummings et al, 1985).

Higher plants all have very large mitochondrial genomes, although this apparently does not reflect a large increase in coding capacity over the mtDNA of animals and fungi (Levings & Brown, 1989). The variation in size can be vast, even within a single family; the Cucurbita species have mitochondrial genomes ranging from 330 kb to 2400 kb (Ward et al, 1981). The smallest mtDNAs are found in the Brassica family, and these are about 200 kb in size.

One distinctive feature of the organisation of plant mitochondrial genomes is the presence of pairs of repeat sequences, in either direct or inverted orientation, between which recombination occurs. The restriction mapping of

various plant mitochondrial genomes has revealed a multicircular genome organisation, in which a population of subgenomic circles is present (Lonsdale et al, 1984). These smaller circles could be derived from a single master circle by recombination between the repeat sequences. Restriction fragment analysis of mtDNA from Brassica oleracea, which contains one pair of directly repeated sequences, has shown that all 3 of the predicted circular forms were present in vivo (Palmer, 1988). However, repeat sequences are not universally found in higher plant mtDNA; B. hirta mtDNA lacks any such sequences, existing simply as a 208 kb circle (Palmer & Herbon, 1987).

The mtDNA of higher plants has been found to contain DNA sequences apparently derived from the chloroplast genome. In maize, sequences related to the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase (Rubisco), and 2 chloroplast tRNAs have been detected in the mitochondrial genome (Lonsdale et al, 1983), and both chloroplast and nuclear DNA sequences have been reported in the mtDNA of Oenothera (Schuster & Brennicke, 1987).

The mitochondrial genome of Oenothera contains an ORF which contains sequences similar to the conserved blocks of sequence characteristic of reverse transcriptases (Schuster & Brennicke, 1987). The detection of transcripts from it has prompted the suggestion that the transfer of sequences between organelles, and to and from the nucleus could occur via an RNA intermediate, which is then reverse transcribed, and the DNA copy integrated into the genome.

The unicellular green alga Chlamydomonas reinhardtii also possesses an ORF in its mitochondrial genome which has sequence similarity to reverse transcriptases (Boer & Gray, 1988). However, the organisation of its mtDNA is radically different from that of the higher plants. The C. reinhardtii genome is a linear 15.8 kb molecule, and so it differs from 2 related Chlamydomonas species, which have circular 16 kb genomes (Boer et al, 1985). The most bizarre feature of C.

reinhardtii mtDNA is, however, the fact that the large and small subunit rRNA genes are present in short segments (modules), spread over a 7 kb region of the genome, which also contains 2 tRNA genes, the ND1 gene, and ORF y (Boer & Gray, 1988). Transcription studies have shown that each of the rRNAs is produced from several transcipts, and that the molecules are not spliced together to form continuous large and small rRNA molecules (Boer & Gray, 1988). From analyses of the secondary structure of the continuous 23S and 16S rRNA molecules of E. coli, and the C. reinhardtii molecules, it was observed that the modular rRNAs from C. reinhardtii mitochondria could base-pair to form the characteristic core rRNA structure found in the continuous rRNA molecules of E. coli. It was therefore postulated that the mitochondrially-encoded C. reinhardtii rRNAs are, in fact, functional.

The organisation of the mtDNA from kinetoplastid protozoa, such as Crithidia, Leishmania, and Trypanosoma species, is unusual, in that the genome is made up of 2 distinct populations of molecules - minicircles and maxicircles (Simpson, 1987). They form a single, catenated network of DNA, which is located in the single, large mitochondrion (the kinetoplast) present in each organism. There are 500-5000 minicircles per kinetoplast, and the sizes vary between species, from 454 bp in T. vivax, to 2515 bp in C. fasciculata. All minicircle DNAs analysed contain a small conserved region and a larger variable region, which shows varying degrees of heterogeneity in different species (Simpson, 1987).

The maxicircles, which are about 20 kb in size, and present in 20-50 copies per kinetoplast, are analogous to the mtDNA molecules of other organisms, though some of the transcripts undergo extensive editing before they can be translated. The process of RNA editing involves the addition and/or deletion of bases (normally U residues) to the primary transcripts in order to create mRNA molecules which can be translated to produce functional proteins. Transcripts which undergo RNA editing include those for COII, COIII, A6 and cyt b in T.

brucei, L. tarentolae, and C. fasciculata (reviewed by Stuart, 1991), with the most extreme example of editing being the COIII gene of T. brucei, in which 58% of the protein coding sequence is the result of the addition and deletion of uridine residues (Feagin et al, 1988).

1.3 MtDNA REPLICATION IN ANIMALS.

1.3.1 MECHANISM OF mtDNA REPLICATION.

The first studies of the replication of mtDNA were carried out using electron microscopy (EM) to examine replicating molecules isolated from a number of different organisms, including mouse, human, rat, Drosophila and sea urchin.

The studies of mtDNA from mouse L cells (Kasamatsu & Vinograd, 1973; Kasamatsu et al, 1974) revealed that in 25-30% of the molecules examined, a small region of the mtDNA molecule was displaced by synthesis of a daughter strand forming a single-stranded loop, the size of which was always approximately 3.5% of the length of the genome. This type of molecule was termed the D-loop form, and was also observed in mtDNA preparations from human HeLa cells (Robberson et al, 1972) and rat hepatomas (Wolstenholme et al, 1973a;b). The actual percentage of molecules having this form varied, depending on the organism studied, the growth phase of the cells, and the isolation procedure used.

In other molecules, the displaced portion of the genome was much larger, extending up to the full length of the circular mtDNA molecule (Robberson et al, 1972; Kasamatsu & Vinograd, 1973), and was either completely or partially single-stranded. By analysing the distribution and extent of the double-stranded portion of the displaced strand in HeLa cell mtDNA, Robberson et al (1972) deduced that replication of the second-strand (initiated on the displaced H-strand) did not occur until at least 55% of the genome had been displaced, but that in a third of molecules with more than 55% D-loop

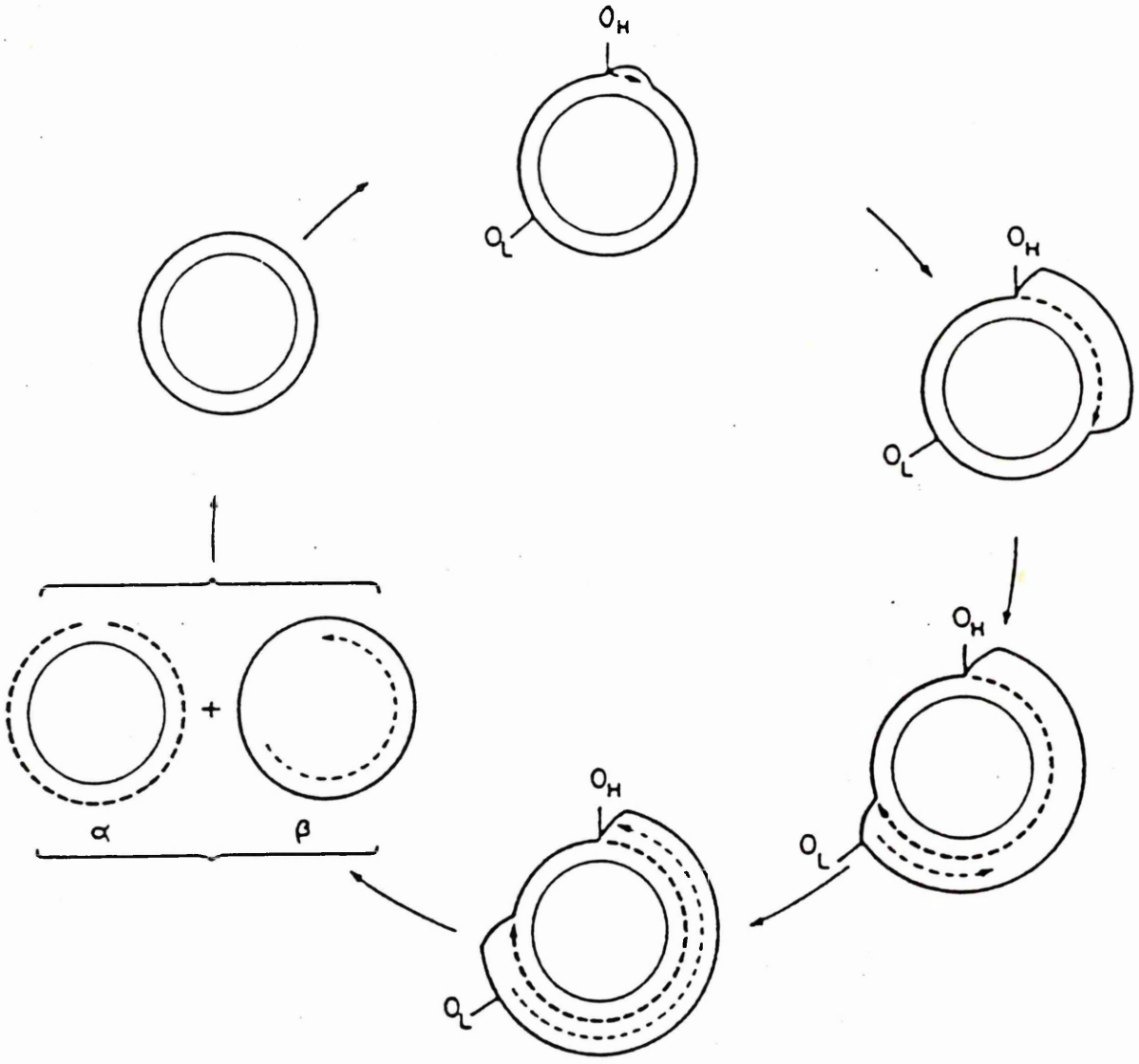


FIGURE 1.2: Model for the replication of animal mtDNA. Thick lines represent the heavy (H) strand, and thin lines, the light (L) strand of the mtDNA molecule. The newly-synthesised strands are shown as dashed lines, and the parental strands as solid lines. O_H and O_L mark the positions of the heavy- and light-strand replication origins, respectively. Adapted from Clayton, 1982.

(facing page 12)

expansion no duplex synthesis had occurred. By assuming that replication by D-loop expansion was unidirectional, they showed that the start of duplex synthesis occurred at 0.6-0.7 genome length (G), and that the displaced strands could be aligned so that the distance between the fixed end of the replication bubble and the distal end of the duplex occurred between 0.5-0.7 G, with an average at 0.6 $\pm$ 0.07 G.

The unidirectional nature of mtDNA replication was confirmed by the elegant experiments of Kasamatsu & Vinograd (1973), in which they studied replication in the double-length mtDNA molecules of mouse LD cells. These molecules exist as head-to-tail dimers, and approximately 15% of molecules examined by electron microscopy contained 2 D-loops, 180^{o} apart on the genome. The presence of molecules containing both a small and an expanded D-loop meant that these genomes contained a marker, visible by electron microscopy, which was fixed in relation to the replication fork. From the measurement of the 2 halves of the molecule between the fixed D-loop fork to the distal end of the expanded D-loop fork, it was clear that these distances were the same, so replication had to be unidirectional. Of all the molecules examined, none were found to have 2 expanded D-loops, and a second D-loop was not found in molecules in which D-loop expansion was greater than one genome length. As with the experiments of Robberson et al (1972), annealing of heavy- or light-strand DNA (the 2 mtDNA strands have different densities) to the replicating molecules revealed that initial displacement was of the heavy (H)-strand (Kasamatsu & Vinograd, 1973).

These electron microscopy studies led to the proposal of the expanded D-loop model for mtDNA replication. Extension of the D-loop molecule by single-stranded, unidirectional DNA synthesis leads to the displacement of the H-strand (Fig. 1.2). Once the replication fork has passed the L-strand origin (O_L), unidirectional replication of the L-strand back towards the H-strand origin can occur, implying that O_L can only function when single-stranded. Separation of the 2 daughter molecules precedes the completion of L-strand

replication (Fig. 1.2), and the ligation of the 2 ends of each daughter strand together, after which D-loop synthesis can recommence.

Electron microscopy studies of rat and chick mtDNA molecules have revealed a similar appearance of replication intermediates, with displacement loops of between 3 and 95% of the genome present, which were either completely or partially single-stranded. In 2 rat hepatoma cell lines studied (Chang & Novikoff), D-loop forms were seen in 9 and 25% of the molecules respectively, and the loop represented 3.7% of the genome length, similar to that seen in mouse (Wolstenholme et al, 1973a). However, unlike in mouse mtDNA, molecules were observed in which the replication intermediates were completely double-stranded, with the replicated portion being up to 80% of the length of the genome. These types of molecules were also seen in mtDNA from the regenerating rat liver cells, so they were not peculiar to the hepatoma cell lines.

From these observations it was proposed that 2 different replication schemes exist for rat mtDNA. The first was the expanded D-loop model put forward for mouse mtDNA (Robberson et al, 1972), whereas the second involved a stepwise replication of the molecule to form the completely double-stranded replication intermediates; structures not observed in mouse mtDNA.

Further analysis of the replicating mtDNA from rat Novikoff hepatoma cells (Wolstenholme et al, 1973b) showed that the sizes of the replicated portions of molecules fell into 8 distinct groups, with the mean lengths occurring 4.1 to 7.6% of the genome length apart. This implies that there were sites present in the genome 650-950 nt apart, at which replication paused, possibly due to the loss of external factors needed for replication during the purification procedure.

In the sea urchin Strongylocentrotus purpuratus, mtDNA

molecules isolated from oocytes were found to have an even higher incidence (25% of all expanded D-loop forms) of completely double-stranded Cairns' form replication intermediates than was seen in rat (Matsumoto et al, 1974). The replication intermediates also differed from those of rat and mouse, in that in the partially single-stranded molecules, multiple regions of duplex DNA were observed. Approximately 25% of the expanded D-loop molecules observed contained multiple regions of duplex DNA, implying that multiple initiations of second- strand synthesis occurs.

Another difference from mouse L cell mtDNA was the presence of duplex synthesis in genomes with much smaller expansion loops. It was observed that in nearly all molecules with 15-20% of the genome replicated, duplex synthesis had occurred (Matsumoto et al, 1974), in contrast to the 60% expansion of the D-loop required in mouse mtDNA (Kasamatsu et al, 1974). The low frequency of gapped circles observed in S. purpuratus mtDNA, which represent an important replication intermediate in mouse mtDNA replication, also suggests differences in the mechanism of replication between the 2 organisms.

In S. purpuratus, the presence of both Cairns' form, and multiple duplex replication intermediates suggested a replication mechanism somewhere between the D-loop expansion, and the stepwise double-stranded mechanisms proposed for mouse and rat mtDNA respectively. The observations point to DNA replication by D-loop expansion with multiple, possibly random lagging-strand origins. Pauses in leading-strand replication at these second-strand origins could lead to the formation of double-stranded, Cairns' form replication intermediates in some instances.

In sea urchin oocyte mtDNA, molecules were observed with displacement loops of 3.2 $\pm$ 0.7% of the genome length, which were taken to be the D-loop form of mtDNA (Matsumoto et al, 1974). However, the subsequent sequencing of S. purpuratus mtDNA (Jacobs et al, 1988) revealed that the major non-coding region of the mitochondrial genome is only 121 bp long. A D-

loop of 121 nt or less would not be visible under the electron microscope, and would be much smaller than the D-loop observed by Matsumoto et al (1974), which was about 500 nt in length, and may therefore represent a discrete pause in DNA synthesis in 4-7% of the molecules examined. The small size of the proposed D-loop could account for the failure of Matsumoto et al (1974) to observe D-loop forms by EM in mature egg mtDNA molecules isolated from S. purpuratus, even though they are present in at least 80% of mitochondrial genomes from unfertilised Xenopus eggs (Hallberg, 1974).

Two stop sites for DNA replication were observed in mouse L cell mitochondrial genomes (Bogenhagen et al, 1978). An accumulation of L-strand daughter molecules of 10.5 kb was observed, which was interpreted as representing the pausing of L-strand synthesis at the H-strand origin. The halting of L-strand synthesis until H-strand synthesis was completed, followed by the separation of the 2 daughter molcules, would lead to the formation of a high percentage of gapped replication intermediates, and could account for their presence in preparations of replicating mouse mtDNA (Robberson et al, 1972). There was also an accumulation of H-strands of 13 kb, which would map the 3′ end of the growing strand close to the 3′ end of the 16S rRNA gene (Bibb et al, 1981). This suggests a possible conflict between replication and transcription, or some type of coordinated regulation of the 2 processes.

1.3.2 ANALYSIS OF 7S mtDNA AND THE D-LOOP REGION.

The newly synthesised DNA strands (D-strand), which are responsible for the formation of D-loops, have been isolated from the mtDNA of mouse, human, and Xenopus cells, and analysed by gel electrophoresis. The initial isolation was carried out by Kasamatsu et al (1971), who identified a 7S DNA species from mouse mtDNA by its sedimentation as a sharp band after sucrose gradient centrifugation. Electrophoretic analysis revealed that this 7S DNA was not a single species,

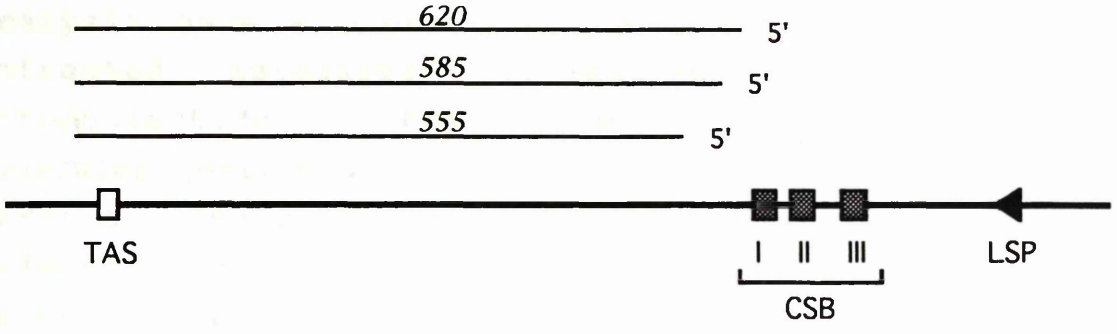
but that it was made up of molecules of several different lengths. Two different analyses of human 7S DNA from HeLa cells gave bands of 680 and 630 nt (Crews et al, 1979), and 620, 585, and 555 nt (Brown et al, 1978), and further experiments on human liver cell and KB cell mtDNA detected bands of the same sizes, although not all of them were always present, and the relative intensities of the different bands varied (Gillum & Clayton, 1978; Tapper & Clayton, 1981).

Restriction enzyme digestion of purified 7S DNA from human cells (HeLa, KB and liver), which had been labelled with ^{32}P at either the 5′ or 3′ end, and then hybridised to excess L-strand DNA, was used to determine the location and extent of the heterogeneity (Gillum & Clayton, 1978; Brown et al, 1978). This revealed that the 5′ end was clearly heterogeneous, while further analysis of the 3′ ends of the molecules from KB cells (Doda et al, 1981) gave a single band on AvaII digestion, indicating a common stop site for D-loop synthesis (Fig. 1.3a).

In mouse mtDNA, the 7S DNA was shown to be a heterogeneous population of molecules (Gillum & Clayton, 1978; 1979), with 5 major forms detected, ranging in size from 550 to 670 nt. However, in contrast to the situation in human mtDNA, they were all shown to have the same 5′ end, except for the longest (670 nt) form. Restriction enzyme analysis revealed that the heterogeneity occurred at the 3′ end of the 7S DNA strands, with 4 different end points present within a 100 nt stretch of the non-coding region (Fig 1.3b; Doda et al, 1981).

End-group analysis of the 7S DNA molecules of both mouse and human by thin layer chromatography, and DNA sequencing (Gillum & Clayton, 1979; Doda et al, 1981; Tapper & Clayton, 1981) revealed microheterogeneity at their 5′ and 3′ ends. Differences of 2-5 nt were commonly observed in the positions of the start and stop sites of molecules previously identified as being the same length, though the 5′ ends of the mouse D-strand molecules showed even greater variation (10-15 nt).

a) Human



b) Mouse

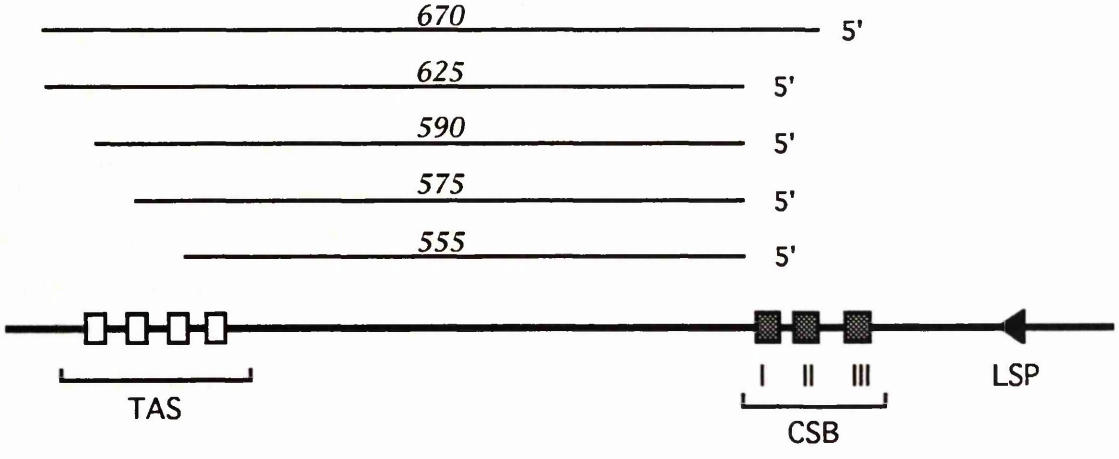


FIGURE 1.3: Maps of the D-loop regions of (a) human and (b) mouse mtDNA, showing the positions of the 5' and 3' ends of the D-strand molecules, and their sizes in nt (from Brown *et al*, 1978, and Gillum & Clayton, 1979, for (a) and (b), respectively), the conserved sequence blocks (CSB I-III; shaded boxes), and the "termination" sequences (TAS; open boxes). The light-strand promoter (LSP) is shown as a filled arrow head, indicating the direction of transcription.

Thin layer chromatographic analysis of the mouse D-strand molecules initiating at the upstream start site (the 670 nt molecules) revealed that either rA or rG was present at the 5′ terminus. Alkali treatment of these molecules prior to analysis gave a doublet up to 10 nt shorter than the untreated molecules, indicating the presence of ribonucleotides at the 5′ end of the 670 nt D-strand molecules (Gillum & Clayton, 1979). In contrast, thin layer chromatographic analysis of the 5′ termini of D-strand molecules at the downstream start site showed that the termini are mostly deoxyribonucleotides.

Similar analysis carried out using 7S DNA from human KB cells failed to detect any ribonucleotides at the 5′ ends of the molecules (Tapper & Clayton, 1981), implying that the removal of the RNA primer for D-loop synthesis is a more effecient process in human than in mouse. No ribonucleotides were detected in Xenopus 7S DNA molecules either (Gillum & Clayton, 1978), though 1-2 nt were present at the 5′ end of the 2 rat D-loop species (Sekiya et al, 1980). As with the mouse D-strand molecules, the major variation in size between the 2 different classes of D-strand molecules in rat occurs at the 3′ end, where there are 2 different stop sites.

Examination of the DNA sequence of the non-coding region around the 3′ ends of the human and mouse 7S DNA molecules revealed a 12-15 nt sequence (the termination associated sequence, or TAS) with more than 75% sequence similarity between the 2 species. The TAS is present in a single copy in humans, and 4 copies in mouse (Doda et al, 1981). In humans, this sequence lies 51-53 nt upstream of the stop site, and in mouse the 4 sequences lie 24-63 nt 5′ to the corresponding end point. It is, however, unclear how these sequences act to terminate DNA synthesis.

The actual role of the D-loop in mtDNA replication is not clear. There is as yet no evidence that the D-strand molecules act as primers for replication. The rapid turnover of the D-strand molecules in mouse (half-life of about 1 hr)

means that the majority cannot act as replication primers, since the frequency of the initiation of productive replication is much lower than the rate of D-strand turnover (Bogenhagen & Clayton, 1978). Analysis of the 5′ ends of nascent H-strands over 1 kb in length (ensuring that all D-strand DNA molecules were excluded), isolated from human KB cells (Tapper & Clayton, 1981), gave the same pattern as that observed for D-strand molecules, confirming that the initiation of D-strand synthesis, and nascent H-strand molecules occurred at the same location. This therefore implies that the D-strand molecules represent aborted replication events, and that DNA replication is not solely controlled at the level of the initiation of DNA synthesis. The production of different length D-loop molecules, and their high rates of turnover suggest that D-loop synthesis could be involved in other processes besides DNA replication. The location of the transcriptional promoters in the non-coding region of the genome suggests the possibility that D-loop synthesis could affect transcription by altering the secondary structure of the mtDNA in the non-coding region.

Sequence analysis of the non-coding region of the mitochondrial genome, which contains the H-strand replication origin, has revealed a large degree of divergence between species. This is particularly striking amongst the vertebrates, given the high degree of conservation of mitochondrial genome organisation and similarity between the protein coding genes (63-79% similarity between bovine and human genes; Anderson et al, 1982).

Comparisons of the D-loop region of mouse, rat, and human mtDNA (Walberg & Clayton, 1981), revealed 3 short conserved blocks of sequence (CSB I, II, and III) around the 5′ ends of the D-loop molecules, and a 225 nt region at the centre of the D-loop region, which showed more than 65% similarity between the 3 genomes. Finally there is the 12-15 nt TAS at the 3′ end of the 7S DNA molecules, mentioned above. The 3 CSBs are also present in Xenopus laevis mtDNA (Roe et al, 1985), although CSB III is absent in the bovine non-coding

region (Anderson et al, 1982).

The role of the CSBs in the mitochondrial genome appears to involve the transition from RNA to DNA synthesis (section 1.3.4). The high number of consecutive C residues in CSB II is of particular note given the ability of homopolymeric runs of G.C base-pairs to form unusual DNA structures (Hanvey et al, 1988; Kohwi & Kohwi-Shigematsu, 1988), and the presence of an extensive run of consecutive G residues in the non-coding region of sea urchin mtDNA (Jacobs et al, 1988; Cantatore et al, 1989).

1.3.3 THE PRIMING OF DNA SYNTHESIS AT O_H.

Although it was clear from the presence of ribonucleotides at the 5′ ends of mouse 7S DNA molecules that replication was primed by RNA, the location of the 5′ end of the primer was not established until 1985 by Chang et al. By using a probe complementary to the D-strand molecules for a region within the non-coding region, upstream of the 5′ ends of the D-strand molecules, abundant short RNA molecules of 140-155 and 80-120 nt were detected in total mitochondrial nucleic acid (Chang et al, 1985a). Other longer transcripts (200-2500 nt) were also observed, but were much less abundant than the shorter RNAs. The short transcripts were not detected using a probe lying within the region covered by the D-strand molecules. Primer extension reactions, performed using a probe located inside the 5′ end of the major D-strand molecules, mapped the 5′ ends of the D-strand molecules to nt 16036, and nt 16093-16103, as before (Gillum & Clayton, 1979). Another new band was observed, generated by RNA molecules, mapping their 5′ ends to nt 16183. RNA primer molecules initiating at this site would therefore generate 2 different classes of RNA molecules, depending on which DNA start site processing of the primer molecules occurred.

Pretreatment of the total nucleic acid sample before primer extension analysis with RNase T1 or alkali showed that some

of the shorter D-strand molecules had ribonucleotides covalently attached to their 5′ termini, and that an RNA to DNA transition event occurred between nt 16029 and 16037 (Chang et al, 1985a). S1 nuclease mapping of the abundant small RNAs located upstream of the D-strand molecules revealed that the 3′ ends of many of them aligned with the 5′ ends of the D-strand molecules, and that some of the RNA molecules had covalently-attached deoxyribonucleotides at their 3′ termini (Chang et al, 1985a).

The same results were obtained using CsCl-gradient purified mtDNA, indicating that the covalent attachment of RNA to DNA occurs in D-loops in vivo. These experiments therefore demonstrated that D-strand synthesis was primed by RNA, and that the RNA primers initiated at a single site, which occurs either 80 or 150 nt upstream from the start of DNA synthesis, depending on the initiation site used for D-strand synthesis.

The non-coding region of vertebrate mtDNA contains the transcriptional promoters for both the heavy (major-coding) and light (minor-coding) DNA strands, and these promoters are designated the HSP and LSP respectively. The localisation of the LSP in mouse mtDNA (Chang et al, 1985a) revealed that the 5′ end of the primer RNA mapped to the same site as the LSP. The site of primer initiation in human KB cell mtDNA (Chang & Clayton, 1985) was mapped using a similar approach to that taken in mouse, although in the case of human mtDNA, no covalently linked ribonucleotides were detected in the D-strand. S1 nuclease mapping of the 3′ ends of the RNA transcripts initiating from the LSP showed a good alignment of 2 groups of ends with the 5′ ends of the D-strand molecules, with the presumed transition site falling within the region containing the 3 CSBs, as in mouse mtDNA. However, the 5′ terminus of one of the D-strand molecules occurred upstream of the LSP, leading to speculation that it was primed from the minor LSP observed in the in vitro transcription studies (Chang & Clayton, 1984), although no transcript was detected initiating there in vivo (Chang & Clayton, 1985).

1.3.4 ENZYMES INVOLVED IN THE RNA TO DNA TRANSITION IN D-LOOP SYNTHESIS.

In mouse LA9 cells, an endoribonuclease has been identified, which will cleave the RNA transcript initiating from the LSP *in vitro* (Chang & Clayton, 1987a). It was identified using an RNA substrate, generated by *in vitro* transcription of a 270 nt region of the mouse non-coding region sequence covering the LSP and the 3 CSBs. A mitochondrial lysate from mouse cells was incubated with the substrate RNA, and found to cleave it at a site which mapped to one of the previously identified 5′ ends of the D-strand molecules mapped *in vivo* (nt 16103; Chang *et al*, 1985a).

When the equivalent region of human mtDNA was used to produce the substrate RNA molecule, which in effect introduced the mouse CSBs into foreign RNA, since these regions are the only ones in common between the 2 species, the mouse endoribonuclease cleaved the RNA (Chang & Clayton, 1987a). The reciprocal reaction, using the equivalent human mitochondrial protein fraction, and the mouse RNA substrate, also led to cleavage of the RNA substrate. This implies that the CSBs, or some feature of RNA structure, which are conserved between the 2 sequences, are being recognised by the processing enzyme. The actual cleavage event was found to occur between CSB II and CSB III, at the RNA consensus sequence 3′-G(G/A)AGA-5′ (Chang & Clayton, 1987a).

The mouse endoribonuclease, termed RNase MRP (mitochondrial RNA processing), was shown to contain an RNA moiety essential for its activity (Chang & Clayton, 1987b). Micrococcal nuclease treatment inactivated its cleavage of substrate RNA molecules, as did RNase A, but DNase I had no effect. The RNA was identified as being a 136 nt molecule, the elution of which from a mono Q FPLC column correlated with the activity profile of the enzyme. Another RNA, 275 nt in length, was also shown to be associated with MRP activity, and contained sequences identical to the smaller RNA. The identification and sequencing of the nuclear gene for the RNA molecules of

RNase MRP in mouse revealed that the 136 nt RNA is encoded by the 3′ end of the gene for the longer molecule (Chang & Clayton, 1989), and that the MRP RNA contains regions of sequence complementary to the CSB II and III regions of mtDNA.

Characterisation of the human MRP enzyme has shown that the gene for the RNA moiety is located in the nucleus, and that it is 84% identical to the mouse gene (Topper & Clayton, 1990b). The 5′ flanking sequences are also highly conserved, showing 70% similarity over a 700 nt region, which is unusual for genes encoding small RNAs. As is the case for the mouse enzyme, 2 different forms of human MRP exist, containing either a large (265 nt) or a small (108 nt) RNA molecule, with the 108 nt RNA encoded by the 3′ end of the gene. Although the cleavage of the RNA to form the small molecule does not occur at the same position in the mouse and human molecules, it does take place at a conserved nucleotide sequence (ANCCCGC). The inhibition of human MRP activity by oligonucleotides complementary to the 5′ end of the 265 nt RNA molecule implies that the complete RNA molecule is assembled into the active enzyme complex (Topper & Clayton, 1990b), but it is not yet clear whether the 3′ portion of the RNA alone can form an active enzyme complex, as is the case in mouse (Chang & Clayton, 1987b).

Deletion analysis of the part of the non-coding region of mouse mtDNA covering the CSBs showed that both CSB II and III were needed in the RNA substrate for it to be processed efficiently by RNase MRP (Bennett & Clayton, 1990). Saturation mutagenesis of the region covering CSB II and III, and the sequence between them, revealed that only a few mutations in the region between the 2 CSBs affected MRP actvity, and these were mainly confined to the cleavage site. Mutations in both CSB II and III had much greater effects on MRP action, with the majority of mutations leading to decreased cleavage efficiency (Bennett & Clayton, 1990).

Structural analysis of the 275 nt RNA molecule from mouse MRP

(Topper & Clayton, 1990a) showed that the region of RNA complementary to the CSB II and III sequences was present largely in a single-stranded region of the molecule. The human MRP molecule could also be folded to adopt a similar conformation, which was consistent with a limited structural analysis of the molecule (Topper & Clayton, 1990a). Oligonucleotides complementary to the single-stranded region of mouse MRP RNA were shown to inhibit MRP action *in vitro* (Bennett & Clayton, 1990; Chang & Clayton, 1987b; Topper & Clayton, 1990b), and mutations in CSB II and III which affected more than one of the potential base-pairs between the MRP RNA and the substrate RNA (Fig. 1.4) had greater effects on MRP activity than mutations removing a single potential pairing (Bennett & Clayton, 1990). All these data suggest that the base-pairing of MRP RNA with its substrate is functionally important, as is the case for the snRNPs U1, U2 and U7 and their target RNA molecules in nuclear pre-mRNA splicing (Parker *et al*, 1987).

When considering the results of the MRP cleavage reactions *in vitro*, it should be noted that the cleavage site seen using mouse MRP and substrate RNA (around nt 16099; Fig. 1.4; Chang & Clayton, 1987a; Bennett & Clayton, 1990) does not occur at the major D-strand 5′ end mapped *in vivo*, which occurs around nt 16030, at the 3′ end of CSB I (Chang *et al*, 1985a). This therefore implies either that the specificity of MRP is altered in the *in vitro* reactions, and the enzyme can act at multiple sites *in vivo*, or that other enzymes are involved in the processing reactions of primer formation.

Analysis of the subcellular distribution of RNase MRP activity in mouse revealed that it was present at about equal levels in the nuclear and mitochondrial compartments (Chang & Clayton, 1987b). Northern blot analysis showed that both the 275 and 136 nt RNA molecules were present in both the nucleus and the mitochondrial fractions, but that their relative abundance differed. The 275 nt RNA was present in at least 100-fold excess over the 136 nt RNA in the nucleus, but the situation was reversed in the mitochondria. In humans the

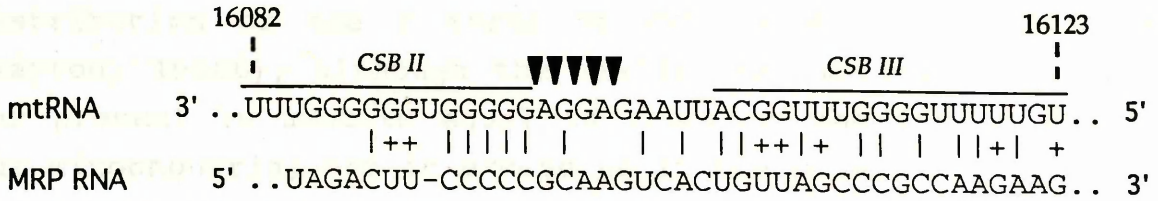


FIGURE 1.4: Proposed base-pairing interaction between mouse MRP RNA and the mtRNA substrate in the formation of the primer for mtDNA replication. The CSB sequences are shown overlined; the sites of cleavage of the mtRNA are shown by the arrowheads; Watson-Crick base-pairs are shown as I, and G-U base pairs as +. The dash (-) in the MRP RNA sequence represents a space introduced to align the 2 sequences. The coordinates of the mouse sequence are given. Adapted from Bennett & Clayton, 1990.

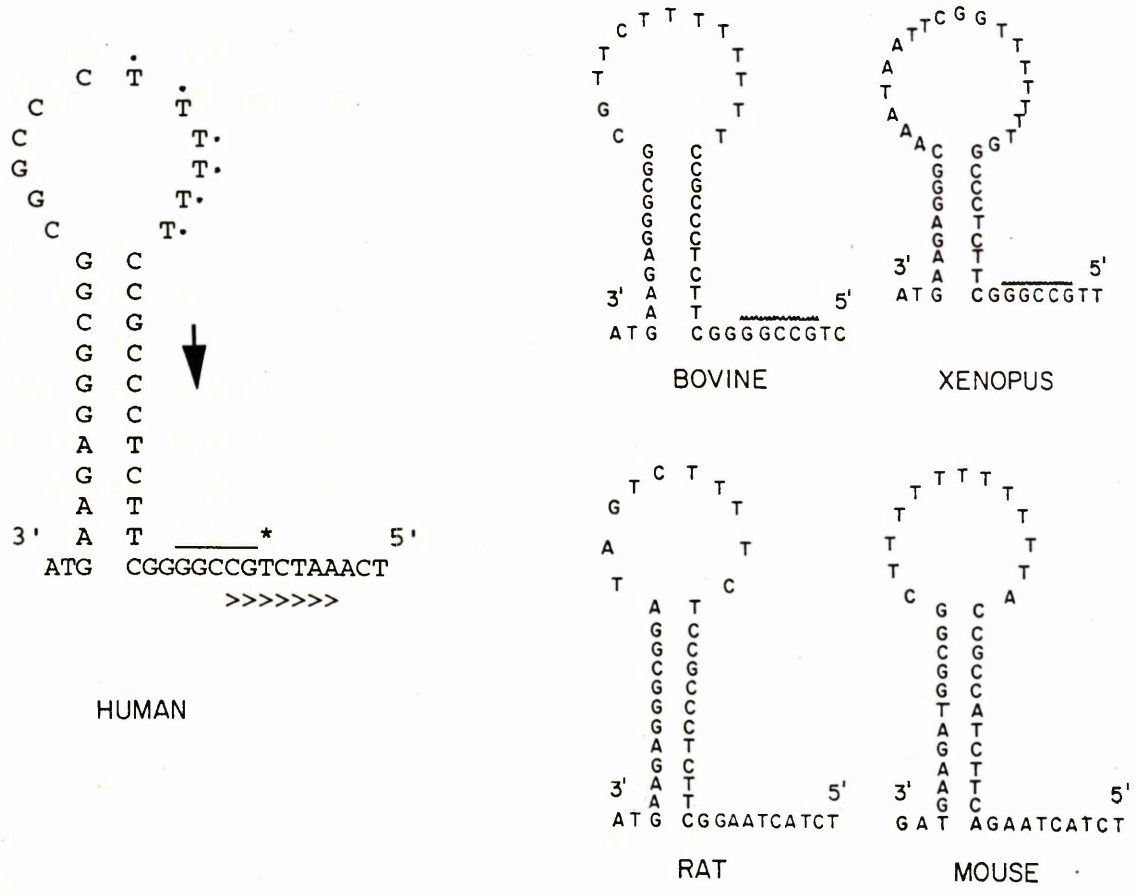


FIGURE 1.5: Sequence, and potential secondary structure of the regions of DNA covering the light-strand replication origins of the human, bovine, *Xenopus*, rat and mouse mitochondrial genomes. On the human origin sequence, the nucleotides at which primer RNA synthesis initiates are indicated by dots, and those at which the RNA to DNA transition occurs are indicated by >. The asterisk indicates the position of the major 5' DNA end, the direction of synthesis is indicated by the filled arrowhead, and the GGCCG sequence conserved between human, bovine and *Xenopus* is overlined. Adapted from Chang *et al*, 1985b; Wong & Clayton, 1985.

distribution of the 2 forms of MRP is similar (Topper & Clayton, 1990b), although the smaller RNA molecule (108 nt) was present in amounts equal to those of the 260 nt RNA in the mitochondria, not in excess as in the mouse. Two possible roles for the RNase MRP in the nucleus have been proposed (Topper & Clayton, 1990b). It could either have a role in pre-ribosomal RNA processing, as suggested by its preferential localisation to the nucleolus (Gold et al, 1989), or it could act in nuclear DNA replication, possibly in the processing of RNA primers.

1.3.5 THE L-STRAND REPLICATION ORIGIN.

The position of the L-strand replication origin (O_L) in mouse and human mtDNA was established by using electron microscopy to analyse the replication intermediates (section 1.3.1). More accurate mapping of O_L in mouse mtDNA was carried out by digesting replicating and non-replicating molecules with restriction enzymes, followed by S1 nuclease to remove any single-stranded DNA present (Martens & Clayton, 1979). The patterns of bands obtained following gel electrophoresis were compared, and both the HincII and the HpaII enzymes used gave an extra band with the replicating mtDNA sample. The position of the L-strand origin relative to the 2 enzyme sites mapped O_L to 67% genome length away from the heavy-strand replication origin (O_H).

The sequencing of a 340 nt region covering O_L revealed a potential stem-loop structure, with a 13 nt T-rich loop, and a G+C-rich stem of 12 nt. The presence of a TaqI site close to O_L allowed the 5′ ends of the L-strands to be mapped more accurately, revealing a considerable degree of heterogeneity in their position (Tapper & Clayton, 1982), though they clustered around one major site. Analysis of the 5′ terminal nucleotides showed that 2 ribonucleotides were present in all the nascent strands, which could reflect inefficient excision of the RNA primer. Ribosubstitution is also commonly observed in closed circular mtDNA in this small region of the genome

(Brennicke & Clayton, 1981), although only in about 1% of all molecules.

Comparisons of the O_L sequences of mouse, human, bovine, rat and Xenopus mtDNA (Chang et al, 1985b) have revealed conservation in the structure of the stem-loop, and its location within a cluster of 5 tRNA genes (Fig. 1.5). The region is not, however, conserved at the DNA sequence level between all 5 species. The base composition of the stem and loop are similar in human, bovine and Xenopus mtDNA, but show considerable differences with the mouse and rat sequences (Fig. 1.5). No ribonucleotides have been detected at the 5′ ends of the nascent L-strands in human mtDNA, indicating that, as at O_H, primer removal is a much more efficient process in human mtDNA than it is in mouse mtDNA. Mapping the 5′ end of the nascent L-strands in human mtDNA molecules also revealed a second minor start site not seen in mouse mtDNA (Tapper & Clayton, 1981). This is 37 $\pm$ 2 nt away from the major 5′ ends, and maps to a small stem-loop structure complementary to the anticodon loop of the $tRNA^{cys}$ gene. As in the mouse L-strands, and the D-strand molecules of both species, there is a degree of microheterogeneity at the 5′ ends of the nascent L-strands in humans.

The development of an in vitro system for the initiation of DNA synthesis at O_L (Wong & Clayton, 1985) has revealed more information about this process in human mtDNA. The system used human mtDNA from the region spanning O_L, cloned into M13, as a single-stranded template, and mitochondrial protein fractions to supply the enzymatic activity. The actual products observed in the reaction were dependent on the amounts of rNTPs present. With all 4 rNTPs added, the RNA to DNA transition occurred between nt 5769-5775, just 3′ to the base of the stem-loop structure (Fig. 1.5), whereas with limiting concentrations of rUTP, rCTP, and rGTP, the transition point was within the stem itself (nt 5761-5763). In both cases, rNTPs were present at the 5′ ends of some of the nascent strands, unlike the situation in vivo (Tapper & Clayton, 1981). This may be due to the low levels of RNase H

present in the enzyme fraction used, resulting in incomplete removal of the RNA primer. A heterologous system was also tried, using the human protein fraction, and either bovine or mouse mtDNA sequences as a template (Wong & Clayton, 1985). The products formed from the bovine mtDNA template were analogous to those seen in the human system, with their 5′ ends at the same relative positions in the origin region. However, the human enzyme did not work on the mouse DNA template. This may reflect the considerable sequence similarity between the bovine and human O_L regions (Fig. 1.5), and the relative divergence of the mouse mtDNA sequence in comparison to the other two.

By using the *in vitro* system for the initiation of DNA replication from O_L, Hixson *et al* (1986) identified the intact stem-loop structure (nt 5730-5763) as being essential for DNA synthesis in human mtDNA, but showed that very little of the flanking sequence was required. Site-directed mutagenesis showed that both C to T, and G to A substitutions within the loop had no effect on DNA synthesis, but that mutations in the 3′-GGCCG-5′ sequence (nt 5766-5770) which lies 3′ to the stem-loop, decreased synthesis by 50-80%.

All except one of the mutants studied had no effect on DNA synthesis *in vitro* using ATP as the sole rNTP. This is consistent with the previous observation that under these conditions the RNA to DNA transition occurs within the stem structure itself (Wong & Clayton, 1985), so the GGCCG sequence no longer has a role in this process. This GGCCG sequence is found in the same position in the bovine and *Xenopus* O_L regions, but not in mouse or rat, which could account for the failure of the human enzyme fraction to initiate synthesis on the mouse template.

The *in vitro* system for the initiation of L-strand DNA synthesis was also used to show that purified L-strand DNA primase from human could prime DNA synthesis in the presence of DNA polymerase , and the absence of mtRNA polymerase (Wong & Clayton, 1985), with the RNA to DNA transition

occurring 3′ to the base of the stem-loop. This clearly demonstrates the differences in the initiation of replication between O_L and O_H, where replication is primed by mitochondrial RNA polymerase.

Isolation of the human mitochondrial L-strand DNA primase (Wong & Clayton, 1986) revealed that its ability to replicate template DNA was sensitive to protease, micrococcal nuclease, and RNase A, but not DNase, indicting that it is a ribonucleoprotein. Two RNA species of 155 nt and 160 nt were isolated from the primase enzyme, and further experiments indicated that the cytosolic 5.8S rRNA molecule could be the RNA component of primase (Wong & Clayton, 1986). Mitochondrial DNA primase, like RNase MRP, therefore provides another example of a nuclear-encoded mitochondrial protein containing an RNA molecule, which is also encoded by a nuclear gene.

1.3.6 DEVELOPMENTAL REGULATION OF mtDNA REPLICATION.

In the somatic cell cycle, the replication of mtDNA and nuclear DNA is controlled such that the amount of mtDNA doubles once per cell cycle (Berk & Clayton, 1974). However, the labelling of HeLa cell mtDNA with 5-bromodeoxyuridine revealed that the control of mtDNA replication is not rigid, in that some molecules do not replicate during a given cell cycle, whereas others replicate twice (Flory & Vinograd, 1973). It is not clear whether the replication of mtDNA molecules is coordinated with a particular phase of the cell cycle, as occurs in some lower eukaryotes (*e.g.* kinetoplastid protozoa). Pica-Mattoccia & Attardi (1972) reported an increased level of mtDNA synthesis during the S and G_2 phases of the cell cycle in HeLa cells, though other studies carried out in mouse L cells have reported a constant level of mtDNA synthesis throughout the cell cycle (Bogenhagen & Clayton, 1977), and studies by Posakony *et al* (1977) found that the growth and division of mitochondria occur throughout the cell cycle in HeLa cells.

The coupling of the mitochondrial and nuclear DNA replication is not, however, maintained in oogenesis and early embryogenesis. In oogenesis, mtDNA replication is prolific (Piko et al, 1967; Matsumoto & Piko, 1971; Webb & Camp, 1979; Webb & Smith, 1977) but no nuclear DNA replication occurs. MtDNA replication is halted in mature eggs, although D-loop synthesis still occurs to varying extents, depending on the organism. In Xenopus laevis at least 80% of the mtDNA molecules have D-loops (Hallberg, 1974), whereas the figure for mouse is around 30% (Piko & Matsumoto, 1976), and D-loop synthesis occurs in sea urchin mtDNA (Jacobs et al, 1989). The block to mtDNA replication is maintained throughout early embryogenesis, when nuclear DNA synthesis is extensive (Bresch, 1974; Chase & Dawid, 1972; Matsumoto & Piko, 1971; Piko, 1969; Piko & Taylor, 1987; Wassarman, 1971). In sea urchins, mtDNA replication does not resume until the pluteus stage (Bresch, 1974; Piko, 1969), and in X. laevis, the levels of nuclear DNA are 10^5 fold greater than those in unfertilised eggs by the time mtDNA replication starts in stage 30-32 embryos (Chase & Dawid, 1972).

Studies into the regulation of nuclear and mitochondrial DNA replication in early development implied that in sea urchins and X. laevis the nucleus was involved in the inhibition of mtDNA replication (Rinaldi & Giudice, 1985). In sea urcin eggs (from P. lividus), the removal of the nucleus allowed parthenogenetic activation of the enucleate egg fragments to stimulate mtDNA synthesis, whereas in whole eggs or nucleated halves, activation did not induce mtDNA replication (Rinaldi et al, 1979a). The same effect was also observed in eggs from X. laevis (Rinaldi et al, 1981).

There are several potential steps in the replication process of animal mtDNA at which regulation could occur. These include the start of RNA primer synthesis; the processing of the RNA molecule to form an active primer; the initiation of DNA synthesis; and the termination of D-strand synthesis, which would generate a D-loop. There is no firm evidence as to which step(s) is controlled either in normal (i.e.

somatic) cell-cycle regulation, or during oogenesis and early embryogenesis, so further discussion of this question is left to the concluding chapter (section 6.2).

1.4 MtDNA REPLICATION IN OTHER ORGANISMS.

1.4.1 MtDNA REPLICATION IN SEA URCHINS.

Relatively little is known about mtDNA replication in sea urchins, and what is known is discussed in detail elsewhere in the thesis in relation to the results obtained in the course of this study, and the replication process in other organisms. This section therefore simply provides a summary of the data.

The EM studies carried out on S. purpuratus mtDNA isolated from oocytes (Matsumoto et al, 1974) suggested a strand-displacement mechanism for replication, similar to that used by vertebrate mtDNA, although differences were observed in the forms of the replication intermediates (section 1.3.1). These differences suggest that multiple second-strand origins are present in the mitochondrial genome. Sequence analysis of the mtDNA of S. purpuratus (Jacobs et al, 1988) and P. lividus (Cantatore et al, 1989) revealed that the major non-coding regions of these genomes were much shorter than those seen in vertebrate mtDNA, and that the mtDNAs did not appear to contain a sequence corresponding to the vertebrate lagging-strand origin. Studies into sea urchin mtDNA replication at different stages of development (section 1.3.6) showed that D-loop synthesis (Jacobs et al, 1989), but not DNA replication (Matsumoto et al, 1974), occurred in mtDNA isolated from eggs. MtDNA replication is also inhibited in embryogenesis, with the nucleus apparently being involved in the inhibition process (Rinaldi et al, 1979a).

1.4.2 MtDNA REPLICATION IN DROSOPHILA.

In Drosophila mtDNA, the leading-strand replication origin is located in a very A+T-rich, non-coding region of the genome. Replication is again by a highly asymmetrical, strand-displacement mechanism, but no molecules have been detected containing a D-loop within the origin region (Goddard & Wolstenholme, 1978). This may either reflect the loss of D-loop strands by branch migration during the purification of mtDNA, or their absence from the replication mechanism of Drosophila mtDNA. The G+C content of the origin region of D. melanogaster mtDNA is less than 5% (Goddard & Wolstenholme, 1978), and that of D. yakuba is less than 8% (Clary & Wolstenholme, 1985), which is in stark contrast to regions of 30% and 41% G+C content covering the start of DNA synthesis in mouse and rat mtDNA repectively. Both rat and mouse genomes also contain CSB II in this region, which is conserved between species, and has over 75% G+C content. In S. purpuratus mtDNA, the overall G+C content of the non-coding region is 52%, and it contains a run of 20 consecutive G residues (Jacobs et al, 1988). It therefore appears that Drosophila mtDNA is highly atypical with regard to the DNA sequences found at the H-strand replication origin.

The Drosophila mitochondrial genome contains no obvious region which could fold to resemble the L-strand origin found in vertebrate mtDNA, so in this respect Drosophila mtDNA resembles that of S. purpuratus. However, electron microscopy studies on replicating mtDNA in D. melanogaster suggest that the L-strand origin is also located within the non-coding region, and that as a result of this, leading-strand synthesis leads to the displacement of over 90% of the genome before L-strand synthesis is initiated (Goddard & Wolstenholme, 1978).

1.4.3 REPLICATION OF KINETOPLAST DNA.

As outlined above (section 1.2.2), kinetoplast DNA (kDNA) in

kinetoplastid protozoa (e.g. T. brucei, L. tarentolae) is made up of 2 different kinds of molecules - minicircles and maxicircles. The replication of kDNA is coordinated with nuclear DNA replication, occurring in the S phase of the cell cycle (Cosgrove & Skeen, 1970; Steinert & Steinert, 1962), and there is also evidence that each minicircle and maxicircle molecule is replicated only once per generation (Hajduk et al, 1984), which does not appear to be the case in animal mtDNA replication (Flory & Vinograd, 1973).

Minicircles are no longer attached to the network when they are replicated, with several hundred being released at any one time (Englund*, 1979). Once replication is complete, they are reattached to the periphery of the network, as either nicked or gapped molecules (Englund*, 1978). The minicircle molecules are closed during the G2 phase of the cell cycle, when the network of molecules splits into 2 halves. The minicircles of each species are heterogeneous, though the extent of the heterogeneity varies between species (Borst, 1991; Simpson, 1987). However, all minicircles contain at least one conserved region of 100-200 nt. In C. fasciculata there 2 conserved regions, and in T. cruzi, 4; in each case they are present in direct repeat, equally spaced around the genome. Within the conserved region are 2 sequences, located 70-95 nt apart, which have been found in all minicircles so far studied. These are the sequences GGGGTTGGTGTA (the universal minicircle sequence or UMS), and ACGCCC (sequences given for the heavy- and the light-strands respectively, as in Ryan et al, 1988).

In T. equiperdum, a 73 nt H-strand DNA molecule is commonly detected on denaturating gel electrophoresis of both progeny and replicating, theta-shaped, molecules (Ryan et al, 1988). The ends of this DNA molecule lie at the 2 conserved sequences common to all minicircle genomes, with the 3′ end within the UMS. It is not yet clear as to whether the replication cycle initiates with L- or H-strand synthesis, though the daughter molecules in which the L-strand was newly synthesised become reattached to the network faster than

* ref listed on p219.

those with a new H-strand.

Analysis of the replication intermediates, by pulse-labelling of the DNA in isolated kinetoplasts from T. equiperdum, revealed that the 2 daughter molecules produced by the replication of each minicircle differ in structure. L-strand synthesis is continuous and is initiated at a UMS. In the newly synthesised molecule a single 6-10 nt gap occurs at the UMS, and the presence of 1 or 2 ribonucleotides at the 5′ end of the L-strand suggests that replication was primed by RNA synthesis (Ntambi et al, 1986). In C. fasciculata, there is a directly repeated sequence of 170 nt, containing a UMS, and replication initiates at either of them with equal frequency (Sheline et al, 1989).

H-strand synthesis initiates from the conserved sequence ACGCCC, which is located 70-95 nt from the UMS. Synthesis is discontinuous, producing 40-90 nt fragments, giving rise to daughter molecules with a gapped H-strand. Repair of the gaps is rapid, and initiated on free minicircle molecules, but it is completed after reattachment to the kDNA network.

In C. fasciculata the conversion of the gapped daughter molecules to closed circular molecules in vitro required the use of DNA Pol I and T4 ligase, and could not be accomplished with T4 polymerase and ligase. This indicates that a 5′ to 3′ exonuclease activity, which is present in Pol I but not in T4 polymerase, was required to process the 5′ terminus of the newly synthesised strand before ligation could occur (Sheline et al, 1989). This final ligation occurs at the same time as the division of the kDNA network, after all the minicircle molecules have replicated, and may therefore represent a feature of the control mechanism of kDNA replication.

Maxicircle replication, in contrast to that of minicircles, occurs by the rolling-circle mechanism of DNA replication (Hajduk et al, 1984). The maxicircle replication intermediates have been visualised by electron microscopy during the S phase of the cell cycle, and often have a

single-stranded tail portion, indicating that the lagging-strand origin is separate from that of the leading-strand. Replication occurs on molecules which are still attached to the network, and proceeds for more than a single genome-length, producing a daughter molecule with repeated terminal sequences. Circularisation of the newly synthesised molecule presumably occurs by recombination between the 2 ends, and attachment to the network can occur either before or after this event. The majority of free maxicircles have been shown to be linear structures, the ends of which appear to be unique, but exhibit a degree of heterogeneity, and reattachment to the network continues throughout G2 phase of the cell cycle (Hajduk et al, 1984). Replicated maxicircles of T. brucei cluster near the centre of the network before division occurs (Hoeijmakers & Weijers, 1980), presumably to ensure an equal partition of the daughter molecules.

The mechanism of separation of the mitochondrial genome between the 2 daughter trypanosomes has recently been shown to be dependent on the segregation of the new and old flagellar basal bodies (Robinson & Gull, 1991), which is mediated by microtubules. By using antibody staining to identify the basal bodies and the flagellae, and DAPI staining to locate the nuclear and kinetoplast DNA, the relationship between basal body and kDNA segregation was analysed in the presence of drugs which inhibited one or the other process. This showed that the inhibition of basal body separation prevented the segregation of the kDNA, but that inhibition of kDNA segregation did not affect basal body movement (Robinson & Gull, 1991).

1.4.3 MtDNA REPLICATION IN YEAST.

The wild-type mitochondrial genome of S. cerevisiae contains 7 very similar copies of an ori sequence, 2 copies of which have been shown to act as origins of mtDNA replication in yeast rho$^-$ mutants (mutants in which the mtDNA contains deletions). Each ori sequence is about 280 nt in length, and

consists of 3 G+C-rich regions (called A, B, and C), which are identical in all 7 *ori* sequences, separated by A+T-rich sequences (Fig. 1.6; de Zamaroczy *et al*, 1984). This gives an overall G+C content of 18.7%, which is approximately the same as that of the wild-type mitochondrial genome as a whole. The *l* sequence resembles A+T-rich spacer DNA found in the rest of the genome, and has only about 70% homology between the different *ori* sequences, but shows a high degree of length conservation. This means that the yeast mtDNA replication origins resemble those of prokaryotes (de Zamaroczy *et al*, 1984) in having highly conserved sequences at either end, separated by an internal region whose size and base composition, but not primary sequence is the important factor. The conserved 16 nt A+T-rich sequence (*r*), adjacent to the C region is a site of transcriptional initiation, and is homologous to other transcriptional promoter sequences located elsewhere in the genome. In 2 of the *ori* sequences (*ori*1 and *ori*2) a partial duplication of the *r* sequence (*r'*) separates *r* and C (Fig. 1.6a), and in 3 further *ori* sequences (4,6 and 7) a G+C-rich sequence (δ) interrupts the *r* region. Another G+C-rich sequence, called β, is also found in *ori*4 and *ori*6, located within the *l* region, and it shows conservation of both its location and sequence (de Zamaroczy *et al*, 1984).

Indirect evidence that the *ori* sequences act as origins of replication in yeast mtDNA has come from the studies of petite mutants (yeast strains which lack functional mitochondria). The overwhelming majority of spontaneous petites studied contain either *ori*1, 2, 3, or 5, with *ori* sequences 4, 6, or 7 very rarely being the sole *ori* sequence present in the deleted genome (Baldacci *et al*, 1984). Crosses of petite strains with wild-type yeast strains have shown that the petite strains containing either *ori*1, 2, 3, or 5 show high levels of transmission of the petite genomes (up to 99%) to the progeny, particularly if the *ori* sequence is present in several copies in tandem repeat. This is presumably due to the replication advantage the petite genomes have over the wild-type mtDNA molecules, which is

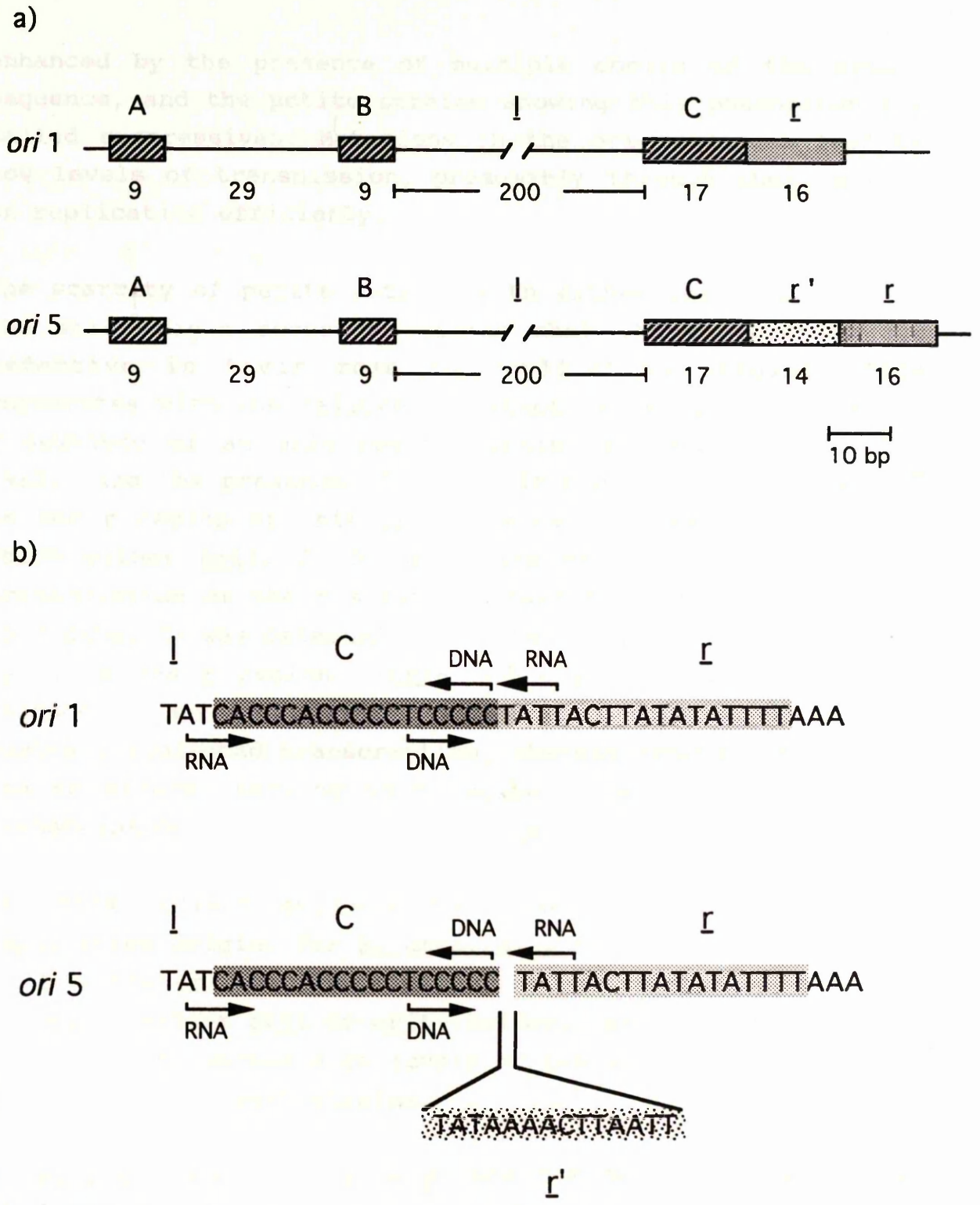


FIGURE 1.6: (a) Organisation of 2 of the origin (*ori*) sequences of *S. cerevisiae* mtDNA, *ori* 1 and *ori* 5. The boxes A, B, and C represent stretches of G+C-rich DNA, and the other regions, stretches of A+T-rich sequence. The lengths of each region (in bp) are given below each box. (b) Sequence of the C and r boxes for the above 2 *ori* sequences, showing the positions of the initiation sites for RNA primer, and DNA synthesis for each strand. The different types of stippled shading indicate the extents of the different sequence boxes. Adapted from Baldacci *et al*, 1984.

enhanced by the presence of multiple copies of the origin sequence, and the petite strains showing this phenomenon are called suppressives. Mutations in the ori sequences lead to low levels of transmission, presumably through their effect on replication efficiency.

The scarcity of petite mutants with either ori4, 6, or 7 as the sole origin sequence implies that these sequences are defective in their role as replication origins. This correlates with the failure to detect transcription from the r sequence of an ori4 petite strain (Baldacci & Bernardi, 1982), and the presence of the additional sequence element Y in the r region of this ori sequence. In petite strains in which either ori1, 2, 3, or 5 was the sole ori sequence, transcription of the r strand (strand containing pyrimidines in cluster C) was detected, with the 5′ end of the transcript lying in the r region (ori6 and ori7 petites not tested; Baldacci & Bernardi, 1982). Mutations in ori1 which removed region C abolished transcription, whereas removal of region A had no effect, showing that region C is also important in transcription.

The first direct evidence that the ori sequences act as replication origins for S. cerevisiae mtDNA came from studies on 2 different petite strains, which contained repeated copies of either ori1 or ori5 (Baldacci et al, 1984). Both of these strains showed high levels of transmission when crossed with wild-type yeast strains.

S1 nuclease mapping, using probes for both strands, showed that DNA synthesis on both strands was primed by RNA, and that the 5′ ends of the DNA strands for the r and non-r strands mapped to the opposite ends of the C sequence (Fig. 1.6b). For both strands the RNA primer is very short compared to those seen in vertebrate mtDNA replication, and the 2 origins are adjacent, so that replication is essentially bidirectional, and appears to be continuous on each strand (i.e. no Okazaki fragments are produced).

Electron microscopy of mtDNA molecules isolated from the yeast Torulopsis glabrata has shown lariat structure replication intermediates, with either single- or double-stranded tail regions (Maleszka et al, 1991). In some molecules the tail portions were longer than one genome length, all of which strongly indicates a rolling circle method of DNA replication. No bubble-form molecules were observed by EM, though it is possible that they were present in small numbers. Y-form molecules were also observed, and were thought to be derived from lariat molecules in which the circle had been cleaved during the preparation procedure for EM. The rolling circle replication mechanism is consistent with the observation of linear mtDNA molecules by pulse-field gel electrophoresis (PFGE) in a number of yeast species (Skelly & Maleszka, 1991). Experiments involving the plasmid pMK2, which contains both mitochondrial and nuclear replication origins for S. cerevisiae, showed that linear molecules were only isolated from the mitochondrion (Maleszka et al, 1991). This strongly suggests that the replication machinery differs in the 2 compartments, and is responsible for rolling circle replication in the mitochondria. Analysis of mtDNA from mouse by PFGE demonstrated that linear DNA molecules were absent, which is consistent with the strand displacement model for mammalian mtDNA replication (Clayton, 1982; section 1.3.1).

Some of the lariat-form replication intermediates of T. glabrata had single-stranded tails, which suggests that second-strand synthesis lags behind that of the leading-strand, and may therefore be the limiting factor in the rate of mtDNA replication. The proposed rolling circle mechanism for yeast mtDNA implies that only one first-strand origin sequence is active per genome, though more than one second-strand origin could be used (Maleszka et al, 1991). It also seems likely, if the T. glabrata model is indicative of the replication scheme in S. cerevisiae, that the second-strand origins are only active when single-stranded.

The rolling circle mechanism for yeast mtDNA replication

could explain the occurrence of an active recombination system in yeast mitochondria, and its virtual absence in the mitochondria of animals. Rolling circle replication needs a recombination mechanism to generate circular molecules from the linear replication products, and aberrant circularisation of the linear mtDNA molecules provides an alternative explanation for the creation of petite mutants observed in yeast.

Definite similarities do, however, exist between the vertebrate and yeast replication systems. Replication of the r strand is primed by an RNA molecule synthesised from a transcriptional promoter (Baldacci et al, 1984), most probably by RNA polymerase, making it analogous to H-strand replication (Clayton, 1982). The differences in sequence at the r and non-r strand primer initiation sites, and the absence of transcription of the non-r strand in petites (Baldacci & Bernardi, 1982), suggest that the priming of non-r strand replication occurs by a different enzymatic system (most probably a primase) as in the replication of the vertebrate L-strand (Wong & Clayton, 1986). Although the wild-type mtDNA of S. cerevisiae contains 7 ori sequences, of which 4 appear to be active (ori1, 2, 3, and 5), it is not clear whether one, some, or all of these sequences are active in the replication of any given mtDNA molecule.

1.5 MITOCHONDRIAL TRANSCRIPTION.

In vertebrate mtDNA, the light-strand promoter (LSP) acts as the promoter for both light-strand transcription and primer formation at the heavy-strand origin of replication (section 1.3.3). Given this connexion between the processes of replication and transcription in the mitochondrion, it seems appropriate to discuss briefly transcription in both vertebrates and yeast mitochondria, and to give a more detailed account of what is known of the process in the sea urchin species S. purpuratus and Paracentrotus lividus.

Mitochondrial RNA polymerase has been purified from humans, Xenopus, and S. cerevisiae. In each case the enzyme consists of 2 subunits; a non-specific core polymerase, and a specificity factor. In humans, the 24 kD transcription factor (mtTF1) directs transcription by binding to a region of the DNA upstream of the promoter sequence (Fisher et al, 1987). In yeast, the specificity factor (MTF1) is a 43 kD protein, which only appears to bind to the DNA template in vitro in the presence of the core polymerase (Schinkel et al, 1988). The specificity factors of yeast and humans do not, however, show any amino acid sequence similarity to each other. The yeast factor (MTF1) has similarities with the eubacterial sigma factors, which interact with core polymerases to modify their promoter specificity (Jang & Jaehning, 1991). In contrast, mtTF1 shows amino acid sequence similarities to the HMG (high mobility group) proteins, and most closely resembles hUBF, the human nucleolar transcription factor (Parisi & Clayton, 1991). A protein (ABF2), with similarity at the amino acid level to the HMG proteins, has also been identified as being important for the expression of the mitochondrial genome in yeast (Diffley & Stillman, 1991). ABF2 binds adjacent to the mitochondrial transcriptional promoter sequences, raising the possibility that it is the yeast equivalent of mtTF1.

Comparisons of the amino acid sequence of the yeast core polymerase, the gene for which (RPO41) has been cloned (Greenleaf et al, 1986), with those of subunits of the multisubunit E. coli RNA polymerase complex revealed no amino acid similarities. However, the yeast mitochondrial RNA polymerase shows considerable amino acid similarity to the RNA polymerases of the bacteriophages T3, T7, and SP6 in the catalytic domain of the enzyme, where the phage polymerases show most similarity to each other (Masters et al, 1987).

The patterns of transcription of the mitochondrial genomes of different species show differences which are, in part at least, related to their differing genome organisation. This is particularly noticeable when the compact animal mtDNA molecules are compared, where the differences in gene order

appear to lead to altered patterns of transcription.

The vertebrate systems of mouse, human and Xenopus have been closely studied, revealing the presence of a single, major promoter for each strand located within the major non-coding region of the genome. A multicistronic transcript is generated from each strand, which is then processed to form the mature RNA molecules. The enzyme RNase P is involved in processing at the 5′ ends of the tRNA genes, and at least one other enzyme must be involved in tRNA 3′ end formation. Studies of the steady state levels of different RNA molecules in mitochondria have shown that the 2 rRNA gene transcripts are present at 10-30 fold higher levels than the mRNAs (Gelfand & Attardi, 1981). In vertebrate mtDNA, the 2 rRNA genes are adjacent, and located next to the non-coding region of the genome. Studies on transcription in HeLa cell mitochondria have revealed that a transcriptional attenuation event occurs just downstream of the 2 rRNA genes (Dubin et al, 1982; Christianson & Clayton, 1988). The detection in vivo of 2 initiation sites for RNA synthesis in HeLa mtDNA led Montoya et al (1983) to propose a model in which the different conformations of the 2 transcripts either permitted or prevented readthrough of the termination site. A protein factor which binds to the region immediately downstream of the 3′ end of the rRNA transcript has been purified, and shown to promote termination of transcription in vitro (Kruse et al, 1989; Hess et al, 1991). However, the detection of 2 distinct transcriptional start sites has not been verified in vitro by other researchers.

The arrangement of genes in the mtDNA of many invertebrates, such as S. purpuratus (Jacobs et al, 1988), P. lividus (Cantatore et al, 1989), and Ascaris suum (Wolstenholme et al, 1987), precludes this mechanism for the maintenance of higher levels of rRNA than mRNA within the mitochondrion, because the 2 rRNA genes are not adjacent. Studies of the pattern of transcription of S. purpuratus mitochondrial RNA, using a combination of primer extension and S1 nuclease mapping have shown that some of the genes overlap slightly

(Elliott & Jacobs, 1989); for example, the 3′ end of 16S rRNA and the 5′ end COI mRNA overlap by 8-10 nt. These transcription studies also mapped the 5′ ends of several mitochondrial RNAs to 3 of the conserved TTATATATAA-like sequences, which occur at 5 locations within the S. purpuratus mitochondrial genome (Fig. 1.7). In vitro capping experiments, using guanylyl transferase, were attempted to determine whether the TTATATATAA-like sequences were the sites of transcriptional initiation, or RNA processing, but proved inconclusive (Elliott, 1990). However, the location of the TTATATATAA-like sequences at the 5′ ends of genes, some of which are divergently transcribed (e.g. ND6 and cyt b) does suggest a potential role as a bidirectional promoter element.

The TTATATATAA-like sequences in S. purpuratus mtDNA resemble the core transcriptional promoter sequences in the mtDNA of S. cerevisiae (ATATAAGTA; Christianson & Rabinowitz, 1983), and N. crassa (TTAGARRA(T/G)R(T/G)A; Kennel & Lambowitz, 1989). However, the conservation of promoter sequences between different organisms is not universal, with the mitochondrial promoters of vertebrates (Chang & Clayton, 1984) not resembling the A+T-rich sequences of fungi. These differences therefore mean that the similarity between the S. purpuratus and fungal sequences may not have any significance for the function of the S. purpuratus TTATATATAA-like motif.

Ultra-violet (UV) irradiation was used to investigate the extent of the transcription units within the S. purpuratus mitochondrial genome (Elliott, 1990). The results are consistent with the idea that the adjacent and overlapping 16S rRNA and COI genes are not transcribed from the same promoter. Treatment of the mitochondria prior to in organello transcription with low doses of UV irradiation (which generates pyrimidine dimers that block transcription) reduced the levels of the COI transcript, but much higher doses were needed to reduce transcription of the 16S rRNA gene. This is consistent with a model in which the promoter for the 16S rRNA gene is located very close to the 5′ end of the gene,

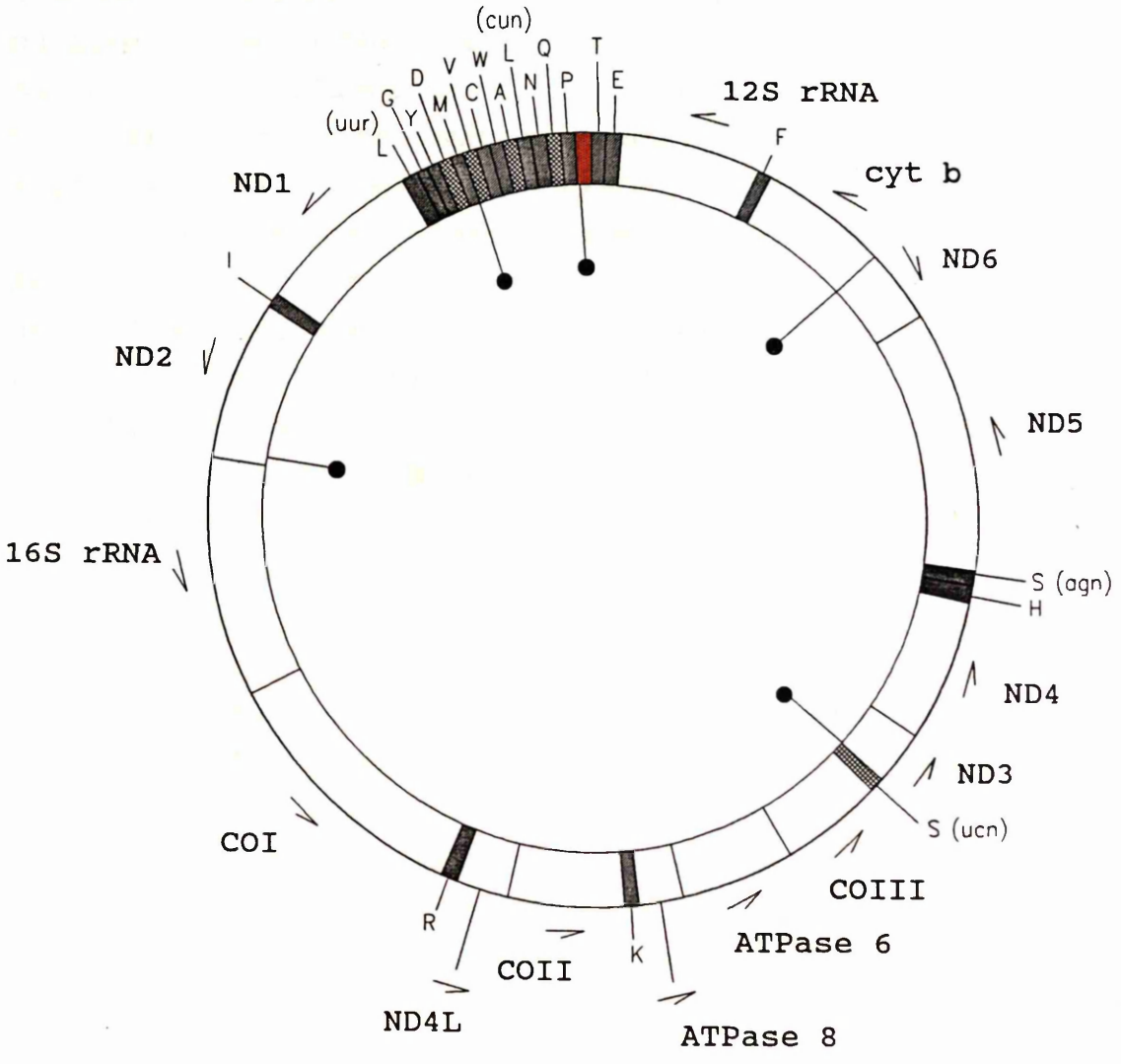


FIGURE 1.7: Organisation of the mitochondrial genome of *Strongylocentrotus purpuratus*. The tRNAs genes are designated by the one letter amino acid code; the large and small rRNA genes are designated 16S and 12S repectively; and the protein-coding genes are abbreviated as follows: COI-III - subunits I-III of cytochrome *c* oxidase; ND1-6, ND4L - subunits 1-6 and 4L of NADH dehydrogenase; A6, A8 - subunits 6 and 8 of ATP synthase; cyt *b* - cytochrome *b*. The direction of transcription of the protein and rRNA genes is shown by the arrows. The non-coding region, within the tRNA cluster, is highlighted in red. Filled circles mark the locations of the TTATATATAA-like sequences (putative transcriptional promoters). From Jacobs *et al*, 1988.

whereas that for the COI gene is far upstream of the start of the gene.

Studies on mtRNA from a different sea urchin species Paracentrotus lividus (Cantatore et al, 1990) revealed a similar transcriptional organisation to that in S. purpuratus, with the transcripts of some genes overlapping. A number of overlapping, high molecular weight RNAs, which together spanned the whole genome, were also detected, and were taken to represent precursor transcripts for the mature RNAs. This suggests the existence of a complex processing mechanism for the generation of mature RNA molecules, probably involving the cleavage of the precursors at the tRNAs or tRNA-like sequences by RNase P-like nucleases; i.e. by the mechanism proposed for RNA maturation in the mitochondria of humans (Ojala et al, 1981), N. crassa (Breitenberger et al, 1985), and D. yakuba (Berthier et al, 1986).

Cantatore et al (1990) favour a model for the generation of the mature RNA molecules from overlapping genes based on the termination of transcription of the upstream genes at the 3′ ends of those genes, with distinct overlapping transcripts existing for the generation of the downstream RNAs. This is based on the mapping of multiple 3′ ends for the 16S rRNA, which are taken to be characteristic of attenuation (on the basis of the mapping of the 3′ end of 16S rRNA in HeLa cells; Dubin et al, 1982), and the observation that no transcripts cross the ND5/cyt b gene boundary. However, the latter 2 genes are separated by ND6, which is transcribed from the opposite DNA strand from ND5 and cyt b. The boundary between ND6 and cyt b would therefore be a potential site for a bidirectional promoter sequence, and the 3′ end of the ND5 gene a likely transcription attenuation site, to prevent the formation of ND6 antisense RNA. The differential processing of different transcripts covering the overlapping genes is therefore a possibility (as proposed for S. purpuratus), although attenuation may be involved in the generation of the 3′ ends of 12S and 16S rRNA, and ND5 and ND6 mRNA.

* ref listed on p219.

The detection of large primary transcripts from the S. purpuratus (Elliott & Jacobs, 1989) and P. lividus (Cantatore et al, 1990) mitochondrial genomes suggests that the processing of these RNAs into functional transcripts occurs after transcription, and not concomitantly, as is the case in mouse. A possible reason for this is the occurrence of the majority (15/22) of the tRNA genes around the non-coding region of the genome. This clustering of the tRNA genes means that they are not located between the genes (as in vertebrate mtDNA), so the tRNAs cannot act as processing signals for cleavage of the polycistronic transcripts into mature mRNAs.

In vitro capping experiments on Saccharomyces cerevisiae mtRNA (Christianson & Rabinowitz, 1983) revealed 20 initiation sites for RNA synthesis within the 75 kb mitochondrial genome. The locations of these sites were confirmed by in vitro transcription analysis using partially purified mitochondrial RNA polymerase. Comparisons of the DNA sequences at the initiation sites produced a nonanucleotide consensus sequence (ATATAAGTA), although other sequences are probably involved in transcription initiation, since the 5′ ends of transcripts have not been detected near all copies of these nonanucleotide sequences present in the mitochondrial genome (Christianson & Rabinowitz, 1983). Competition assays between promoter regions for limiting amounts of mtRNA polymerase in vitro (Mueller & Getz, 1986) have demonstrated that strong and weak promoters exist, with up to a 20-fold difference in strength.

Transcription in yeast mtDNA produces multicistronic transcripts, which are then processed to form the mature RNAs. S1 nuclease and primer extension analysis of the 3′ and 5′ ends of RNA molecules from the polycistronic transcript covering the COXI, A6 and A8 genes showed that processing occurred at a dodecamer sequence (AAUAAUAUUCUU; Osinga et al, 1984). This sequence has been found at the 3′ end of nearly all the protein-coding genes present in the yeast mitochondrial genome, including that for the FIT1 endonuclease encoded by the omega intron (Zhu et al, 1987).

It is not, however, clear whether this sequence also acts as an attenuator, terminating transcription, and therefore resembles the mammalian mtDNA sequence located downstream of the rRNA genes. Pulse-labelling experiments have shown that transcription attenuation does occur in some transcription units (Mueller & Getz, 1986), with lower levels of RNA synthesis from the promoter-distal genes.

The combination of different strength promoters, and transcriptional attenuation allows the levels of the mRNAs for different genes to differ by up to 50-fold (Attardi & Schatz, 1988); a very different situation from that observed for vertebrate mitochondrial mRNAs. The transcriptional organisation of the mtDNA of the filamentous fungi Aspergillus (Dyson et al, 1989), and Neurospora (Kennel & Lambowitz, 1989) appears to be similar to that of S. cervisiae, with multiple, polycistronic transcription units detected.

1.6 AIMS OF THE PROJECT.

In oogenesis and embryogenesis, the control of mtDNA replication, which ensures that the amount of mtDNA doubles once per cell division, is altered. In oogenesis, there is extensive mtDNA replication, but no nuclear DNA replication, whereas the reverse is true in embryogenesis (section 1.3.6). The question therefore arises as to what controls mtDNA replication, and how this control mechanism is altered in development. Sea urchins provide a suitable organism for this study, because large numbers of mitochondria can be easily isolated from eggs, and large, essentially synchronous, embryo cultures can be set up in vitro.

This study was carried out in order to provide basic information about the method of mtDNA replication in S. purpuratus. These preliminary data are essential to enable further reasearch to be carried out into the developmental regulation of this replication system. The initial aims of

the project were as follows:

1. To determine the location of the replication origin(s) in S. purpuratus mtDNA.
2. To map as accurately as possible the positions of the 5′ ends of the nascent DNA strands.
3. To determine whether DNA synthesis is bidirectional or unidirectional, and if the latter, the direction of replication fork movement.
4. To determine whether DNA synthesis is by a semidiscontinuous, or a continuous (strand displacement) mechanism.

The experiments showed that leading-strand DNA synthesis paused at several sites in the genome, and the most prominent of these occurred near the A6/COIII gene boundary, which was also the site of a prominent lagging-strand replication origin. I therefore decided to study this region of the genome in more detail, to assess whether the replication pause site could also have a role in transcription. To accomplish this I set out to:

5. Accurately map the 3′ and 5′ ends of the A6 and COIII transcripts in vivo.
6. Determine whether the region could terminate transcription in vitro.
7. Determine whether mitochondrial proteins could affect in vitro transcription termination in this region.

I decided to use the recently developed technique of 2D agarose gel electrophoresis (Brewer & Fangman, 1987) to locate the replication origin(s) in S. purpuratus mtDNA. This technique maps the position of a replication origin with respect to restriction sites within the molecule, on the basis of the structure of the replication intermediates. A considerable amount of information about the replication mechanism can also be deduced from the structures of the replication intermediates, making 2D agarose gel electrophoresis an attractive method to use in this study.

RNase protection experiments were carried out to map the ends of DNA and RNA molecules in the regions of the genome containing the leading-strand origin, and the A6/COIII gene junction. A chimeric in vitro transcription system (using SP6 RNA polymerase) was also set up using this pause site region, to assess its possible effects on transcription.

CHAPTER

TWO

MATERIALS AND METHODS

2.1 BACTERIAL STRAINS.

The only bacterial strain used in this study was DS941, a derivative of *Escherichia coli* K-12. Its genotype is given below (D.J. Sherratt, pers. comm.):

DS941 *recF143*, *thr-1*, *leuB6*, *thi-1*, *lacY1*, *tsx-33*, *supE44*, *galK2*, *ara-14*, *xyl-5*, *mtl-1*, Δ(*gpt-proA*)62, *hisG4*, *argE3*, *rpsL31*, *lacZ*ΔM15, *lacI*^{q}.

2.2 PLASMIDS AND BACTERIOPHAGES.

The recombinant plasmid and bacteriophage clones used and constructed in this study, are described in Table 2.1. This table also gives details of the plasmid cloning vectors used.

2.3 CHEMICALS.

Chemicals	Source
General chemicals and solvents (analytical grade unless stated)	BDH, Hopkins and Williams, Kochlight Laboratories, May and Baker, Fluka.
Microbiological media	Difco, Oxford.
Biochemicals	Sigma, Pharmacia, BRL.
Agaroses, acrylamide	BRL.
Radiochemicals	NEN (Dupont).
Antibiotics	Sigma.
Restriction enzymes	BRL, Boehringer Mannheim, NBL.
DNA modifying enzymes	BRL, Boehringer Mannheim, NEN (Dupont), Pharmacia, NBL.
RNA polymerases	Boehringer Mannheim, Promega.

PLASMID VECTORS

Vector	Derived from....	Reference
pSP64	pUC12	Melton *et al*, 1984
pSP65	pUC12	Melton *et al*, 1984
pTZ19R	pUC19	Vieira & Messing, 1987
pUC8	pBR322	Vieira & Messing, 1982
pUC18	pBR322	Yanisch-Perron *et al*, 1985
pUC19	pBR322	Yanisch-Perron *et al*, 1985

PLASMID CLONES

Name	Vector	mtDNA insert	Genes	Reference
pPZ0.6	pUC19	*Pst*I (254)-*Sst*I (840)	part 12S rRNA	Jacobs *et al*, 1988
pR3	pUC8	*Eco*RI (8060)-*Eco*RI(9778)	3' COII/A8/A6/5' COIII	Jacobs *et al*, 1988
pRB3	pUC8	*Bgl*II(13250)-*Eco*RI (14547)	3' ND5/ND6/5' cyt *b*	Jacobs *et al*, 1988
pZH0.8	pUC19	*Sst*I (840)-*Hin*dIII (1603)	3' 12S/NCR/tRNAs (E,T,P,Q,N,L,A,W)	H.T.Jacobs, pers.comm.
pZH1.1	pUC19	*Hin*dIII (11618)-*Sst* I(12676)	3' ND4/5'ND5	Jacobs *et al*, 1988
pZH2.0	pUC19	*Hin*dIII (5622)-*Sst* I (7768)	3' 16S/COI/ND4L/5' COII	Jacobs *et al*, 1988
pSP3	pSP64	*Hin*dIII (8424)-*Eco*RI (9778)	A8/A6/5' COIII >>>	Chapter 5
pSP31	pSP64	*Hin*dIII (8424)-*Eco*RI (9778)	A8/A6/5' COIII <<<	Chapter 5
pSPV3	pSP64	*Pvu*II (9148)-*Eco*RI (9778)	3' A6/5' COIII >>>	Chapter 4
pSPV31	pSP64	*Pvu*II (9148)-*Eco*RI (9778)	3' A6/5' COIII <<<	Chapter 4
*pAM35	pSP64	exoIII deletion series derived from pSPV31	3' A6/5' COIII - deleted from COIII end of insert <<<	Chapter 4

*pAM6-35	pSP65	members of pAM35 deletion series cloned into pSP65	3' A6/5' COIII - deleted from COIII end of insert >>>	Chapter 4
*pAMV1	pSP64	exoIII deletion series derived from pSPV3	3' A6/5' COIII - deleted from A6 end of insert >>>	Chapter 5
pTZV3	pTZ19R	*Pvu*II (9148)-*Eco*RI (9778)	3' A6/5' COIII >>>	Chapter 5
pTZV31	pTZ19R	*Pvu*II (9148)-*Eco*RI (9778)	3' A6/5' COIII <<<	Chapter 5
pZH64	pSP64	*Sst*I (840)-*Hin*dIII (1603)	3' 12S/NCR/tRNAs (E,T,P,Q,N,L,A,W) <<<	Chapter 4
pZH65	pSP65	*Sst*I (840)-*Hin*dIII (1603)	3' 12S/NCR/tRNAs (E,T,P,Q,N,L,A,W) >>>	Chapter 4

BACTERIOPHAGE CLONES

A45	M13mp19	*Sst*I(10306)-*Hin*dIII(11617)	3' ND3/5' ND4 (**C**)	H.T.Jacobs, pers.comm.
N18	M13mp9	*Xho*I(14048)-*Eco*RI(14547)	5' ND6/5' cyt *b* (**NC**, cyt *b* / **C**, ND6)	Elliott & Jacobs, 1989
SP16	M13mp18	*Sal*I (15020)-*Pst*I(254)	3' cyt *b* /5' 12S (**C**)	H.T.Jacobs, pers.comm.

TABLE 2.1: The recombinant plasmid and bacteriophage clones, and the plasmid vectors used in this study. The coordinates of the restriction enzyme sites are given according to Jacobs *et al*, 1988, and the abbreviations of the gene names are given in the legend to Figure 1.7. For the plasmid clones containing the SP6 or T7 promoter, the direction of transcription of the mitchondrial insert is indicated by >>> or <<<, which represent transcription of the major- and minor-coding strands, respectively. The plasmid clones marked by an asterisk (*) are, in each case, the first members of an exonuclease III deletion series. Further details of the other clones in each series, and the extents of the deletions of the mtDNA insert are given in the text. For the bacteriophage clones, the strand synthesised by primer extension on single-stranded M13 templates is indicated: **NC** stands for the non-coding strand, and **C**, the coding-strand.

2.4 MICROBIOLOGICAL CULTURE MEDIA.

2.4.1 MEDIA.

L-broth: 10 g bacto-tryptone, 5 g yeast extract, 5 g NaCl, 1 g glucose, and 20 mg thymine were made up to 1 litre in distilled water and adjusted to pH 7.0 with sodium hydroxide.

L-agar: As L-broth, without glucose plus 15 g/l agar.

Top agar: 2 parts of L-broth and 1 part of L-agar, mixed under sterile conditions before use.

2.4.2 ANTIBIOTICS.

Ampicillin was the only antibiotic used for bacterial work in this study. It was stored at -20°C as a 50 mg/ml solution in distilled water, and used at a final concentration of 50 ug/ml.

2.4.3 STERILISATION.

All growth media and glassware were sterilised by autoclaving at 120°C for 15 minutes, supplements and buffer solutions at 108°C for 10 minutes, and 100 mM $CaCl_2$ at 114°C for 10 minutes. Heat sensitive reagents were sterilised by filtration using disposable 0.45 um membrane filters. Plasticware was sterilised by autoclaving at 120°C for 9 minutes.

2.5 BUFFER SOLUTIONS.

2.5.1 ELECTROPHORESIS BUFFERS.

50 x TAE buffer: 242 g Tris-base, 57.1 ml glacial acetic acid, 100 ml 0.5 M EDTA pH8.0, made up to 1 litre in

sterile distilled water. (0.4M Tris acetate, 0.05M EDTA, pH 8.3)

10 x TBE buffer: 109 g Tris-base, 55 g boric acid, 40 ml 0.5 M EDTA pH 8.0, made up to 1 litre in sterile distilled water. (0.4 M Tris borate, 0.2 M EDTA, pH 8.3)

Sequencing gel loading buffer (1x): 90% de-ionised formamide; 20 mM EDTA; 0.05% (w/v) each of xylene cyanol and bromophenol blue.

Agarose gel loading buffer (10x): 0.25% (w/v) each of bromophenol blue, xylene cyanol, and orange G; 25% (w/v) Ficoll 400.

2.5.2 ENZYME BUFFERS.

BRL REact restriction buffers were generally used for restriction digests, as follows:

10 x REact 1: 500 mM Tris.HCl pH 8.0, 100 mM $MgCl_2$.

10 x REact 2: 500 mM Tris.HCl pH 8.0, 100 mM $MgCl_2$, 500 mM NaCl.

10 x REact 3: 500 mM Tris.HCl pH 8.0, 100 mM $MgCl_2$, 1M NaCl.

10 x REact 4: 200 mM Tris.HCl pH 7.4, 100 mM $MgCl_2$, 500 mM KCl.

10 x Ligation buffer: 660 mM Tris.HCl pH 7.6, 50 mM $MgCl_2$, 50 mM DTT, 10 mM ATP. Stored at -20°C.

10 x Nick-translation buffer: 500 mM Tris.HCl pH 7.2, 100 mM $MgSO_4$, 1 mM DTT, 500 ug/ml BSA (enzyme grade). Stored at -20°C.

10 x HIN buffer: 500 mM NaCl, 70 mM Tris.HCl pH 7.5, 100 mM $MgCl_2$, 30 mM DTT. Stored at -20°C.

5 x SP6 transcription buffer: 200 mM Tris.HCl pH 7.5, 30 mM $MgCl_2$, 25 mM NaCl, 10 mM spermidine. Autoclave, then add DTT to a final concentration of 50 mM.

2.5.3 GENERAL BUFFERS.

20 x SSC : 3 M NaCl, 300 mM tri-sodium citrate.

20 x SET : 400 mM Tris/HCl pH 7.8, 3 M NaCl, 20 mM EDTA.

TE : 10 mM Tris/HCl pH 8.0, 1mM EDTA.

Phosphate buffer: Made by mixing 1 M Na_2HPO_4 with 1M NaH_2PO_4 until the pH was 6.8.

EDTA: 0.5 M solution was made up, and adjusted to pH 8.0 with NaOH.

50 x Denhardt's solution: 5 g Ficoll, 5 g polyvinyl-pyrrolidone, 5 g BSA, made up to 500 ml in dH_2O, filtered and stored in aliquots at -20°C.

2.6 ORGANIC SOLVENTS.

Phenol: Phenol was equilibrated several times with 1 M Tris/HCl pH 8.0, and then with 0.1 M Tris/HCl pH 8.0. 8-hydroxyquinoline was added to 0.1% (w/v), and aliquots were stored in the dark under 0.1 M Tris/HCl pH 8.0, at -20°C.

Sevag: 24:1 (v/v) mixture of chloroform and isoamyl alcohol. Stored at room temperature.

Phenol/sevag: 1:1 mixture of phenol and sevag. Stored in the dark, at 4°C.

Formamide: Analytical grade formamide was stored at 4°C.

Other solvents: Ethanol, butanol, and isopropanol were stored at room temperature.

2.7 BACTERIAL CULTURES.

E. coli DS941 cultures were grown up from single colonies on plates, or from glycerol frozen stocks, in L-broth overnight at 37°C, with vigorous shaking. Ampicillin was added to a final concentration of 50 ug/ml for the maintenance of plasmids, if appropriate.

2.8 BACTERIAL TRANSFORMATION.

2.8.1 PREPARATION OF COMPETENT CELLS.

0.5 ml of an overnight culture of bacteria was added to 50 ml L-broth, and this was then incubated for 90 min at 37°C, with shaking. The cells were then spun at 4000 rpm for 5 min at 4°C in a bench-top centrifuge, the supernatant removed, and the resulting pellet resuspended in 25 ml ice-cold 50 mM $CaCl_2$. After a 20 min incubation on ice, the cells were respun as above, and the pellet resuspended in 2.5 ml ice-cold 50 mM $CaCl_2$. The resulting competent bacteria were kept at 4°C, and used within 2 days.

2.8.2 TRANSFORMATION.

1-5 ul of the DNA solution (10-50 ng) was added to 100 ul competent cells and left on ice for 40 min, before the mixture was subjected to a heat-shock at 43°C for 90 sec. These transformation reactions were then plated out on L-agar plates containing 50 ug/ml ampicillin, and incubated overnight at 37°C to select for transformants.

2.9 PREPARATION OF PLASMID DNA.

The isolation of covalently closed circle plasmid DNA was carried out both on a small scale, to characterise transformant clones, and on a large scale, for subcloning and the production of radioactive probes.

2.9.1 SMALL SCALE (STET) PREPARATION.

STET buffer - 8% (w/v) sucrose; 5% (v/v) Triton X-100; 50mM Tris.HCl pH 8.0; 1 mM EDTA.

1.5 ml of an overnight bacterial culture were spun in a microfuge for 20 sec, and the supernatant discarded. The pellet was resuspended in 350 ul STET buffer, 25 ul of freshly prepared lysozyme (10 mg/ml) was added, and the solutions mixed by vortexing. Next the tube was put into boiling water for 40 sec, then immediately transferred to ice for 1 min, before being spun in a microfuge for 15 min. The white pellet was removed using a toothpick, and 400 ul isopropanol added to the remaining solution. This was mixed, and the DNA left to precipitate at RT for 10 min. Again the tube was spun for 15 min, the supernatant discarded, and the pellet washed with 200 ul 70% EtOH. This was discarded, and the pellet dried, before being resuspended in 50 ul TE. 5 ul of the DNA solution was used for each restriction enzyme digestion.

2.9.2 LARGE SCALE (BIRNBOIM AND DOLY) PREPARATION.

GET - 50 mM glucose; 25 mM Tris.HCl pH 8.0; 10 mM EDTA.
Soln II - 0.2 M NaOH; 1% SDS.
KAc - 300 ml 5.0 M KAc; 57.5 ml glacial acetic acid; 124.5 ml dH_2O.

A 100 ml overnight bacterial culture was spun in a 250 ml centrifuge bottle at 7.5 krpm in a JA14 rotor for 10 min at

4°C, and the supernatant discarded. The pellet was resuspended in 4 ml ice-cold GET, transferred to a 35 ml Oak Ridge tube, and incubated on ice for 5 min. 8 ml freshly-made solution II was added, and the resulting solution mixed gently by inversion. After a further 5 min incubation on ice, 6 ml ice-cold KAc solution was added, and the tube gently inverted to mix the contents. Centrifugation of the tube was then carried out at 18 krpm for 20 min, at 4°C, in a JA20 rotor with no brake applied to stop the spin. The supernatant was poured through a glasswool-plugged funnel into another Oak Ridge tube containing 12 ml isopropanol, the solutions mixed, and the DNA left to precipitate at RT for 15 min. The nucleic acid was then spun down by centrifugation at 15 krpm for 10 min at 20°C. The pellet was washed with 70% EtOH, and dried in vacuo, before being resuspended in 7 ml TE. 7 g solid CsCl and 200 ul (10 mg/ml) EtBr were added, and the density of the solution adjusted to between 1.57 and 1.60 g/ml. The solution was subsequently loaded into Beckman Ti70 ultracentrifuge tubes, and the tubes filled with CsCl solution (made in TE; density 1.60 g/ml), then balanced and sealed. The tubes were spun at 49 krpm in a Ti70 rotor at 22°C for a minimum of 16 hrs.

After centrifugation the tubes were viewed under long-wave UV light to visualise the plasmid DNA band (the lower prominent band of the two), which was then recovered using a 5 ml syringe, attached to an 18 gauge needle. The EtBr was removed by 3 extractions with isopropanol (section 2.11.2), and the remaining DNA-containing solution was dialysed overnight against at least 1000 volumes TE, with 2 changes of buffer.

2.10 DNA MANIPULATIONS.

2.10.1 RESTRICTION ENZYME DIGESTS.

DNA was digested in the appropriate BRL REact buffer for the enzyme, using at least 2 U enzyme per ug DNA. Incubations were all carried out at 37°C, for either 1 hr (plasmid DNA)

or 3 hrs (mtDNA).

2.10.2 DNA LIGATIONS.

DNA ligations were carried out in a final volume of 10 ul, using 100-500 ng DNA, with a 3-fold molar excess of insert DNA to vector. 1 ul 10 x ligation buffer, and 5 U T4 DNA ligase were added to the DNA, and the volume made up to 10 ul with dH_2O. The reaction was incubated at room temperature (RT) for 4-16 hrs.

2.10.3 NUCLEIC ACID PRECIPITATION.

The precipitation of nucleic acids from solution was either performed by the addition of $1/10^{th}$ volume 3 M NaAc, pH 6.0 and 2.5 volumes absolute EtOH, and incubation at $-20^{o}C$ for 1 hr, or the addition of an equal volume of isopropanol, and incubation at RT for 15 min. The precipitate was spun down either in a JA-20 rotor at 15 krpm, or a microfuge, for 15 min. The supernatant was then discarded, and 70% EtOH added to the pellet, which was vortexed briefly. The tube was then respun for 5 min as above, and the supernatant again discarded, before the pellet was dried *in vacuo*.

2.10.4 DENATURATION OF DNA.

In a few experiments, the DNA samples were denatured before agarose gel electrophoresis. Denaturation was achieved by the addition of $1/15^{th}$ volume of a 3:1 (v/v) 5 M NaOH : 0.5 M EDTA solution, followed by incubation at $70^{o}C$ for 5 min, and immediate transfer to ice. Agarose gel loading buffer was then added to the sample before electrophoresis.

2.11 EXTRACTIONS USING ORGANIC SOLVENTS.

2.11.1 PHENOL/SEVAG.

Proteins were removed from solutions containing nucleic acid by successive extractions with phenol/sevag and sevag (section 2.6). An equal volume of phenol/sevag was added to the solution, which was then mixed by shaking for 1 min, and spun either in a microfuge, or at 4 krpm in a bench top centrifuge, for 5 min. The upper phase containing the nucleic acid was then transferred to another tube, avoiding the white proteinaceous material at the interface. This procedure was repeated until the interface was clear, and then a final extraction was carried out in the same way using an equal volume of sevag.

2.11.2 ISOPROPANOL.

Isopropanol was used to remove the EtBr from plasmid DNA purified on CsCl/EtBr gradients (section 2.10.2). An equal volume of isopropanol was added to the DNA solution, and the 2 phases mixed by shaking. The phases were then left to separate, and the upper, pink layer removed and discarded. This procedure was repeated twice more, with the original volume of the DNA solution being restored, by the addition of dH_2O, before the final extraction to avoid precipitation of the CsCl.

2.12 CONSTRUCTION OF AN EXONUCLEASE III DELETION SERIES OF A PLASMID RECOMBINANT.

10 x exoIII buffer - 660 mM Tris.HCl pH 8.0; 6.6mM $MgCl_2$.
ExoIII stop buffer - 200 mM NaCl; 5 mM EDTA.
Mung bean nuclease premix - 50 ul 10 x MBN buffer (Stratagene); 200 ul dH_2O; 0.5 ul (10 U) mung bean nuclease.
MBN stop buffer - 500 mM Tris.HCl pH 8.0; 125 mM EDTA.

The plasmid DNA was digested with the appropriate restriction enzymes required to ensure the action of exoIII only into the insert region of the plasmid (i.e. leaving the plasmid DNA molecules with one 5′ protruding end, which will be attacked, and one 3′ protruding end, which will not). The DNA was then EtOH precipitated, and resuspended in 30 ul exoIII buffer per deletion time-point. 1-2 ug DNA was used per deletion time point. 20 U exoIII were then added per ug of DNA, and 30 ul aliquots removed at each time point, and transferred into 90 ul exoIII stop buffer. Next the samples were incubated at 65°C for 10 min, and then the DNA was precipitated using EtOH. The dried pellets were each resuspended in 15 ul dH_2O, and 15 ul mung bean nuclease premix added to each sample. They were then incubated at 37°C for 30 min, before 4 ul MBN stop buffer was added to each. Aliquots of 4 ul were removed for gel electrophoresis, and the rest of each sample was stored at 4°C. If the extent of the exoIII deletion was sufficient, then a phenol/sevag extraction was performed on the samples, and the DNA was precipitated. Klenow DNA polymerase was then used to fill in any overhanging 5′ ends (section 2.13) that had been left by the mung bean nuclease, before the DNA was ligated and transformed into E. coli.

2.13 KLENOW DNA POLYMERASE TREATMENT OF DNA TO PRODUCE FLUSH ENDS.

The DNA was resuspended in a solution containing 1.5 ul 10 x REact 1, 2 ul of each of 20 mM dATP, dCTP, dGTP, and dTTP, 3.5 ul dH_2O, and 1 ul Klenow DNA polymerase. This reaction mix was incubated at RT for 15 min, then 70°C for 15 min, before undergoing 1 phenol/sevag and 1 sevag extraction, followed by EtOH precipitation.

2.14 GEL ELECTROPHORESIS.

2.14.1 ONE-DIMENSIONAL AGAROSE GEL ELECTROPHORESIS.

DNA was run on agarose gels, which were made and run in 1 x TAE buffer. Minigels (50 ml; 10 x 6.5 cm) and maxigels (300 ml; 24 x 16 cm) were used. The minigels were usually run at 11 V/cm for 30-45 min; the maxigels were run at between 1 and 4 V/cm, for up to 20 hrs.

2.14.2 TWO-DIMENSIONAL AGAROSE GEL ELECTROPHORESIS.

This protocol was based on that of Brewer & Fangman (1987). DNA samples digested with restriction enzymes were run in a 0.4% agarose/TAE maxigel at 1 V/cm for 16-20 hrs. The gel was stained with EtBr and photographed (section 2.15.4). The relevant lane was then excised from the gel, placed on the maxigel electrophoresis apparatus at right angles to the direction of DNA migration, and a 1.0% agarose/TAE gel poured round it. The gel was subsequently run at 4 V/cm for 3 hrs.

2.14.3 POLYACRYLAMIDE/UREA GELS.

RNA samples from RNase protection, and *in vitro* transcription experiments were analysed on 3.5% polyacrylamide gels containing 8 M urea. The gels were prepared using 10 ml 5 x TBE, 5.8 ml 30% acrylamide (29:1 acrylamide : bis-acrylamide) solution, and 25 g urea, in a final volume of 50 ml. After heating gently to dissolve the urea, dH_2O was added to make the solution up to volume. 330 ul 10% (w/v) ammonium persulphate, and 25 ul TEMED were added, and the 10 x 10 cm, 35 ml gels poured in an LKB 2001 vertical gel electrophoresis apparatus. The gels were pre-run at 250 V/30 mA for 15 min before loading the samples, and run until the bromophenol blue loading dye reached the bottom of the gel. The gels were dried for 1 hr at 80^{o}C on a Bio-Rad gel drier, before auto-radiography.

2.14.4 VISUALISATION OF DNA ON AGAROSE GELS.

The DNA was visualised by soaking the agarose gels in EtBr solution (final concentration 1 ug/ml) for 15 min, then destaining in dH_2O for 20 min. The gels were then viewed under short wave (254 nm) or long wave (302 nm) UV illumination, and photographed using either a Polaroid camera, loaded with type 667 land film, or a Pentax 35 mm SLR camera, loaded with Ilford HP5 film. In both cases the camera was fitted with a Kodak Wratten filter no. 23A.

2.15 RECOVERY OF DNA FROM AGAROSE AFTER GEL ELECTROPHORESIS.

2.15.1 GENECLEAN.

DNA restriction fragments were isolated from agarose/TAE gels using the GENECLEAN (BIO 101 Inc.) kit. The relevant DNA band was cut out of the gel under long-wave UV illumination, and the gel slice weighed. A volume of NaI solution equal to 2.5 times the weight of the gel slice was added to it, and the tube containing both was incubated at 55°C for 10 min to melt the agarose. 5 ul vortexed glassmilk solution was then added, and following vortexing, the mixture was incubated on ice for 5 min. The tube was then microfuged for 10 sec, and the supernatant discarded. The pellet was resuspended in 500 ul "New Wash" solution (10 mM Tris.HCl pH 7.5; 12.5 mM NaCl; 0.5 mM EDTA; 50% (v/v) EtOH), respun for 10 sec, and the supernatant again discarded. This was repeated twice more, and all the supernatant removed before resuspending the pellet in 10 ul TE. Next this solution was incubated at 55°C for 5 min, spun for 30 sec, and the supernatant containing the DNA recovered. This step was then repeated to increase the recovery of the DNA.

2.15.2 LMP AGAROSE.

After electrophoresis, the band was excised from the gel, 50

ul TE was added to it, and it was incubated at 65oC until the agarose had completely melted. An equal volume of phenol was then added, and the tube vortexed continuously for 1 min. It was then microfuged for 5 min, and the upper layer removed, avoiding the interface. This DNA containing solution was transferred to another tube, and then extracted once each with phenol, phenol/sevag and sevag, before being precipitated with EtOH.

2.16 SOUTHERN TRANSFER.

0.5 M D - 1.5 M NaCl; 0.5 M NaOH.
1 x N - 1.5 M NaCl; 0.5 M Tris.HCl pH 7.4.

Agarose gels containing DNA for transfer were soaked successively (with gentle agitation) for 30 min in 0.5 M D, 30 min in 1 x N, and then 15 min in either 1 x SSC or 20 x SSC depending on the type of filter used (NEN-GeneScreen or Pall-Biodyne, respectively). DNA was transferred onto either GeneScreen or Biodyne nylon membrane by capillary blotting overnight in either 1 x SSC or 20 x SSC respectively. Filters were then baked at 80^{o}C for 2 hrs.

2.17 HYBRIDISATION.

Hybridisation buffer - 4 x SET; 5 x Denhardts; 0.02 M phosphate buffer pH 6.8; 5% (w/v) dextran sulphate; 50-100 ug/ml denatured salmon sperm DNA; 10-20 ug/ml poly A; 10-20 ug/ml poly C; 0.1% SDS.
5 x W - 0.1% SDS, 0.1% sodium pyrophosphate, 0.025 M sodium phosphate buffer pH 6.8, 5 x SET.
1 x W - As for 5 x W, but with 1 x SET.
0.1 x W - As for 5 x W, but with 0.1 x SET.

Filters from Southern blots were pre-hybridised at 65^{o}C for at least 1 hr in a Hybaid hybridisation oven, using 15 ml hybridisation buffer per large tube. The denatured

radioactive probe was then added and the hybridisations carried out overnight at 65°C. After hybridisation, the probe and buffer were discarded, and the filters washed at 65°C in a shaking water bath for 30 min each in 5 x W, 1 x W, and finally 0.1 x W. The filters were then sealed in plastic bags and subjected to autoradiography.

2.18 RADIOACTIVE PROBES.

2.18.1 NICK-TRANSLATION.

The reactions were set up using 0.5 ug DNA in a reaction mix containing 5 ul 10 x nick translation buffer, 50 uCi ^{32}P-dATP (800 uCi/mmol), 10 ul nucleotide mix (500 uM dCTP; 500 uM dGTP; 500 uM dTTP), 3 ul DNase I (100 ng/ml), and 10 U DNA polymerase I, in a total volume of 50 ul. This was incubated at 14°C for 2-4 hrs, before the addition of 2.5 ul 0.5 M EDTA and 1 ul proteinase K (20 mg/ml), and subsequent incubation at 65°C for 5 min. Unincorporated counts were separated from the probe using a Sephadex G50 chromatography column, and the probes were used after being denatured by the addition of 100 ul formamide, followed by incubation at 70°C for 5 min, and then on ice for 1 min.

2.18.2 HU/MESSING PROBES.

Strand-specific probes were prepared from single-stranded M13 clones using the method developed by Hu and Messing (1982). 0.5-1 ug DNA was added to 1.5 ul 10 x HIN buffer, and 1 ul BRL 17-base probe primer, in a final volume of 10 ul, and then incubated at 90°C for 5 min. This reaction mix was then allowed to cool slowly to <30°C, before the addition of 1 ul 0.1 M DTT, 1 ul nucleotide mix (as for section 2.18.1), 10 uCi ^{32}P-dATP (800 uCi/mmol) and 10 U Klenow DNA polymerase. The reaction was then incubated at RT for 1 hr, before 1 ul 0.5 M EDTA was added to stop the reaction. The probe was separated from unincorporated counts as above (section

2.18.1), and used without denaturation.

2.19 AUTORADIOGRAPHY.

Autoradiography of probed filters and dried gels was carried out at -70°C, using 2 intensifying screens and Kodak X-Omat S film, for up to 3 weeks. Films were developed using a Kodak X-Omat film processor.

2.20 IN VITRO TRANSCRIPTION USING SP6 AND T7 POLYMERASES.

In vitro transcription was used both to generate riboprobes for RNase protection experiments, and in transcriptional run-off assays. In each case the reaction used to produce the radiolabelled RNA was basically the same.

2.20.1 IN VITRO TRANSCRIPTION REACTION.

The reactions were carried out using 100-500 ng plasmid DNA, linearised by restriction enzyme digestion, and either SP6 or T7 RNA polymerase, as appropriate. The reaction mixture contained 10 ul 5 x SP6 transcription buffer, 2.5 ul ^{32}P-GTP (10 uCi/ul; 3000 uCi/mmol), 2 ul 1.5 mM UTP, 1.25 ul of each of 10 mM ATP, 10 mM CTP, and 10 mM GTP, 2.5 ul 1 mg/ml BSA, 0.5 ul (15 U) RNasin (Promega), and 2 ul (2-4 U) RNA polymerase. The final volume was increased to 50 ul using dH_2O, and the reaction was carried out at 30°C for 30 min. Where the RNA was to be used for RNase protection experiments, 2 U RQ1 RNase-free DNase (Promega) was added to the reaction, which was then incubated for a further 15 min at 37°C; otherwise this step was omitted. One phenol/sevag and one sevag extraction were then carried out on the reaction mixture, after which the nucleic acid was precipitated for 1 hr at -20°C by the addition of 1/10th volume 3 M NaAc pH 6.0 and 2.5 volumes EtOH.

2.20.2 TRANSCRIPTIONAL RUN-OFF ASSAYS.

The precipitated nucleic acid, generated by *in vitro* transcription (section 2.21.1 above) was spun down in a microfuge for 20 min, the supernatant discarded, and the pellet washed in 70% EtOH. The tube was respun for 5 min, the supernatant pippetted off, and the pellet air dried. It was then resuspended in 10 ul sequencing gel loading buffer, and the RNA denatured by incubating at 70°C for 2 min before loading onto a polyacrylamide/urea gel.

2.20.3 RNase PROTECTION REACTIONS.

Formamide hybridisation buffer - 80% formamide; 40 mM PIPES pH 6.7; 0.4 M NaCl; 1mM EDTA.
RNase digestion solution - 10 mM Tris.HCl pH 7.5; 5 mM EDTA; 300 mM NaCl.

The precipitated nucleic acid, generated by *in vitro* transcription (section 2.21.1 above) was spun down in a microfuge for 15 min, the supernatant discarded, and the pellet washed in 70% EtOH. The tube was respun for 5 min, the supernatant pippetted off, and the pellet air dried. It was then resuspended in 20 ul dH_2O, and stored at -20°C if not required immediately.

The RNA or DNA sample (1-2 ug, and 100-200 ng, respectively) to be studied was precipitated and resuspended in 30 ul formamide hybridisation buffer. An excess amount of probe was added, and the sample was denatured by incubation at 85°C for 5 min, and then transferred to a 45°C heating block and left to hybridise overnight. The sample was subsequently cooled to RT, and 300 ul RNase digestion solution added. RNase A and RNase T1 were added to final concentrations of 40 ug/ml and 2 ug/ml, respectively. After incubation at 30°C for 60 min, 20 ul 10% SDS and 10 ul freshly prepared proteinase K (10 mg/ml) were added, and the sample incubated for a further 30 min at 37°C. Next 400 ul phenol/sevag was added, and an extraction

was carried out, after which the nucleic acid was precipitated at -20°C for 1 hr by the addition of 1 ml ice-cold EtOH. The sample was then centrifuged for 15 min, the pellet washed with 70% EtOH, and then dried. It was resuspended in 10 ul sequencing gel loading buffer, heated at 95°C for 5 min, then chilled on ice for 1 min, before being loaded onto a polyacrlyamide/urea gel.

2.21 SEA URCHIN MANIPULATIONS.

2.21.1 SEA URCHINS.

Adult sea urchins of the species Strongylocentrotus purpuratus were obtained from Pacific Biomarine, Venice, California, and maintained in sea water aquaria for up to 3 weeks.

2.21.2 SPAWNING.

ASW - 26.3 g NaCl, 0.7 g KCl, 11.9 g $MgSO_4.7H_2O$, 1.5 g $CaCl_2.2H_2O$, and 0.5 g $NaHCO_3$, made up to 1 litre in dH_2O and stored at 14°C.

Sea urchins were spawned by the repeated injections of 0.5 ml 0.5 M KCl. Eggs were collected by inverting the urchins on beakers full of ASW, whereas sperm were collected dry by inverting the urchins on petri-dishes. Eggs from each female were checked for premature envelope elevation, indicating fertilisation or activation, by microscopy, and any batches which were positive were discarded. Eggs were also checked for their ability to be fertilised, by the addition of 1 drop of diluted sperm (1 drop in 10 ml ASW) to a drop of eggs on a microscope slide. Any batches failing to show at least 95% envelope elevation 2 min after the addition of the sperm were also discarded.

2.21.3 ENUCLEATE EGG FRAGMENTS.

Enucleate egg fragments were prepared by a method based on that of Wilt (1973). Sea urchin eggs collected in ASW were passed through a Nitex 90 um filter, and washed 3 times in 500-800 ml ASW cooled to 14°C, allowing the eggs to settle under gravity each time on ice. On each occasion the ASW was aspirated off, and after the final wash the eggs were transferred to 50 ml conical centrifuge (Falcon) tubes, and again left to settle under gravity on ice. 1 ml packed eggs were then layered onto a 35 ml 0.5-1.0 M continuous sucrose gradient, prepared from 1.0 M sucrose in dH_2O, and a 0.5 M sucrose solution, which was made by mixing 1.0 M sucrose in dH_2O with an equal volume of ASW. The gradients were poured in ultra-clear Beckman SW28 tubes, cooled to 4°C, and after loading, spun at 4°C in a pre-cooled SW28 rotor in a Beckman ultra-centrifuge, at 7 krpm for 8 min, then 15 krpm for 11 min. The lowest band(s) in the gradient, containing the enucleate egg fragments, was collected by puncturing the bottom of the ultracentrifuge tube and transferred to 50 ml Falcon tubes. The fragments were then washed to remove the sucrose by adding an equal volume of cooled (14°C) ASW, and spinning at 2 krpm for 5 min. The supernatant was then discarded, and the pellets resuspended in ASW and centrifuged at 2 krpm for 5 min.

2.21.4 ACTIVATION OF ENUCLEATE EGG FRAGMENTS.

This was carried out following the method of Epel (1974), using 2 alternative sets of conditions. In both cases the egg fragments were incubated in ASW (100 ml ASW per 100 ul fragments) at 14°C for 2 hrs, with constant stirring and aeration. The ASW was either adjusted to pH 9.0 with Tris.HCl pH 9.4, and the egg fragments activated by the addition of 1 M NH_4Cl pH9.0 to a final concentration of 1 mM, or the pH of both the ASW and the NH_4Cl was adjusted to pH 8.0, and the NH_4Cl added to a final concentration of 10 mM. After incubation the fragments were spun down at 2 krpm for 2 min,

the supernatant discarded, and the pellets frozen at -20°C.

2.21.5 PREPARATION OF NUCLEIC ACID FROM OOCYTES/AEEFs.

Lysis buffer - 20 mM Tris.HCl pH 7.4; 100 mM NaCl; 10 mM EDTA; 10 mM EGTA.

The pellets were resuspended in 10 volumes ice-cold lysis buffer, and both proteinase K and 20% N-lauroyl sarcosine added to final concentrations of 50 ug/ml and 1% respectively. This solution was then incubated at 37°C for 30 min, and then phenol/sevag extractions were carried out until the interface was clear. One final sevag extraction was then performed, before the nucleic acid was precipitated using NaAc and EtOH. The final, dried nucleic acid pellet was resuspended in TE, in 0.5x the volume of the starting material.

CHAPTER

THREE

ANALYSIS OF REPLICATING mtDNA MOLECULES BY AGAROSE GEL ELECTROPHORESIS

3.1 INTRODUCTION.

Mitochondrial DNA replication has been most extensively studied in vertebrates, where it has been discovered that the two DNA strands (referred to as the heavy and light strands) have replication origins located in different regions of the genome. The heavy-strand (H-strand) origin is located in the major non-coding region of the mtDNA (Clayton, 1991). This area of the genome is much less well conserved between species than the coding regions, except for 3 highly conserved blocks of sequence called CSB I,II and III. The light-strand origin (O_L) is, on the other hand, located within a cluster of 5 tRNA genes in a short region of non-coding DNA between the $tRNA^{cys}$ and $tRNA^{asn}$ genes. Although the DNA sequence at O_L is not conserved between species, this region can fold to form a conserved stem-loop structure (Chang *et al*, 1985b).

In the sea urchin, the major non-coding region of the mtDNA is much shorter than those seen in other animal species (121 bp in *S. purpuratus* (Jacobs *et al*, 1988), compared to 879 bp in mouse (Bibb *et al*, 1982), 1122 bp in human (Anderson *et al*, 1981), and 0.8-5.4 kb in *Drosophila* (Goddard & Wolstenholme, 1978; 1980; Clary & Wolstenholme, 1985)), but it still contains blocks of sequence similar to the vertebrate CSB sequences, making it a likely site for a replication origin. Other experiments carried out in the laboratory (Jacobs *et al*, 1989) have located a D-loop to this region in egg mtDNA, supporting the idea that this is the location of a replication origin, and that the mechanism of replication initiation is similar to the strand displacement method of vertebrate mtDNA replication (Clayton, 1982).

From DNA sequence analysis (Jacobs *et al*, 1988), the *S. purpuratus* mitochondrial genome does not appear to contain a region analogous to the O_L sequences of vertebrates, although the presence of a minor L-strand start site within the $tRNA^{cys}$ gene of humans (Tapper & Clayton, 1981) does not rule out a tRNA gene acting as a lagging-strand origin in the sea

urchin. This would appear unlikely though, as the RNA primers for both the major and minor O_L in human mtDNA replication initiate at the same position within the non-coding region of this origin.

Initially I studied mtDNA molecules isolated from S. purpuratus in order to locate the origin(s) of replication, and determine the mechanism of DNA replication. To achieve this, a combination of gel electrophoresis techniques was used. The main technique utilised was the 2D agarose gel electrophoresis system developed by Brewer and Fangman (1987), which localises replication origins on the basis of the structures of the replication intermediates produced. Consequently, this also provides a considerable amount of information about the actual mechanism of DNA replication. To complement this technique, Southern blots of both native and denatured replicating molecules were performed to look for single-stranded regions of DNA, and to determine the lengths of nascent DNA strands.

3.2 ISOLATION OF REPLICATING mtDNA MOLECULES FROM S. PURPURATUS.

Before any of the electrophoresis experiments could be attempted, it was essential to develop a method for the isolation of replicating mtDNA molecules. Two potential sources of replicating molecules were examined; oocytes, and activated enucleate egg fragments (AEEFs).

The replication of mtDNA is extremely prolific in oogenesis, when the store of mitochondria, which are partitioned in early development, is being built up (Piko & Matsumoto, 1971; Piko et al, 1967). Therefore, I initially attempted to isolate oocytes as the source of mtDNA molecules. These attempts to isolate oocyte mtDNA, following the procedure of Matsumoto et al (1974), were, however, unsuccessful, probably due to the failure to obtain enough oocytes of sufficient purity by this method. I therefore decided to turn to

activated enucleate egg fragments (AEEFs) as my source of replicating mtDNA molecules.

In normal sea urchin development, fertilisation results in dramatic increases in mitochondrial RNA and protein synthesis, and the onset of nuclear DNA replication, but mitochondrial DNA replication does not occur until the pluteus larval stage of developmemt (Piko, 1969; Bresch, 1974), due to the apparent inhibition by the nucleus (Rinaldi & Giudice, 1985). Studies of the fertilisation of sea urchin eggs have revealed that many of the events occurring immediately after fertilisation can be induced by the action of NH_4^+ ions on eggs in culture (Epel et al, 1974). Removal of the nucleus, followed by activation of the eggs in culture, has been shown to stimulate mtDNA replication in the sea urchin, Paracentrotus lividus (Rinaldi et al, 1979a), and also in Xenopus (Rinaldi et al, 1981). AEEFs therefore provide a potential source of replicating mtDNA molecules from sea urchin eggs.

Enucleation was initially carried out following the method of Wilt (1973), in which 1.0 ml of a 20% (v/v) solution of eggs in artifical sea water (ASW) was layered on top of a sucrose gradient (see section 2.21.3 for details). Centrifugation produced 3 distinct bands of material in the gradient, the fastest migrating of which was of enucleated egg fragments, the middle one a mixture of unfractionated eggs and enucleated fragments, with nucleated fragments and some unfractionated eggs forming the top band.

To increase the amount of material collected from each gradient, fractionations were performed using different quantities of eggs. Fractionations using either 1.0 ml or 1.5 ml of settled eggs were performed, and the gradients examined by eye for the production of 3 bands, and by light microscopy for the formation of egg fragments. The loading of 1.0 ml of eggs consistently gave a clear separation of the enucleate fragments from the unfractionated eggs, whereas with 1.5 ml the 2 lower bands were poorly defined, and often appeared as

a large, single band. I therefore decided to use 1.0 ml of settled eggs per gradient in the enucleation step.

Activation of the enucleate fragments was carried out by the method of Epel *et al* (1974), in which NH_4Cl is added to a culture of the fragments in artificial sea water (ASW). Two different sets of conditions were tried, in which a final concentration of either 1 mM NH_4Cl, pH 9.0, or 10 mM NH_4Cl, pH 8.0, was used in ASW, adjusted to the appropriate pH. No obvious differences in the appearance of the replication intermediates on the blots of 2D agarose gels, or their amounts relative to the non-replicating DNA present (see section 3.4 for explanation), were apparent between mtDNA molecules activated using the different conditions.

The mtDNA isolated from these egg fragment was readily digested by the restriction enzyme *Sst*I, but was cut very poorly by many other enzymes including *Eco*RI, *Pst*I, *Kpn*I and *Hin*dIII. This is probably due to the presence of contaminating carbohydrate, though it may be related to the structure of the replicating molecules produced. Other researchers have also reported difficulties in digesting mtDNA from marine invertebrates, but no universal method for overcoming this problem has been reported in the literature.

The proportion of replicating molecules present in the isolated *S. purpuratus* mtDNA, based on the relative amounts of replicating and non-replicating DNA in the 2D gel analysis, is low compared to that reported by Rinaldi *et al* (1979a) for *P. lividus*, where electron microscopy studies suggested that about 20% of the mtDNA molecules examined were replication intermediates. This discrepancy could either be an effect of the 2D gel electrophoresis method used on *S. purpuratus* mtDNA, or could reflect differences between the 2 species in the efficiency of activation, or in the stability of the replication intermediates during isolation.

3.3 TWO-DIMENSIONAL AGAROSE GEL ELECTROPHORESIS.

In order to map the location of the replication origin in S. purpuratus mtDNA, the 2D agarose gel electrophoresis system developed by Brewer and Fangman (1987) was used. They developed this system to map the position of the replication origin of the yeast 2 um plasmid, and also used it to study DNA replication in the yeast ribosomal RNA gene cluster (Brewer & Fangman, 1988).

The system utilises the fact that non-linear DNA molecules migrate abnormally under agarose gel electrophoresis compared to linear molecules, and that these differences in mobility are exaggerated under conditions of high voltage and high agarose concentration (Bell & Byers, 1983). The gel system exploits this observation by first separating the DNA molecules according to mass, using a low percentage gel run at low voltage, then running them in the second dimension under high voltage and high agarose concentrations, to separate the molecules on the basis of their structure.

A heterogeneous population of replicating molecules is required for the system to work, since molecules at all stages of replication need to be examined. The replicating DNA is cut with restriction enzymes before electrophoresis, and the behaviour of each particular restriction fragment analysed by probing Southern blots of the resulting gels. The pattern produced by the replication intermediates (RIs) from a given restriction fragment depends on the position of this fragment in the genome with respect to the replication origin. A combination of different restriction enzyme digests and probes can therefore be used to map the origin relative to the restriction enzyme sites.

The easiest way to understand these differences is by considering the behavour of hypothetical 1 kb restriction fragments, located at different positions in the genome with respect to the replication origin. When the origin is located outside the restriction fragment, replication (assuming a

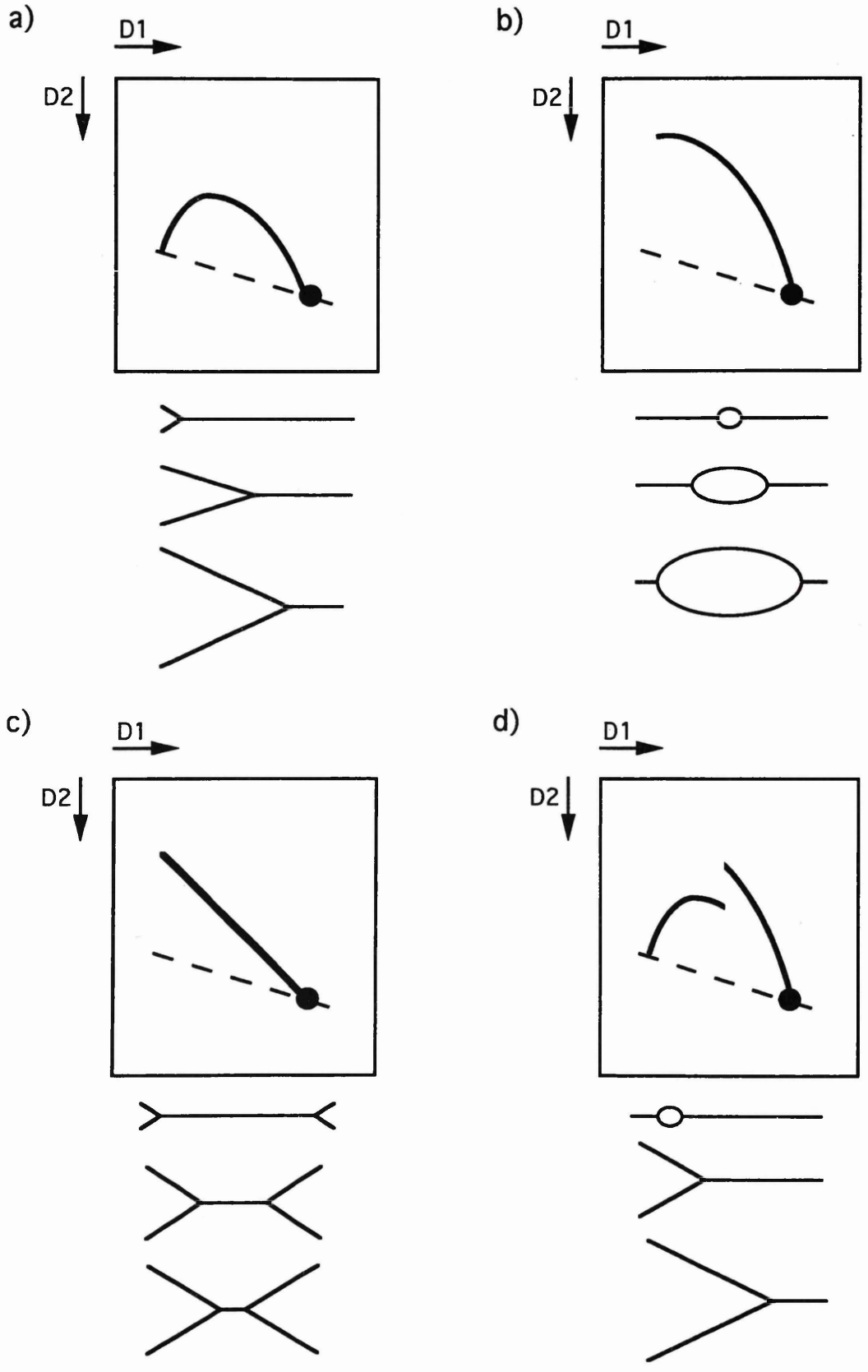


FIGURE 3.1: Diagrammatic representations of the effects of 2D agarose gel electrophoresis on the intermediates generated by replication of hypothetical 1 kb restriction fragments, and the structures of the replication intermediates (RIs). (a) Y-shaped; (b) bubble-shaped; (c) double-Y (or X-) shaped; (d) asymmetric (see text for explanation). On the 2D gel pictures, a solid line represents the RIs; a dashed line shows the position of migration of linear DNA molecules; and a filled circle represent linear, unreplicated, 1 kb molecules. Arrows show the direction of electrophoresis (D1- first dimension; D2 - second dimension). From Brewer & Fangman, 1987.

(facing page 74)

simple bi-directional mechanism) occurs by the unidirectional movement of the replication fork from one end of the restriction fragment to the other. A heterogeneous population of replicating molecules will therefore contain a spectrum of intermediates, from those in which the replication fork has just entered the fragment, to those in which the fragment is almost completely replicated.

As the replication fork passes through the restriction fragment, the Y-shaped RIs become less linear in structure as their molecular weight increases, so becoming more retarded in the second dimension of the gels, compared to linear molecules (Fig. 3.1a). When the fragment is exactly half replicated, all the arms will be of equal length, so it will be at its least linear, and therefore maximally retarded. Further replication increases the lengths of the replicating arms, compared to the non-replicated portion, so the molecules become more linear again (Fig. 3.1a), and are therefore retarded less under electrophoresis. This generates an arc of RIs which rises from the diagonal formed by linear molecules, peaks at 1.5 kb when the fragment is half-replicated and maximally retarded, and then returns to the diagonal as replication is completed (Fig. 3.1a).

If the replication origin is located at the exact centre of the 1 kb restriction fragment, a replication bubble would be generated, increasing in size until both forks passed the ends of the molecule. As the bubble grows in size, the RIs become less linear, and therefore increasingly retarded compared to linear molecules of the same size, resulting in the generation of an arc of RIs which does not home in on the linear diagonal (Fig. 3.1b).

The same is also true for the converse situation, in which the two replication forks meet in the centre of the restriction fragment. The double-Y-shaped molecules would become increasingly non-linear as replication proceeded, and so be more and more retarded as they got larger (Fig. 3.1c). The actual shapes of the curves for bubbles and double-Ys

differ, shown in Figure 3.1b,c are based on the observations of Brewer and Fangman (1987).

The location of the replication origin asymmetrically within a restriction fragment would generate a more complex pattern of intermediates than the situations outlined above. In this case a replication bubble would form initially, increasing in size until one of the replication forks passed the end of the fragment. At this point the curve of RIs produced would change from the bubble pattern to a simple-Y shape, mirroring the change in shape of the molecules, and generating a discontinuity in the arc of RIs at the point at which this change occurred (Fig. 3.1d).

By measuring the size of the molecules at the point of transition, the distance from the origin to one end of the restriction fragment can be calculated, assuming that the 2 replication forks move at the same rate. If the size of the RI at the transition was 1.3 kb, for example, then a total of 300 bp of DNA would have been replicated before the fork passed one end of the fragment. For bidirectional replication, this would put the origin 150 bp from one end of the restriction fragment. Analysis of a number of different restriction fragments using this procedure therefore enables the replication origin to be located within the genome.

The asymmetrical meeting of 2 replication forks within a fragment would also generate a complex pattern of RIs, though in this case the transition from a simple-Y to a double-Y-shaped curve would evolve gradually as the second fork slowly exerted its effect.

Further investigations into the behaviour of RIs in this 2D gel system have been carried out using the plasmid pBR322, in which replication is unidirectional (Martin-Parras et al, 1991). Replicating molecules were linearised at different enzyme sites within the plasmid, and then subjected to 2D gel electrophoresis. The results showed that the shapes of the curves depended on the distance between the replication

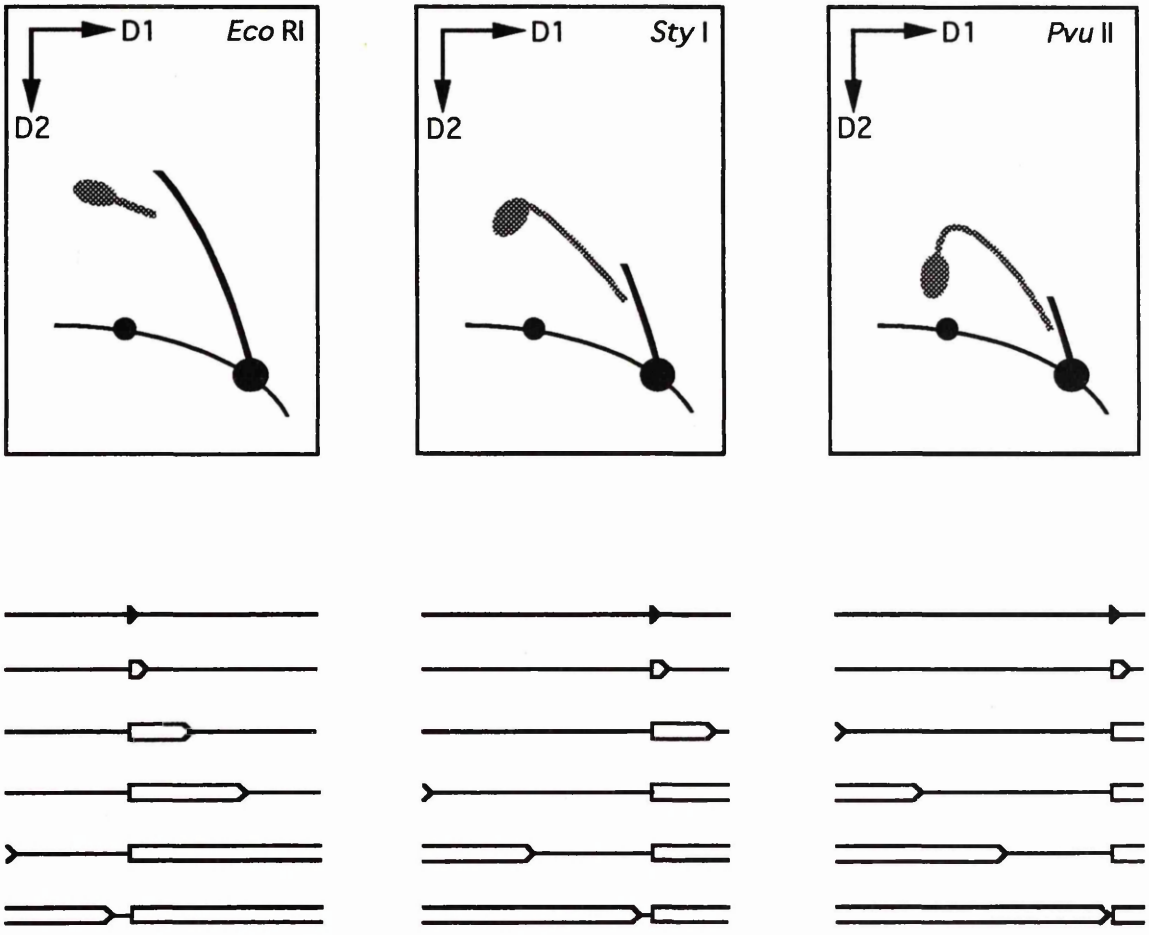
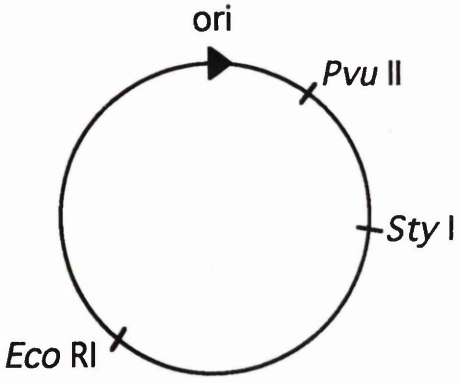


FIGURE 3.2: Diagrammatic representations of the effects of 2D agarose gel electrophoresis on the replication intermediates (RIs) produced by the plasmid pBR322, which have been linearised by restriction enzyme digestion. In these digrams the thick solid line represents bubble-shaped RIs; the shaded line and oval, double-Y shaped RIs; and the thin solid line, linear DNA molecules, on which the filled circles represent the unreplicated and fully-replicated plasmid molecules. Below the diagram of each gel are a series of diagrams of the structures of the RIs, in which the direction of synthesis is shown by the arrowhead. A diagram of the plasmid DNA shows the relative positions of the 3 enzyme sites and the origin of replication (filled arrowhead). Adapted from Martin-Parras *et al*, 1991.



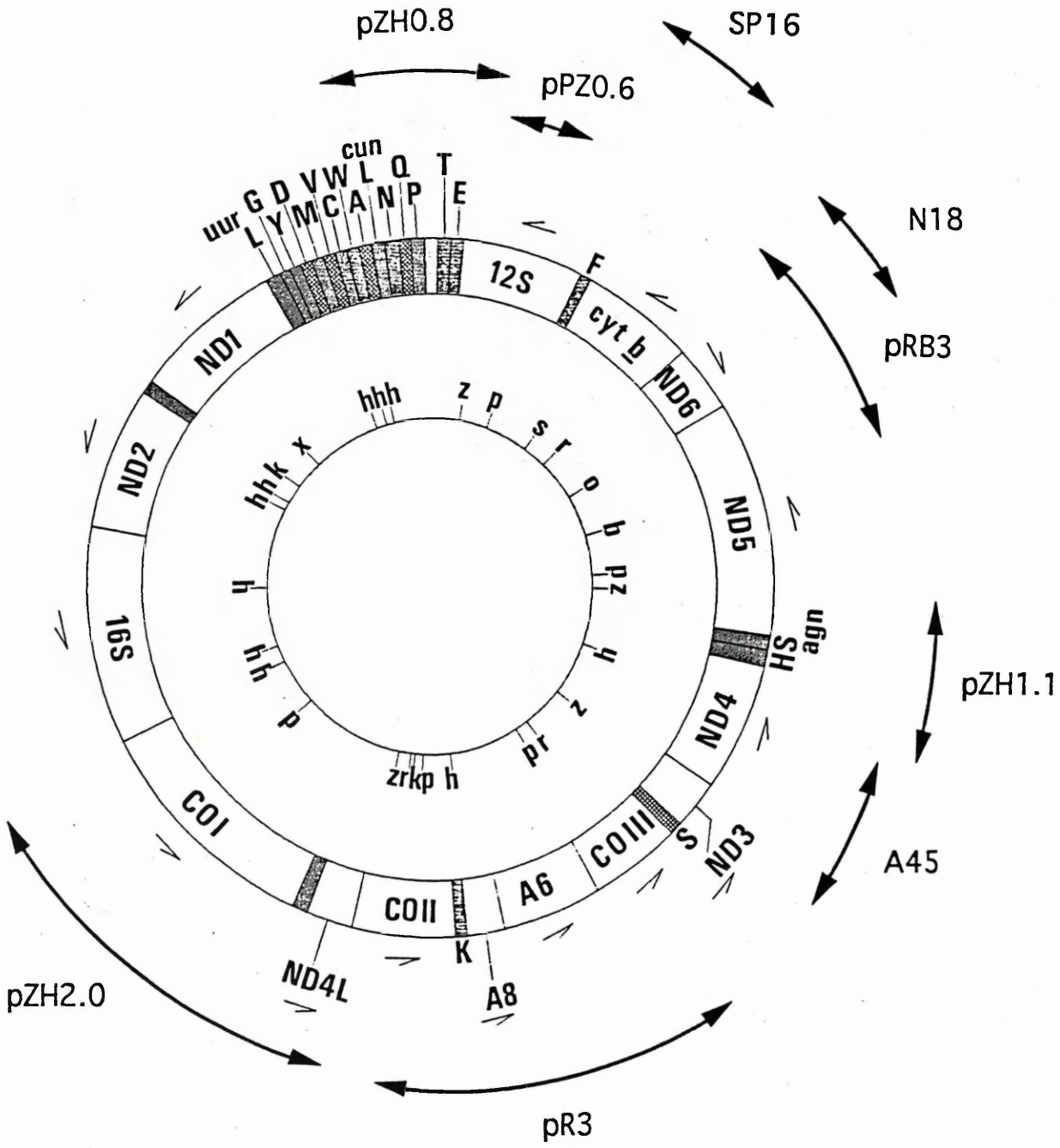


FIGURE 3.3: Map of the mitochondrial genome of *Strongylocentrotus purpuratus*, showing the locations of the various mtDNA subclones used as probes in the hybridisation experiments described in this chapter. The map also shows the positions of the restriction enzyme sites for *Bgl* II (b), *Eco* RI (r), *Hin* dIII (h), *Kpn* I (k), *Pst*I (p), *Sal* I (s), *Sst* I (z), *Xba* I (x), *Xho* I (o). For the abbreviations of the names of the various mitochondrial genes, see the legend to Figure 1.7.

(facing page 77)

origin and the enzyme site. The RIs change from a bubble form to a double-Y form after the replication fork has passed the enzyme site. The distance from the origin to this site determines the length of the fixed arm of the replicated portion of the double-Y intermediates, which influences the mobility of the intermediates generated by the movement of the replication fork through the rest of the molecule (Fig. 3.2).

3.4 LOCALISATION OF THE LEADING-STRAND REPLICATION ORIGIN.

A combination of 2D agarose gel electrophoresis, and conventional one-dimensional (1D) agarose gels of both neutral, and denatured DNA were used to map the position of the leading-strand replication origin in replicating mtDNA isolated from AEEFs.

3.4.1 TWO-DIMENSIONAL GEL ANALYSIS OF LINEARISED AEEF mtDNA MOLECULES.

The restriction enzyme BglII has a single recognition site in the S. purpuratus mitochondrial genome, located within the gene for NADH dehydrogenase subunit 5 (ND5; Fig. 3.3). Replicating mtDNA prepared from AEEFs, was digested with BglII, and then subjected to 2D agarose gel electrophoresis. The resulting Southern blot was probed with nick-translated pPZ0.6, a clone containing a portion of the 12S rRNA gene (Fig. 3.3).

The autoradiograph (Fig. 3.4a) shows a very prominent spot of linear mtDNA molecules, migrating at 15.7 kb, from which emerges a steep arc of replication intermediates. The curve ends at a clear discontinuity, resuming as a linear arc. A diagrammatic interpretation of the autoradiograph is shown in Figure 3.4b. The same pattern of RIs was seen with 2 different DNA samples from separate batches of enucleate fragments, with the only difference being in the position of

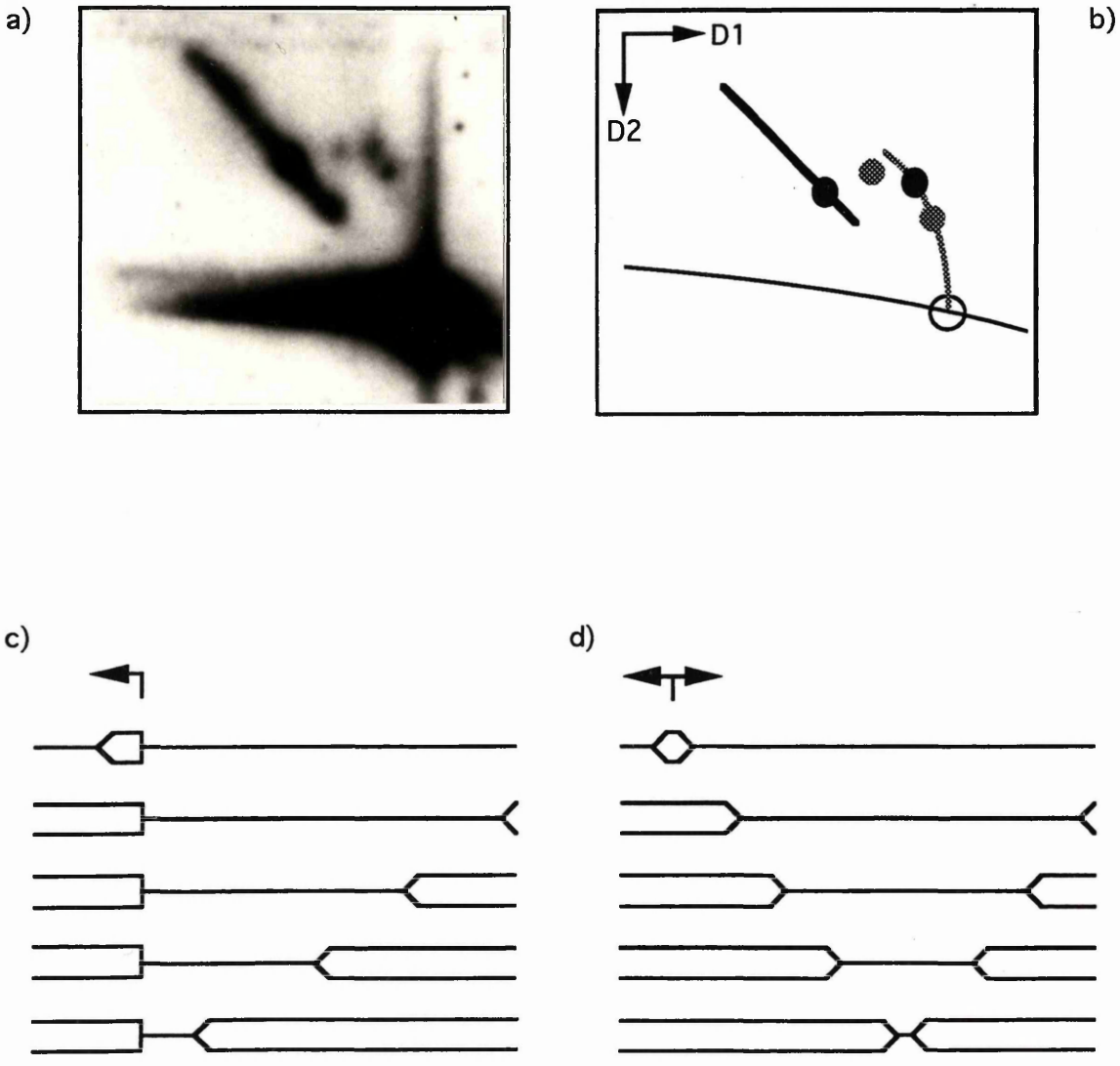


FIGURE 3.4: (a) Autoradiograph showing the results of the hybridisation of pPZ0.6 probe to a Southern blot of a 2D gel, on which *Bgl* II-digested AEEF mtDNA had undergone electrophoresis. (b) Diagrammatic representation of the autoradiograph. Thin solid line represents the position of migration of linear DNA; thick solid line, the replication intermediates (RIs); thick shaded line, the inferred positions of RIs; open circle, unreplicated (*i.e.* linear) AEEF mtDNA molecules; other circles, RIs of the same structure, detected in all (black), or some (shaded) of the DNA samples used. The proposed structures of the RIs are shown for (c) unidirectional and (d) bidirectional replication. No distinction is made between single- and double-stranded DNA, and the arrowheads indicate the direction of synthesis.

(facing page 78)

some of the spots lying on the arcs of RIs. These spots represent accumulations of RIs of the same structure, and have also been observed in further 2D gel studies on mtDNA cut with other restriction enzymes (section 3.5). The implications of this phenomenon for the mode of replication of S. purpuratus mtDNA are discussed below (see sections 3.6 and 3.7).

The pattern of replication intermediates observed can be interpreted on the basis of the patterns discussed earlier for the hypothetical 1 kb restriction fragment (section 3.3). A replication fork migrating from an origin located asymmetrically within the linearised mtDNA molecule would create a replication bubble, which increases in size, generating the steep curve of intermediates seen emerging from the linear DNA spot (Fig. 3.4a). The discontinuity arises when the replication fork passes the BglII site, creating Y-shaped intermediates. However, since the molecule is simply the linearised genome, further movement of the replication fork produces double-Y-shaped molecules, resulting in the linear arc of RIs seen in Figure 3.4a.

Schematic diagrams of the shapes of the possible RIs for both unidirectional, and bidirectional replication are shown in Figures 3.4c and 3.4d respectively. Both mechanisms generate similar double-Y-shaped intermediates, so it is not possible to distinguish between them on the basis of the pattern of RIs observed. The position of the discontinuity can, however, be used to calculate the distance of the replication origin from the BglII site for both uni- and bi-directional replication, although further experiments are needed to distinguish between the 2 alternatives, and determine on which side of the BglII enzyme site the origin lies.

From the interpretation of the 2D gel pattern outlined above, the end of the linear arc of RIs represents almost completely replicated mtDNA, which would be approximately 31.3 kb in size. However, calibration curves plotted from linear size markers (uncut and HindIII-digested phage lambda DNA) run on

the gels, indicate that the molecules at the end of the curve migrate at an apparent strand length of roughly 50 kb. This discrepancy is probably due to significant retardation of the RIs in the first dimension, caused by their large size, coupled with their highly non-linear structure (Fig. 3.4c,d). This reasoning is supported the 2D gels of SstI-digested mtDNA described below (section 3.5).

Taking the end of the linear arc as 31.3 kb, an estimate of the size of the replication intermediates at the discontinuity can be obtained, that works out to be 19 kb. This maps the origin to a site either 3.4 kb, or 1.7 kb from the BglII site, for unidirectional, and bidirectional replication, respectively, but does not indicate on which side of the BglII site the origin lies.

3.4.2 ONE-DIMENSIONAL GEL ANALYSIS OF DENATURED, LINEARISED AEEF mtDNA MOLECULES.

In an effort to map further the position of the replication origin in relation to the BglII site, and so determine its location within the mitochondrial genome, linearised mtDNA was alkali-denatured and then analysed by 1D agarose gel electrophoresis. This treatment separates the nascent strands from the full-length molecules, so by using probes for regions either side of the BglII site, the size and position of these molecules can be determined. The distance from the replication origin to the restriction enzyme site should be represented by a nascent strand of about 3.4 kb if unidirectional replication is occurring, although Okazaki fragments could be present if DNA replication is semi-discontinuous, rather than by strand displacement. For bidirectional replication, the situation is rather more complicated, with either a strand of 1.7 kb, or one of variable length being generated, depending on whether replication is simply by strand displacement, or involves the generation of Okazaki fragments.

Denatured, BglII-digested AEEF mtDNA was subjected to 0.7% agarose gel electrophoresis, and transferred onto nylon membrane by Southern blotting. Hybridisations were then carried out using 2 plasmids, pPZ0.6 and pZH1.1, which lie on either side of the BglII site (nt 13245) as probes (Fig. 3.3). A band of 3.55 ± 0.05 kb (sized against denatured DNA markers) was detected with ^{32}P-labelled pPZ0.6 (Fig.3.5a), but no clear band was seen using pZH1.1 as a probe (Fig. 3.5b). The 3.55 kb band was also faintly detected in the lane containing BglII-digested, native mtDNA with the pPZ0.6 probe, probably due to branch migration by some of the replicating molecules during or after isolation of the mtDNA, leading to release of the growing strand.

The 3.55 kb band was not detected in uncut, denatured mtDNA probed with pPZ0.6, though a band of 7.5 ± 0.1 kb was present (Fig. 3.5a), which was also detected by pZH1.1 (Fig. 3.5b).

The fact that pPZ0.6, but not pZH1.1 detected a discrete band when hybridised to Southern blots of denatured, BglII-digested DNA implies that the origin lies on the 12S rRNA side of the BglII site. The size of the band (3.55 kb) suggests that it has been generated by unidirectional replication, and maps the origin to nt 1145 ± 50, which is within the non-coding region of the genome (nt 1085-1205). This was the location implied by comparisons with the organisation of vertebrate mtDNA (Jacobs et al, 1988), and the position of the D-loop in S. purpuratus egg mtDNA (Jacobs et al, 1989).

The 7.5 kb band detected in uncut mtDNA by both probes may represent nascent DNA strands stalled at a particular site in the genome. This is consistent with the observation of spots present on the arcs of replication intermediates in the 2D gel analysis of BglII-cut mtDNA (Fig. 3.4a,b), and in the other 2D gels discussed below (section 3.5). With the origin located at nt 1145, the 7.5 kb band maps a pause site to nt 9300 ± 100, which is close to the position inferred from the 2D gel analysis of BglII-digested mtDNA. The 3.95 kb band

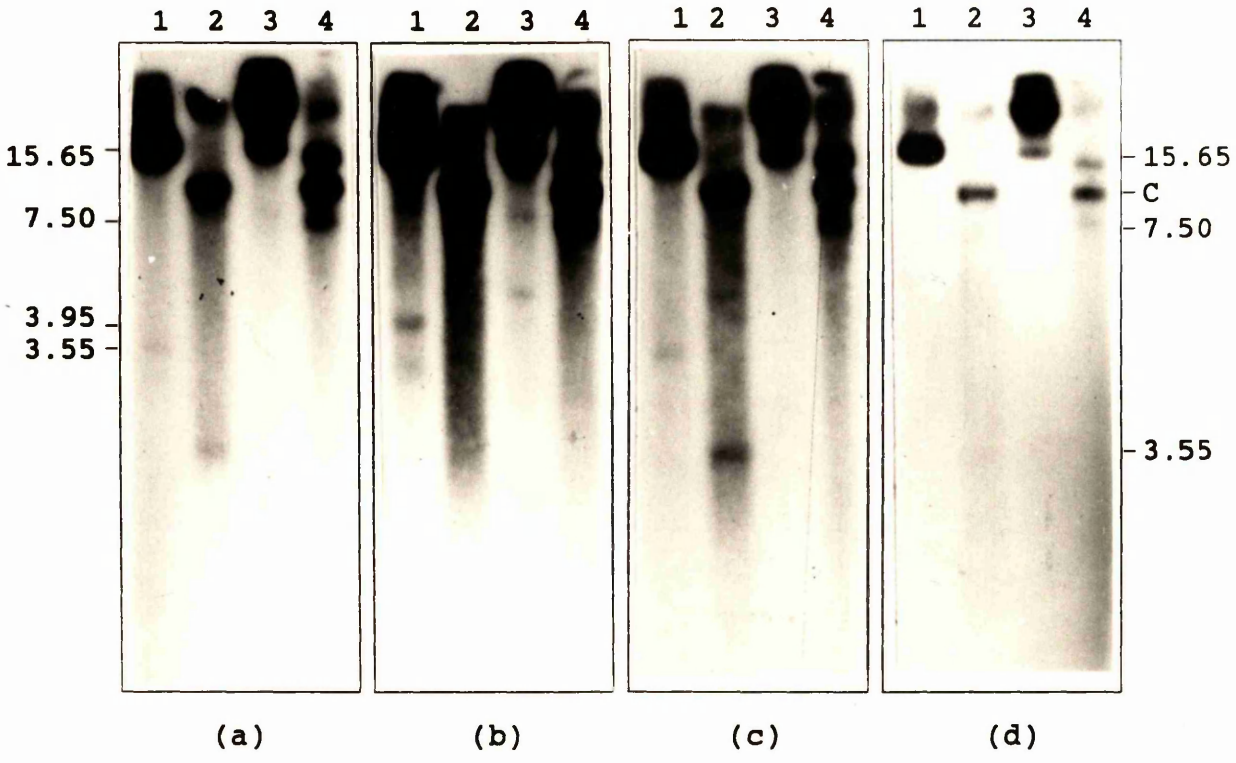


FIGURE 3.5: Autoradiographs showing the results of the hybridisations of various DNA probes to Southern blots of agarose gels, on which AEEF mtDNA had been run. Lane 1 - native, *Bgl* II-digested DNA; 2 - denatured, *Bgl* II-digested DNA; 3 - native, uncut DNA; 4 - denatured, uncut DNA. The probes used were: (a) pPZ0.6; (b) pZH1.1; (c) Hu/Messing probe SP16; (d) Hu/Messing probe N18. The sizes of the major bands discussed in the text are indicated either side of the figure for native (left) and denatured samples (right). C denotes the band believed to correspond with circular, single-stranded mtDNA in lane 4.

seen in native, BglII-digested DNA probed with pZH1.1 (Fig. 3.5b), could represent the nascent DNA strands between the BglII site (nt 13245) and the pause site (nt 9300), but the failure to detect this band in the denatured DNA sample, and the presence of other bands, makes it impossible to be certain.

DNA from 2 clones (SP6 and N18) from the region of mtDNA between the BglII site and the non-coding region (Fig. 3.3) was used to produce ^{32}P-labelled strand-specific probes. These probes were used to determine whether replication was proceeding by strand displacement, as in vertebrate mtDNA replication, or on both strands simultaneously, by Okazaki fragments.

Leading-strand (SP16), and lagging-strand (N18) probes were hybridised to Southern blots of alkali-denatured, BglII-cut DNA, and both detected a 3.55 kb band (Fig. 3.5c,d). However, the leading-strand probe, SP16, gave a stronger signal than probe N18 at exposures giving comparable signals for the 15.7 kb band of full-length mtDNA molecules. This is consistent with a D-loop expansion model of mtDNA replication, with the lagging-strand signal being due to the presence of molecules which have initiated at a lagging-strand origin, and extended back to, but not beyond, the leading-strand origin.

3.5 DETERMINATION OF THE MECHANISM OF mtDNA REPLICATION USING TWO-DIMENSIONAL AGAROSE GEL ELECTROPHORESIS.

Further analysis of the structure of replication intermediates was carried out on SstI-digested AEEF mtDNA, using the 2D gel technique. SstI cuts the mtDNA of *S. purpuratus* at 4 sites, and each of the restriction fragments was analysed using specific probes (Fig. 3.3). However, the analysis of the results was complicated by the occurrence of polymorphisms within the mtDNA molecules, even though eggs from only a few female sea urchins were used for each egg fragment preparation. A further problem was the failure of

SstI to digest single-stranded DNA, which increased the complexity of some of the 2D gel patterns.

3.5.1 ANALYSIS OF THE THREE SstI RESTRICTION FRAGMENTS OF mtDNA NOT CONTAINING THE NON-CODING REGION.

The SstI-digestion of S. purpuratus mtDNA produces 4 restriction fragments, 3 of which (the 2.4, 2.5, and 3.8 kb fragments) do not contain the 121 bp non-coding region. These 3 fragments all produced similar patterns of RIs when studied by 2D gel electrophoresis. In each case the curve of RIs formed an arc which moved away from the diagonal of linear molecules, then peaked, and returned to the diagonal at approximately double the size of the restriction fragment (Fig. 3.6a-d). This pattern is characteristic of the unidirectional movement of a replication fork through each restriction fragment (Fig. 3.1a).

The patterns of replication intermediates are complicated by the presence of restriction site polymorphisms. Another complication is the fact that the shapes of the curves differ from those predicted from the model fragments. Consequently, each figure will be discussed separately in the following paragraphs.

A ^{32}P-labeled probe from the plasmid pRB3, which is specific for the 3.8 kb restriction fragment, detected spots corresponding to linear DNA molecules of 3.8 and 6.2 kb, and hybridised to each with approximately equal efficiency (Fig. 3.6a). This is probably the result of a polymorphism at the SstI site in about half of the mtDNA molecules present in the sample, although it could conceivably be the result of partial digestion, in which one site is preferentially resistant to digestion. Curves of RIs can clearly be seen arising from both the 3.8 and 6.2 kb spots, although in each case the first part of the arc is poorly defined. The curves return to the diagonal at about double unit length (7.6 and 12.4 kb respectively), and as with the 2D gels of BglII-

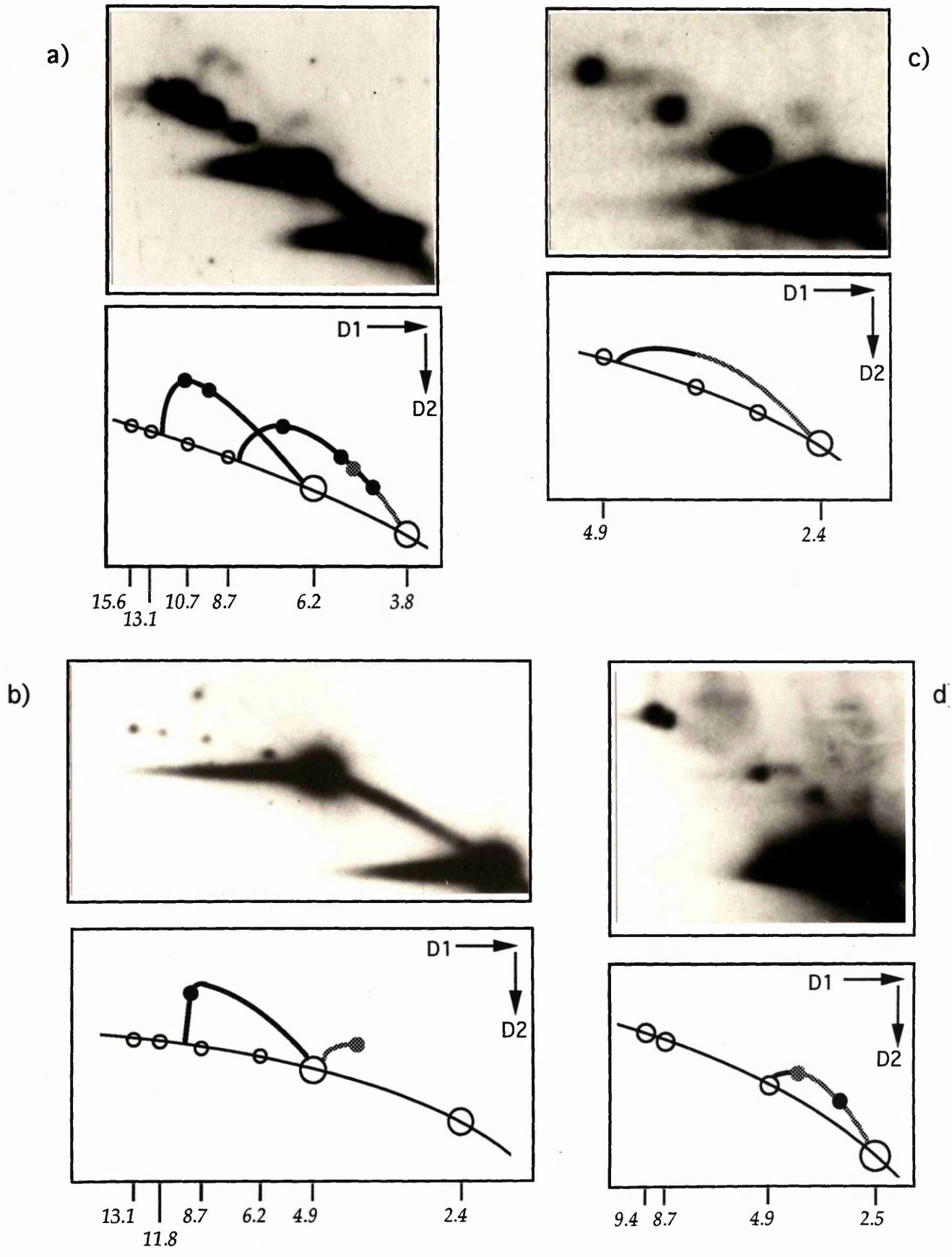


FIGURE 3.6: Autoradiographs (with accompanying diagrammatic interpretations) showing the results of the hybridisations of DNA probes to Southern blots of 2D agarose gels, on which *Sst* I-digested AEEF mtDNA had been analysed. The following probes were used: (a) pRB3; (b) A45; (c) pZH1.1; (d) pR3. See legend to Figure 3.5 for explanation of the symbols used in the diagrammatic interpretations. The sizes of the linear DNA molecules (in kb) generated by restriction enzyme digestion (including the partial digestion products), are given below each diagram.

digested mtDNA described above, there are other spots of RIs lying along these curves.

The 2 arcs do, however, differ slightly in shape from those expected on the basis of the Brewer and Fangman studies for the classic simple-Y replication intermediates. The RIs are most retarded when the restriction fragment is more than half replicated, so that the curve of the RIs is skewed. For the 3.8 kb fragment, the molecules at the apex of the curve migrate at 6.2 ± 0.2 kb, against a predicted size of 5.3 kb. For the 6.2 kb fragment, the apparant size is 10.7 ± 0.2 kb, in contrast to the predicted value of 9.3 kb. This skewing of the curves sugests either a significant degree of retardation in the first dimension for the large, branched molecules (which is consistent with the 2D gel analysis of the BglII-digested DNA; section 3.4.1), or the presence of abnormally-shaped RIs, which are at their least linear when more than half replicated.

This skewing of the arc of RIs is clearly visible for the 2.5 kb restriction fragment in a blot probed with A45 (Fig. 3.6b). An SstI restriction site polymorphism generated spots of equal intensity for the 2.5 and 4.9 kb linear molecules. Interestingly, the curve of RIs from the 2.5 kb fragment is barely visible, except at the very end where the molecules are nearly completely replicated, in contrast to the complete arc detected from the 4.9 kb fragment. In another DNA sample, which appeared to lack the SstI site polymorphism, the probe pZH1.1 detected a more complete curve of RIs emanating from the 2.5 kb restriction fragment (Fig. 3.6c). The pattern is very similar to that detected by pR3 for the 2.4 kb SstI fragment (Fig. 3.6d), and in both cases the skewing of the curve of RIs is less marked than for the larger restriction fragments.

On the basis of the models of 2D gel analysis described above (section 3.3), I propose that these 2D gel patterns, although rather complex, be interpreted as demonstrating that a replication fork passes unidirectionally through the entire

region of the genome covered by these 3 SstI fragments, and consequently, that this part of the mitochondrial genome does not contain a replication termination site, or the origin of leading-strand synthesis. However, it does not exclude the occurrence of lagging-strand replication origins, in which synthesis initiates on an already partially-replicated template molecule.

The question does arise as to whether a termination site could be located very close to the end of one of the restriction fragments, where the presence of the short arm of the double-Y intermediate would have little effect on the overall mobility of the molecules (Fig. 3.3). The presence of polymorphisms at the enzyme sites between both the 3.8 and 2.4 kb, and the 2.4 and 2.5 kb fragments would appear to rule out a termination site close to either of these 2 SstI sites (nt 10302, and 12669), since it would occur in the middle of the larger restriction fragment generated by the molecules lacking the intervening restriction sites. This would leave the only possible location for this termination site as being within the 2.5 kb restriction fragment,near the SstI site at nt 7765. The analysis of the 6.9 kb SstI fragment does (section 3.5.2 below), however, rule out this possibility.

3.5.2 ANALYSIS OF THE 6.9 kb SstI RESTRICTION FRAGMENT OF AEEF mtDNA.

Southern blots of SstI-digested mtDNA, run on 2D agarose gels, were probed with either pZH0.8, or pZH2.0, to detect the 6.9 kb restriction fragment and its pattern of replication intermediates (Fig. 3.7). The patterns detected are complicated by the presence of curves of replication intermediates emanating from the partial digestion products, which to some extent obscure the arc from the 6.9 kb fragment.

The arc of RIs detected by the probes for the 6.9 kb restriction fragment rises steeply from the spot of linear

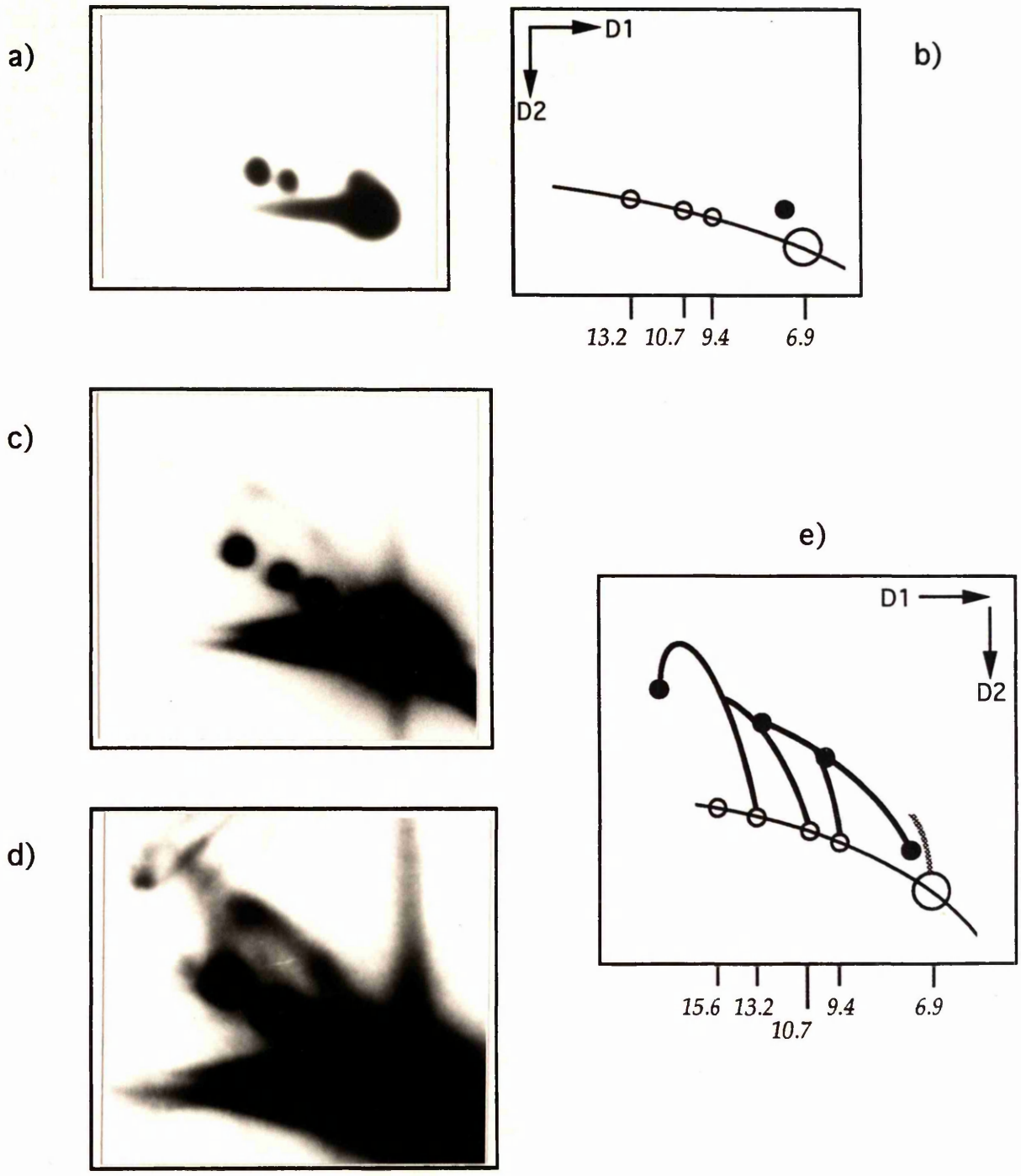


FIGURE 3.7: Autoradiographs showing the results of the hybridisations of pZH0.8 and pZH2.0 probes to Southern blots of 2D agarose gels, on which *Sst* I-digested AEEF mtDNA had been analysed. (a) Short autoradiographic exposure. (c,d) Long autoradiographic exposures. (b,e) Diagrammatic interpretations of the autoradiographs (see legend to Figure 3.5 for explanation of symbols). The sizes of the linear DNA molecules (in kb) generated by restriction enzyme digestion (including the partial digestion products), are given below each diagram.

DNA molecules, gradually becomes less steep as the size of the intermediates increases (Fig. 3.7c-e). This indicates that the RIs become less linear as replication proceeds, and are therefore either bubble, or double-Y structures (see section 3.3, and Figure 3.1b,c for explanation). In the short exposures of the autoradiographs, a spot of 7.2 kb is clearly visible, lying very close to the 6.9 kb spot of linear molecules, but off the diagonal formed by linear DNA molecules (Fig. 3.7a,c). This implies that the 7.2 kb spot is formed by non-linear replication intermediates.

The partial digestion products, visible as spots forming a diagonal of linear molecules, also give rise (in some cases) to curves of RIs. None of these arcs returns to the diagonal, and they have slopes of differing steepness. Although the curves are less intense than that emanating from the 6.9 kb fragment, their intensities are higher, relative to those of the linear DNA from which they emerge, than is the case for the curve of RIs from the 6.9 kb fragment (Fig. 3.7c-e). This points towards specific non-cutting of the replication intermediates, presumably due to their structure.

The shape of the curve of RIs detected for the 6.9 kb, SstI-digested, restriction fragment is consistent with the presence of a replication origin within the restriction fragment. This interpretation is supported by the analysis of the other 3 SstI restriction fragments (section 3.5.1), and the experiments using BglII-digested AEEF mtDNA (section 3.4). Based on its size, the 7.2 kb spot represents molecules in which 300 nt of DNA has been replicated, with the replication fork either having passed the restriction site at the end of the fragment (model A), or having been blocked at a specific point (model B). In either case, unidirectional replication must have occurred, unless the movement of both replication forks is affected, which seems unlikely.

Taking the first possibility (model A), a classic displacement-loop replication model (as occurs in vertebrate mtDNA; Clayton, 1982) can be envisaged, with the replication

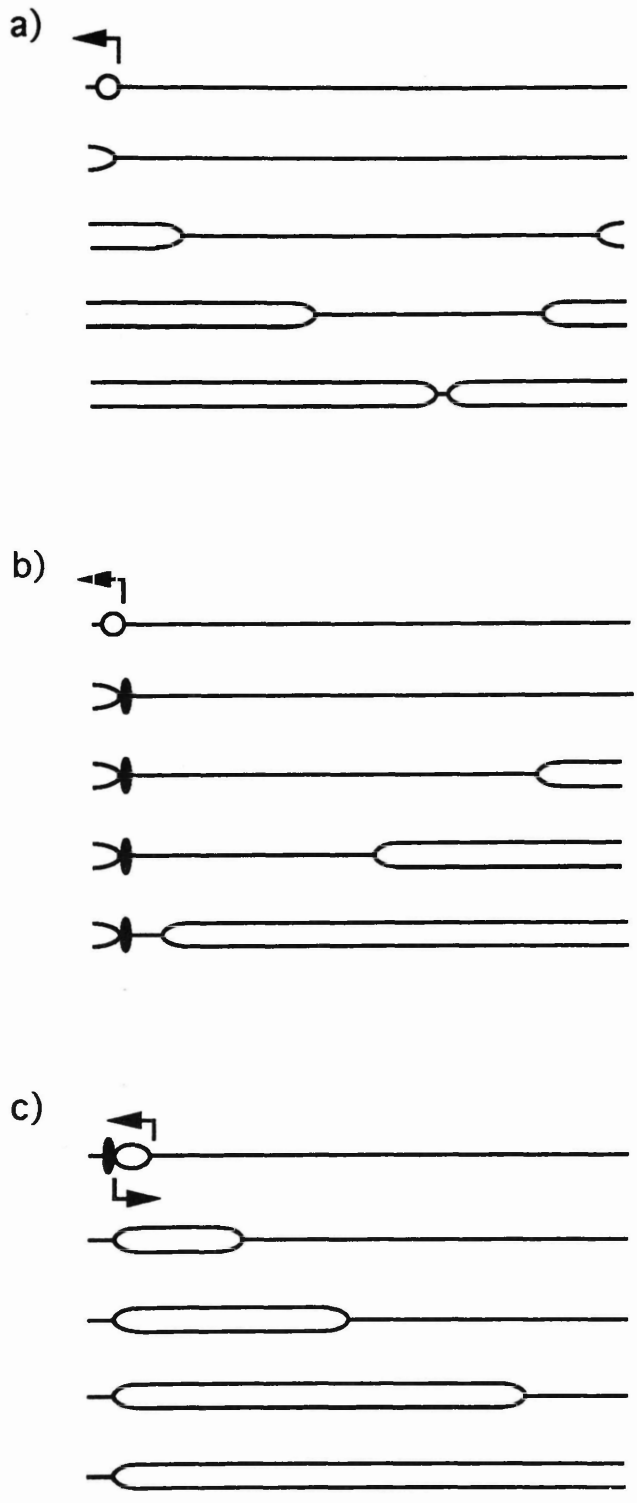


FIGURE 3.8: Diagrams of the possible structures of the replication intermediates generated by the 6.9 kb *Sst* I restriction fragment of AEEF mtDNA. (a) double-Y; (b) modified single-Y; (c) bubble. (a) & (b) correspond to variants of Model A, and (c) corresponds to Model B (see text for explanation). No distinction is made between single- and double-stranded regions of DNA. The filled ovals represent factors blocking replication fork movement, and the arrows show the direction of synthesis.

(facing page 89)

fork passing unidirectionally through the SstI site at nt 840 and on around the genome. This would then generate the patterns of RIs observed for the other 3 SstI restriction fragments. The unmasking of the lagging-strand replication origin(s), would then allow synthesis of the other strand. A situation can easily be imagined in which both replication forks move in opposite directions through the 6.9 kb fragment, creating double-Y-shaped intermediates (Fig. 3.8a), which would generate the pattern of RIs observed. A variant of this situation could arise, in which lagging-strand replication is blocked at the leading-strand origin (Fig. 3.8b), as occurs in vertebrate mtDNA, leading to the production of modified single-Y form intermediates (compare molecules in Fig 3.8b with Fig. 3.2).

Model B predicts a block to movement of the replication fork, possibly similar to that which blocks D-loop expansion in non-replicating molecules. Movement of a replication fork for lagging-strand synthesis back through the 6.9 kb SstI fragment would then produce a bubble, which increased in size, generating the observed pattern of RIs. The movement of the lagging-strand replication fork beyond the restriction site would, however, generate an abrupt discontinuity (Fig. 3.8c), but the simple-Y shaped RIs produced would be almost linear, so the continuation of the curve might be obscured by the partial digestion products, and their curves of RIs.

Model B is, however, at odds with the analysis of BglII-cut mtDNA, since it predicts movement of the replication fork in the opposite direction, which would produce a discontinuity at 27.8 kb, not 19 kb as seen (section 3.4.1; Fig. 3.4). It also does not account for the 3.55 kb band observed with denatured DNA, analysed by 1D gel electrophoresis (section 3.4.2; Fig. 3.5).

Model A would also account for the high proportion of intermediates observed emanating from the partial digestion products. On the basis of the displaced-strand model, supported by the 2D gel analysis outlined above, the movement

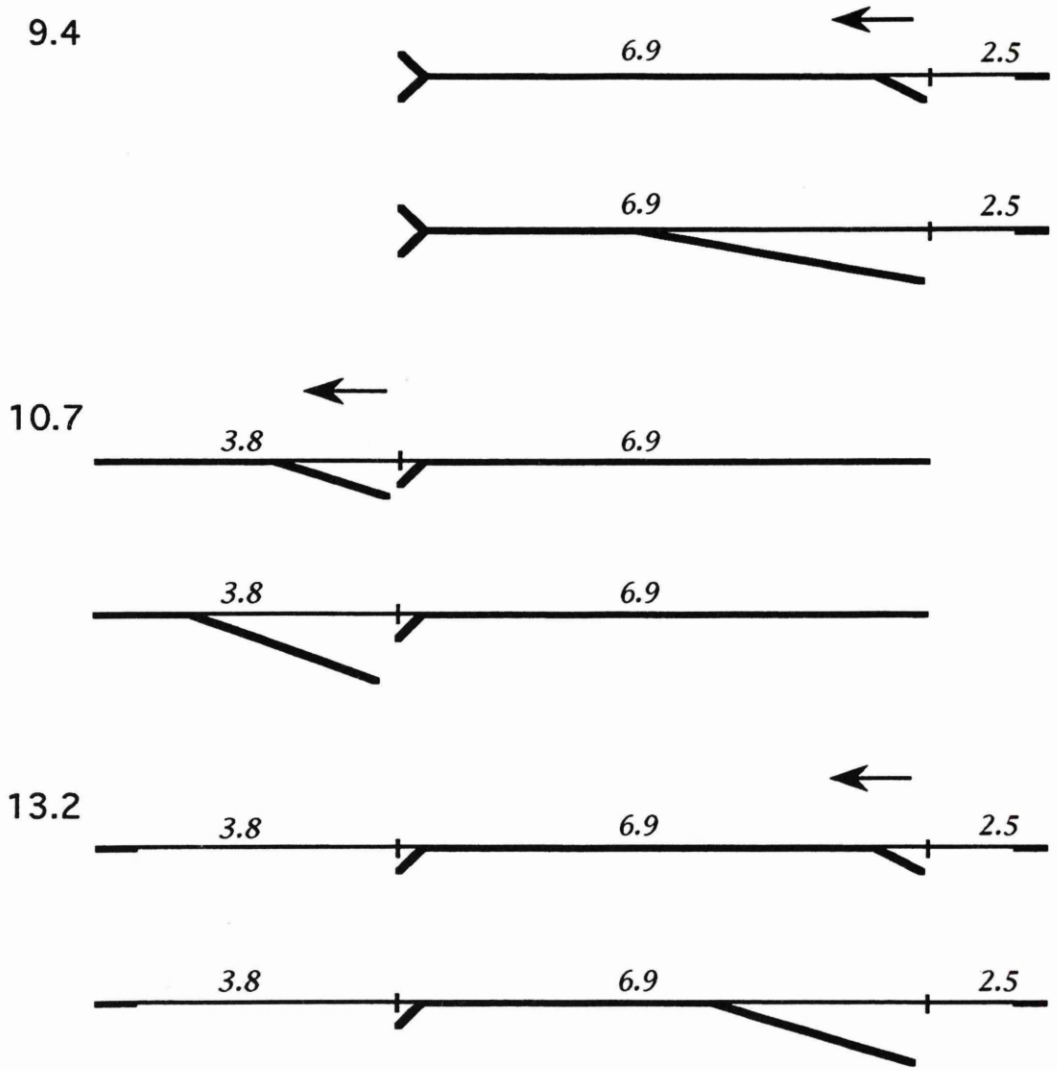


FIGURE 3.9: Proposed structures of the replication intermediates detected in Southern blots of 2D gels of *Sst* I-digested AEEF mtDNA, probed with pZH0.8. These RIs emanated from the spots representing the partial digestion products, the sizes of which are given next to each set of diagrams. The sizes of the *Sst* I restriction fragments present in each partially-digested molecule are given in *italics* above relevent portion of each molecule. Regions of single- and double-stranded DNA are represented by thin and thick lines, respectively, and the cross-bars represent the positions of the undigested *Sst* I sites on the single-stranded regions of DNA. The direction of replication fork movement is shown by the arrow above each diagram. Lagging-strand synthesis is assumed to stall at the leading-strand replication origin until resolution of the daughter molecules occurs on the completion of leading-strand synthesis (as in vertebrate mtDNA).

(facing page 90)

of the replication fork across the SstI site creates a triplex structure at this site. Only the double-stranded region of these RIs would be digested specifically by SstI (though some nicking of the single strands will probably also occur), producing a complex set of molecules containing both single- and double-stranded regions of DNA (Fig. 3.9). These intermediates would essentially be Y-shaped molecules, with a growing double-stranded arm branching off at different postions within the molecules. These RIs could therefore account for the arcs seen in the 2D gels, emanating from the partials (Fig. 3.7c-e).

3.5.3 ANALYSIS OF REPLICATION INTERMEDIATES PRODUCED BY OTHER ENZYME DIGESTS.

MtDNA was also digested with a variety of other restriction enzymes (KpnI, PstI, and SalI), in order to produce different restriction fragments for 2D gel analysis, so the results could be compared with those for the SstI-digested samples, and any ambiguities in interpretation resolved. Unfortunately, none of the enzymes listed cut AEEF mtDNA efficiently, and the absence of any complete arcs of replication intermediates suggested the presence of single-stranded nuclease activity, which would attack the replicating structures.

The aim of these further digests was to generate a series of restriction fragments, in each case with the proposed replication origin near to its centre, or offset, so that the patterns of RIs could be compared with those for the 6.9 kb fragment. However, when blots of the 2D gels of these double-digests were probed with pZH0.8, the patterns of RIs observed were simply faint spots lying off the diagonal (Fig. 3.10). Based on the interpretation of the structure of the 7.2 kb replication intermediate given in model A above (section 3.5.2), these spots were taken to represent intermediates in which the replication fork had moved from the leading-strand origin, located within the non-coding region, and past the

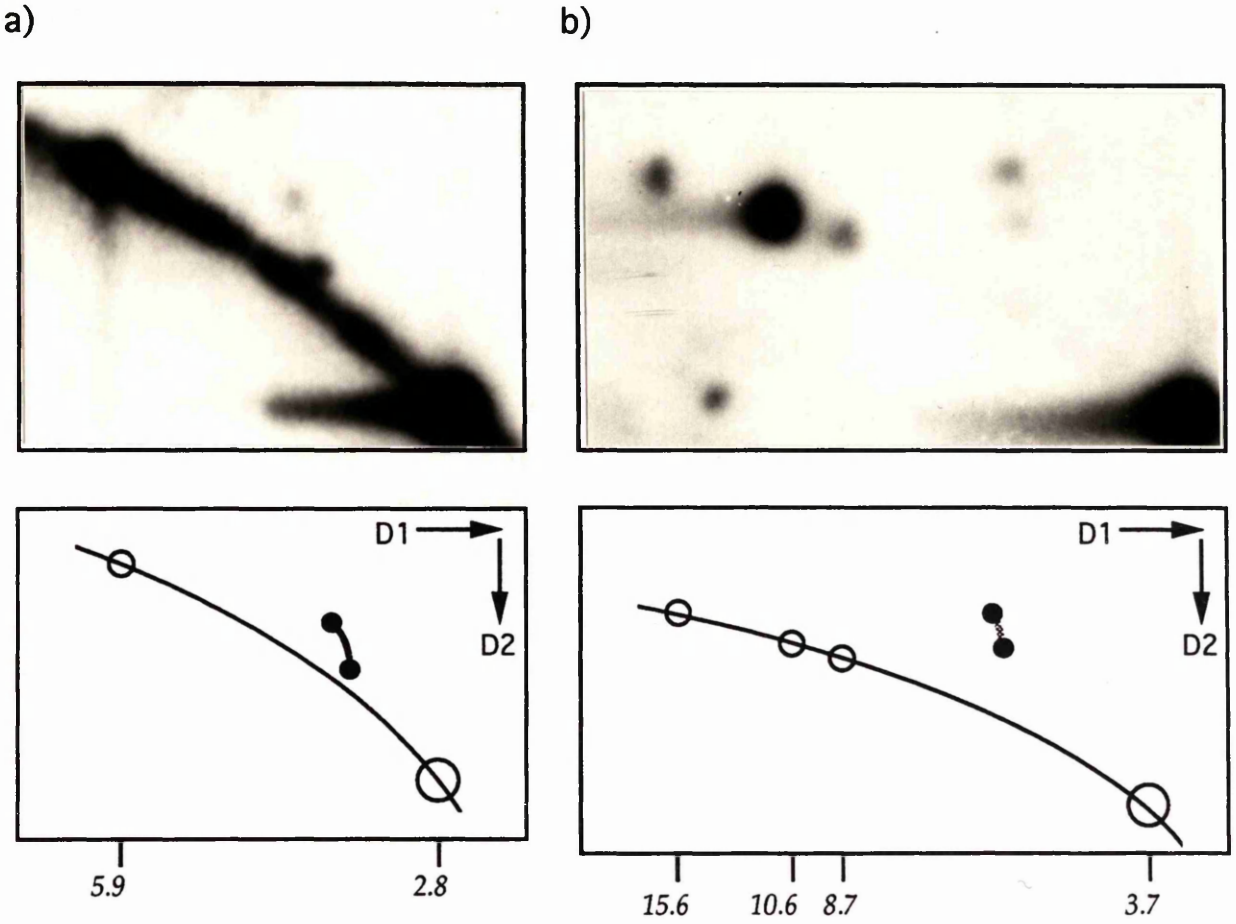


FIGURE 3.10: Autoradiographs (together with accompanying diagrammatic interpretations) showing the results of the hybridisations of pZH0.8 probes to Southern blots of 2D agarose gels, on which AEEF mtDNA had been analysed. (a) *Kpn* I/*Pst* I- digested DNA. (b) *Kpn* I/*Sal* I-digested DNA. See legend to Figure 3.5 for explanation of the symbols used in the diagrammatic interpretations. The sizes of the linear DNA molecules (in kb) generated by restriction enzyme digestion (including the partial digestion products), are given below each diagram.

relevant restriction site. This allowed the position of the origin to be mapped with respect to the different restriction enzyme sites.

In autoradiographs of the 2D gels of KpnI/PstI- and KpnI/SalI-digested AEEF mtDNA (Fig. 3.10), there appear to be 2 spots of RIs lying close together. In each case the size of the smaller one, which is closer to the diagonal, is consistent with uni-directional movement of a replication fork from within the non-coding region towards the 12S rRNA gene. For the KpnI/PstI autoradiograph (Fig. 3.10a), the spot is 3.7 kb in size: the expected size for such molecules, with an origin at nt 1145, is 3680 bp. With the KpnI/SalI digest (Fig. 3.10b), the observed size of the intermediate is 5.5 kb, against an expected size of 5460 bp.

Unfortunately neither of the gels gave a clear arc of RIs, so it is not possible to assess the structures of these molecules in any great detail. The only indication is the presence of the second spot in both of these blots, and the hint of a curve joining the two (most obvious for the KpnI/PstI digest; Fig. 3.10a). One possible interpretation is that the second, larger spot represents replication intermediates in which replication of the lagging-strand back to the origin has occurred, making these molecules completely double-stranded. This does, however, raise the problem of sizing the RIs, and the question of the relative rates of migration of partially double-stranded, and fully double-stranded molecules of the same strand length, which are the structures hypothesised for these 2 types of molecules.

PstI-digested and KpnI/BglII-digested mtDNA, analysed on 2D agarose gels gave similar patterns of RIs to those described above for the KpnI/PstI and KpnI/SalI digests (data not shown). The sizes of the spots were consistent with the replication model outlined in model A (section 3.5.2 and this section).

3.6 LOCATION OF THE REPLICATION PAUSE SITES.

The 2D agarose gel electrophoresis system used to map the position of the leading-strand replication origin, revealed the occurrence of spots on the curves of RIs (Figs. 3.4 & 3.6). Each of these represents a build up of RIs of the same structure within the heterogeneous population of replicating molecules. In other words, replication would appear to pause at certain points around the genome, and this temporary blockage to movement of the replication fork exists for a sufficient length of time for a significant proportion of the molecules to be isolated in these conformations.

The positions of the pause sites were analysed from Southern blots of the 2D gels of SstI- and BglII-digested AEEF mtDNA, and the results for at least 2 independently-isolated mtDNA samples correlated for each restriction fragment analysed. The 6.9 kb SstI restriction fragment was not included in the analysis, even though pause sites were detected (Fig. 3.7e). This is because the additional curves from the partial digestion products made it unclear as to on which curve of RIs the spots were located. The analysis of the positions of the replication pause sites therefore only applies to just over half the genome.

The positions of the pause sites within the genome are given in Figure 3.11, based on the strand-lengths of the molecules corresponding to each pause site, as judged from their migration in first dimension gels. There are several problems involved in the precise mapping of the pause sites, resulting from the difficulty of sizing the RIs accurately. Firstly, the skewing of the Y-shaped curves seen with the SstI-digests, implies that there is significant retardation of the highly branched molecules in the first dimension. This presents the problem of whether to size the molecules according to the linear DNA size markers, or to use the predicted sizes of the apex, and ends of the curve to plot a calibration curve.

SITE	NUCLEOTIDE POSITION	GENES
A	9350 ± 50	A6/COIII
B	12000 ± 200	ND4/tRNAs H,S/ND5
C	13400 ± 200	ND5
D	13900 ± 200	ND5/ND6
E	14900 ± 200	5' cyt *b*

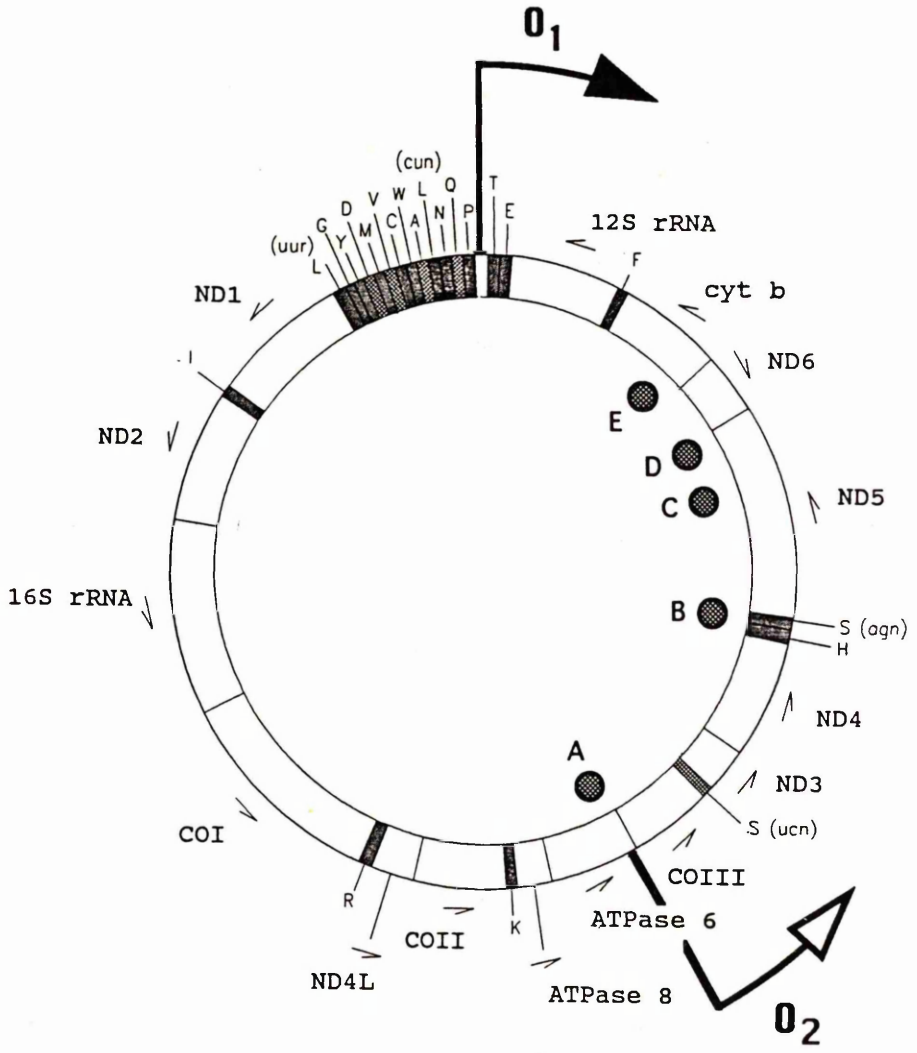


FIGURE 3.11: The positions of the replication pause sites for leading-strand DNA synthesis in the mitochondrial genome of *Strongylocentrotus purpuratus*. The table gives the nucleotide positions and the location of these sites in the genome, and the map, below, shows the positions of the sites (shaded circles) relative to the leading-strand replication origin (O_1; filled arrowhead indicates direction of DNA synthesis) and a prominent lagging-strand replication origin (O_2; open arrowhead).

The second problem is the effect on migration of the single-stranded regions of DNA present in the RIs, and how these molecules migrate compared to their fully double-stranded counterparts. The fact that it is not possible to judge the extent to which molecules are single-stranded has meant that all intermediates are assumed to migrate as completely double-stranded molecules, which contradicts the interpretation of the structures of the RIs observed in the KpnI/PstI and KpnI/SalI 2D gels (section 3.5.3).

A third, but less serious, problem is the fact that many of the spots are fuzzy, making it hard to size them accurately. This suggests either that some of the pause sites are relatively long regions of DNA, or that lagging-strand synthesis is occurring within a fragment in which leading-strand replication has stalled.

Taking into account inaccuracies involved in the mapping process, probably the most striking thing about the locations of the pause sites is that the majority occur close to the boundaries between genes. One particularly strong site is seen using the A45 probe (Fig. 3.6b), just beyond the apex of the arc of intermediates. The skewing of the arc makes it difficult to map the site accurately from this gel. However, by using a non-polymorphic DNA sample probed with pR3 (Fig. 3.6d), it was mapped more accurately to near the junction between the ATP synthase subunit 6 (A6), and cytochrome c oxidase subunit 3 (COIII) genes (nt 9350 $\pm$ 50). The digests of AEEF mtDNA with BglII also imply that a pause site is located in this area; the 2D gel analysis showed a prominent spot of RIs on the linear arc of about 23 kb (Fig. 3.4), implying that about 7.35 kb of the genome has been replicated in these intermediates, mapping the site to nt 9450 $\pm$ 150. The denatured, uncut mtDNA run on the 1D gels, gave a band of 7.5 kb with probes pPZ0.6 and pZH1.1 (Fig. 3.5a,b). A build up of nascent strands of this length indicates a pause site located at nt 9300 $\pm$ 50, again localising a prominent pause site to this region of the genome.

The SstI-digested mtDNA, probed with pRB3 which detects the 3.8 kb restriction fragment, revealed 4 pause sites (Fig. 3.6a). These mapped in the region of the 5′ end of the cytochrome b gene, the 3′ ends of the convergently-transcribed genes for subunits 5 and 6 of NADH dehydrogenase (ND5 and ND6), the 2 tRNA genes between ND4 and ND5, and within the ND5 gene itself. The possible roles of the pause sites in mtDNA replication, and gene expression, are explored in the discussion section below (3.7), and the concluding chapter (section 6.3).

3.7 DISCUSSION.

3.7.1 A SUMMARY OF THE GEL ELECTROPHORESIS DATA.

The data from the gels of AEEF mtDNA isolated from S. purpuratus are complex. In addition, many of the experiments cannot be interpreted unambiguously, but together they provide an internally consistent picture of the replication mechanism of the sea urchin mitochondrial genome.

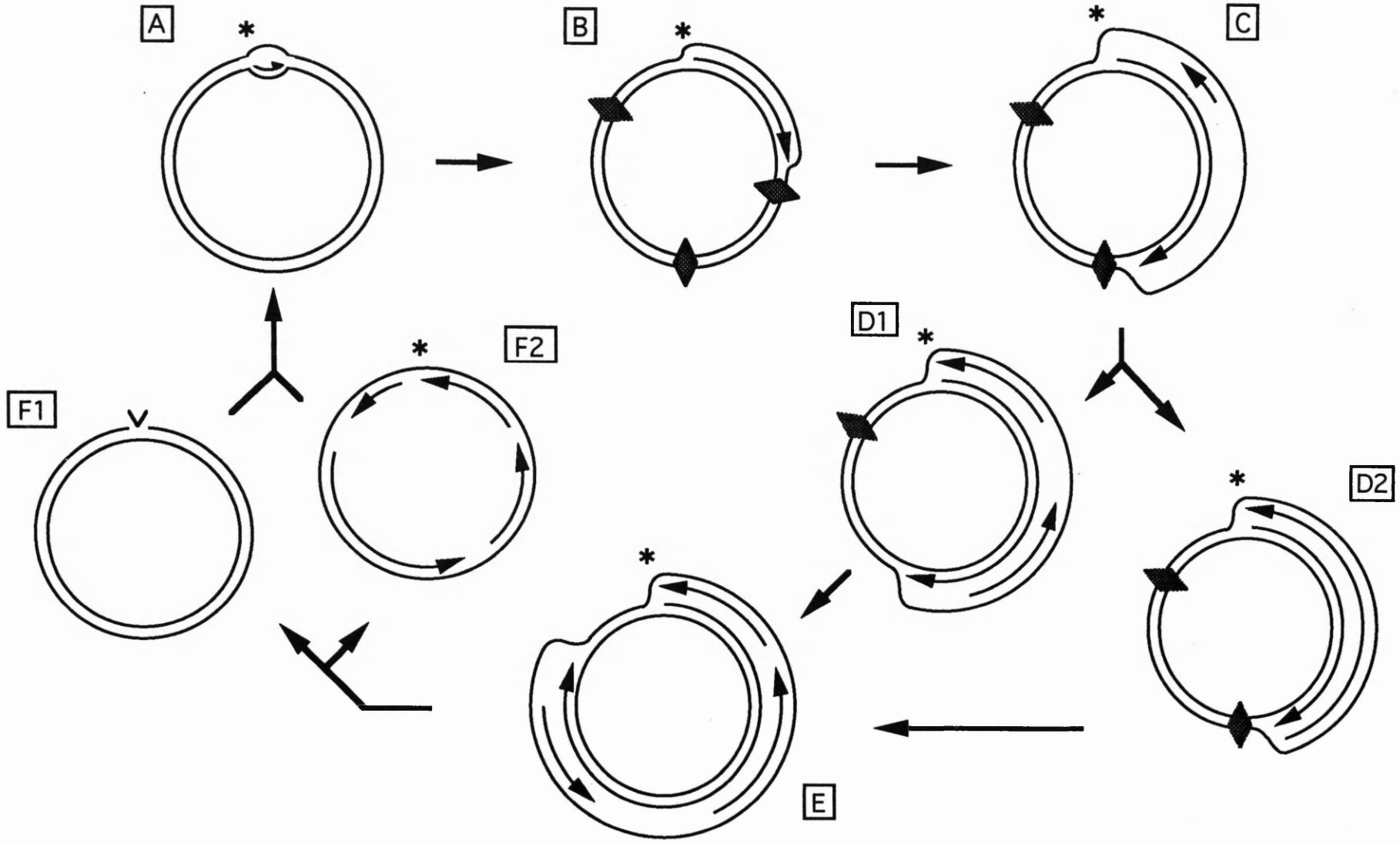
The 2D gels of the linearised (BglII-digested) mitochondrial genome gave a curve of RIs which contained a single discontinuity (Fig. 3.4), indicating that replication initiates from a single location in the genome. On the basis of the size of the RIs at this discontinuity, the origin is located either 3.4 or 1.7 kb from the BglII site, depending on whether replication is uni- or bi-directional, respectively. In order to locate the origin unambiguously, 1D agarose gel electrophoresis of alkali-denatured AEEF mtDNA was performed, and the resulting Southern blots probed for the sequences on either side of the BglII site (section 3.4.2). This procedure confirmed that replication was unidirectional, and that the origin was located 3.55 kb from the BglII site, which mapped it to within the non-coding region, at nt 1145 $\pm$ 50.

The 2D gel data from SstI-digested AEEF mtDNA are consistent

with the above interpretation of the data from the BglII-digested mtDNA. Analysis of the RIs from 3 of the restriction fragments (2.4, 2.5, and 3.8 kb) implied that there is unidirectional movement of a replication fork through each of these restriction fragments, and that none of them contains a termination site for replication (section 3.5.1). When the 6.9 kb restriction fragment from SstI-digested AEEF mtDNA was analysed by 2D gel electrophoresis (section 3.5.2), the curve of RIs did not return to the diagonal formed by linear DNA molecules (Fig. 3.7c-e), and an accumulation of RIs, 7.2 kb in length, was observed. The 7.2 kb intermediate may be interpreted as having been generated by the movement of the replication fork from the origin, past the SstI site at nt 840. Further replication of the fragment would then produce double-Y shaped intermediates. This interpretation is consistent with the analysis of BglII-digested AEEF mtDNA.

High levels of RIs were detected emerging from the partial digestion products seen in the 2D gels of SstI-cut AEEF mtDNA probed for the 6.9 kb fragment (Fig 3.7c-e). The high abundance of these molecules, compared to the spots representing the linear DNA molecules from which they emerge, can be explained as being the result of non-digestion of the single-stranded portion of the RIs by SstI. The 7.2 kb spot of RIs mentioned above therefore appears to represent molecules in which the SstI site has been readily digested by the enzyme, implying that this site is completely double-stranded. This in turn implies that lagging-strand synthesis has occurred back through the SstI site, and has stalled at the leading-strand replication origin, generating completely double-stranded molecules.

Several other digests of AEEF mtDNA were analysed by 2D gel electrophoresis (section 3.5.3), and the Southern blots probed for the restriction fragment containing the non-coding region. The results obtained were consistent with unidirectional replication from an origin within the non-coding region of the genome, past the enzyme site lying downstream from the 12S rRNA gene.



(facing page 98)

FIGURE 3.12: Proposed model for the replication of *S. purpuratus* mtDNA. The position of pause sites for leading-strand replication, and the corresponding lagging-strand replication origins, are shown as shaded diamonds. The leading-strand replication origin is marked by an asterisk (*), and the arrows indicate the direction of synthesis of the daughter strands. The V represents the position of the gap at the leading-strand replication origin before ligation of the daughter strand. See text for a full explanation.

The observation of RIs containing multiple regions of double- and single-stranded DNA by electron microscopy (Matsumoto et al, 1974) implied that multiple sites exist for the initiation of lagging-strand DNA synthesis in S. purpuratus mtDNA. My gel electrophoresis results are consistent with this observation, and the 1D gel data suggested that one of these origins is located in the region of the genome containing the A6/COIII gene boundary (section 3.4.2). A replication pause site was also detected in the same region of the genome, raising the possibility that the several pause sites observed by 2D gel electrophoresis (Fig. 3.11) correspond to the sites of lagging-strand initiation. The 1D gel experiments were consistent with the stalling of lagging-strand synthesis at the leading-strand origin, detecting a band of 7.5 kb in uncut mtDNA, which would represent the distance between the leading- and lagging-strand origins.

3.7.2 A MODEL FOR mtDNA REPLICATION IN S. PURPURATUS.

On the basis of the interpretation of the data from the gel electrophoresis experiments summarised above, I propose the following model for mtDNA replication in S. purpuratus. A diagram of this model is shown in Figure 3.12, and the different stages (A-F) are referred to by letter in the following description. A proportion of the mtDNA molecules isolated from eggs exist in the D-loop form (Jacobs et al, 1989), with the D-loop located within the non-coding region of the genome (A). Replication of the mtDNA molecules occurs from an origin at the same location, and proceeds by D-loop expansion towards the 12S rRNA gene (B). Replication is, therefore, initially unidirectional by strand displacement. At several sites around the genome leading-strand synthesis pauses (shown as shaded diamonds; B-D), so replication intermediates of the same structure accumulate. These pause sites correspond to sites of lagging-strand initiation (as shown in Fig. 3.12), which would be predicted, on the basis of the mammalian model, to act when single-stranded. A pause, or termination site for lagging-strand synthesis occurs at

the leading-strand replication origin (D & E), as in vertebrate mtDNA replication, which can lead to the formation of completely double-stranded (Cairns form) intermediates by some molecules (D2). The completion of leading-strand replication, and separation of the daughter molecules produces 2 different classes of molecules (F1 & F2), which are similar to those proposed for the vertebrate model (Fig. 1.2). The completion of lagging-strand DNA synthesis, and ligation of the daughter strands (at the Vs shown in Fig. 3.12) completes the replication cycle.

Apart from the occurrence of multiple lagging-strand replication origins, which is in accordance with the earlier results of Matsumoto et al (1974), the proposed model for S. purpuratus mtDNA replication is very similar to that for vertebrate (Clayton, 1982) and Drosophila (Goddard & Wolstenholme, 1978; 1980) mtDNA, and kDNA minicircles (Sheline et al, 1989). There are also similarities with mtDNA replication in yeast, and these points are discussed further in Chapter 6 (section 6.1.2).

3.7.3 FURTHER EXPERIMENTS TO TEST THE MODEL.

The location of the leading-strand replication origin was determined more accurately by mapping the 5′ end of the nascent DNA strand by RNase protection assays, using probes covering the non-coding region of the genome. This approach is described in Chapter 4 below.

Although the 2D gel results are internally consistent, certain assumptions have been made about the behaviour of the replication intermediates under electrophoresis, especially the effects of regions of single-stranded DNA on the migration of molecules. The analysis of plasmid DNA replication (Martin-Parras et al, 1991) indicated that the structure of the intermediates can influence migration in the first dimension, as well as the second. This was shown by the behaviour of X-shaped recombinant forms of the same molecular

weight, but different structure. These molecules showed an apparent increase in molecular weight due to first dimension retardation as they became less linear in structure (i.e. more retarded in the second dimension). The discrepancy in the apparent and expected sizes of the RIs seen with BglII-digested AEEF mtDNA from S. purpuratus is therefore not surprising, given the large size of the molecules (15.65 kb unreplicated), and the formation first of bubble, and then of double-Y-shaped intermediates.

The behaviour of the Y-shaped RIs seen with SstI-cut AEEF mtDNA, probed with A45 and pRB3 (Fig. 3.6b,a) could also be explained by retardation of the highly-branched RIs during first dimension electrophoresis. The curves are skewed, so that the apex, at which the molecules are least linear and so most retarded, apparently occurs when the restriction fragments are more than half replicated.

Alternatively, the skewing of the curves of Y-shaped RIs could be due to the occurrence of regions of single-stranded DNA in the molecules, which affect their structure such that they are most retarded when more than half replicated. If molecules of the same strand length containing single-stranded DNA migrate significantly differently from completely double-stranded molecules, then the question arises as to what exactly are the structures of the RIs observed in the different gels. The appearance of multiple lagging-strand replication origins implies that the extent of lagging-strand synthesis will vary for a given amount of leading-strand replication (as it does with just the single O_L in vertebrates; Robberson et al, 1972), and could therefore produce a wide range of molecules of the same strand length with differences in mobility. This would be predicted to lead to the production of fuzzy, ill-defined curves of RIs, or even no distinct curves at all. If this is the case, do the curves of RIs represent only the most commonly occurring forms of intermediates?

If predictions about the structure of the adjacent spots of

RIs seen in the 2D gels of KpnI/PstI- and KpnI/SalI-digested AEEF mtDNA are correct, then the molecules migrate slightly differently in the first dimension, when they are partially single-stranded, than they do when they are completely double-stranded. However, the difference in mobility is more marked in the second dimension, under conditions of high voltage and agarose concentrations. The curve of RIs between the 2 spots in the KpnI/PstI gel (Fig. 3.10a) may therefore indicate the conversion of molecules containing a region of single-stranded DNA to ones which were completely double-stranded.

Two approaches could be adopted to investigate whether the interpretations of the 2D gel electrophoresis analysis are correct. Firstly, artificial RIs could be constructed, and their behaviour under different gel electrophoresis conditions compared to that of completely double-stranded molecules of the same strand length. The easiest molecules to construct would be pseudo-intermediates: double-stranded molecules with a single-stranded tail of differing lengths, which could be produced by exonuclease III digestion. The migration of these molecules could be compared against completely double-stranded molecules of the same strand lengths as the intermediates, and the double-stranded molecules produced by removal of the single-stranded tails from the intermediates. Different gel conditions representing first and second dimension 2D gel electrophoresis could be tried.

Bubble intermediates could also be constructed by creating heteroduplexes, resulting in the formation of completely single-stranded bubbles. These molecules could potentially be used to create intermediates with partially or completely double-stranded bubbles by oligonucleotide-primed DNA synthesis. More complex molecules could potentially be constructed using combinations of different oligonucleotides hybridised together, but problems could arise in verifying that the expected structures are actually formed, and creating large enough intermediates to ensure that any

changes in mobility are observed.

The second possible approach would be to carry out 2D gel electrophoretic analysis on replicating vertebrate mtDNA (e.g. from Xenopus oocytes). In this case a replication mechanism has been proposed based on electron microscopy and the mapping of the 5′ ends of the daughter DNA strands. This means that the shapes of the RIs generated by different restriction enzyme digests of replicating vertebrate mtDNA can be predicted from the model, and their performance under 2D gel electrophoresis observed. The interpretations of the structures of the sea urchin RIs, based on 2D gel patterns, could then be reassessed in the light of the patterns produced by the vertebrate RIs.

There is, however, the danger that the electron microscopy data did not reveal the complete picture of the replication mechanism of vertebrate mtDNA. The occurrence of minor lagging-strand replication origins at other sites in the genome away from O_L would upset any predictions as to the structures of the vertebrate RIs, and could therefore confuse the interpretation of the S. purpuratus data. It could, in fact, turn out that the experiments on vertebrate mtDNA would yield new information about mtDNA replication in vertebrates, rather than resolving the uncertainties in the model proposed for S. purpuratus.

Electron microscopy could also be used to to examine the structures of replicating mtDNA molecules in the sea urchin, with linearised DNA being used to provide a reference point in the molecules. The distribution of single- and double-stranded regions of the RIs could then be investigated, which could reveal the locations of the lagging-strand replication origins and pause sites. The regions of single-stranded DNA could be highlighted by the use of single-stranded DNA binding proteins, possibly in conjunction with immunogold staining. One major problem with this method may be the isolation of sufficient material for the study, given the apparently small number of replication intermediates compared

to non-replicating molecules, seen by the 2D gel analysis. Another potential problem could be branch migration occurring in the replication intermediates during their isolation and preparation for electron microscopy.

CHAPTER

FOUR

MAPPING DNA AND RNA TERMINI AT THE REPLICATION ORIGIN AND PAUSE REGIONS

4.1 INTRODUCTION.

In mammalian mtDNA, H-strand replication is primed from the light-strand promoter (LSP; Chang & Clayton, 1985; Chang *et al*, 1985), with the RNA to DNA transition occurring near the conserved blocks of sequence (CSBI, II, III; Chang *et al*, 1987). In mammals this transition has been shown to be the result of cleavage of the RNA primer by the enzyme MRP (Chang & Clayton, 1987a,b). In D-loop synthesis in the human and rat mtDNA molecules, the 5′ end of the DNA strand has variable start sites (3 in human (Brown *et al*, 1978), 2 in rat (Seikya *et al*, 1980)) whereas in mouse (Gillum & Clayton, 1978; 1979) and *Xenopus* (Gillum & Clayton, 1978), it is the position of the 3′ end that is variable. The gel analysis of replicating mtDNA from *S. purpuratus*, described in Chapter 3, assigned the leading-strand origin to the non-coding region of the genome, but did not reveal anything of the organisation of this region in terms of the number of start sites, and the location of transcripts, terminators, or transition sites.

The vertebrate mitochondrial genome contains a single second-strand (L-strand) replication origin, which is a short non-coding region of DNA located within a cluster of 5 tRNA genes. The region forms a stem-loop structure, which is conserved between several species, including man, mouse, and *Xenopus* (Chang *et al*, 1985b). As with H-strand replication, DNA synthesis is primed by an RNA molecule, but the mechanism of primer formation is different. At O_L, the RNA primer is formed by a primase (Wong & Clayton, 1986), and not RNA polymerase as occurs at O_H. The 5′ end of the RNA primer is located within the loop structure, with the RNA to DNA transition occurring near the base of the stem (Wong & Clayton, 1985). In human this switch occurs at a pentanucleotide sequence, GGCCG (Hixson *et al*, 1986), which is homologous to that found in *Xenopus*, but different from those in rat and mouse (Chang *et al*, 1985b).

In *S. purpuratus* analysis of the mtDNA sequence reveals no

region of the genome corresponding to the vertebrate O_L sequences (Jacobs *et al*, 1988), and both the agarose gel electrophoresis analysis of the replicating mtDNA molecules, described in Chapter 3, and the EM studies of Matsumoto *et al* (1974) were consistent with multiple initiation sites for second-strand DNA synthesis. One of these sites appears to be located close to the boundary of the genes for ATP synthase subunit 6 (A6) and cytochrome *c* oxidase subunit III (COIII; section 3.4.2). In the sea urchin species *P. lividus*, Cantatore *et al* (1989) proposed that this region of the genome formed the second strand origin on the basis that its potential secondary structure resembled that of the vertebrate O_L region.

The analysis of replicating mtDNA molecules, isolated from AEEFs from *S. purpuratus*, located the leading-strand origin of replication to the 121 bp non-coding region of the mitochondrial genome. The experiments also implied that the replication fork for leading-strand synthesis paused at several sites in the genome. One of these pause sites, located near the junctions of the A6 and COIII genes (Fig. 4.1), could, on the basis of the denatured DNA blots, be at the site of the prominent lagging-strand replication origin already referred to. In addition, the blots implied that a termination or pause site for second-strand synthesis was present at the first-strand replication origin, as occurs in vertebrate mtDNA (Bogenhagen *et al*, 1978).

I therefore decided to study the A6/COIII boundary and the non-coding region of the mitochondrial genome in more detail, in order to locate the ends of any DNA and RNA molecules present. This should allow the site of the leading-strand origin to be mapped more precisely, and the pause site and prominent second-strand origin to be located, as well as revealing more about the mechanism of mtDNA replication in sea urchins. In addition, the 3′ end of the A6 transcript and the 5′ end of the COIII transcript were mapped, in order to analyse further the transcriptional organisation of the genome. The analysis of the D-loop region also meant that the

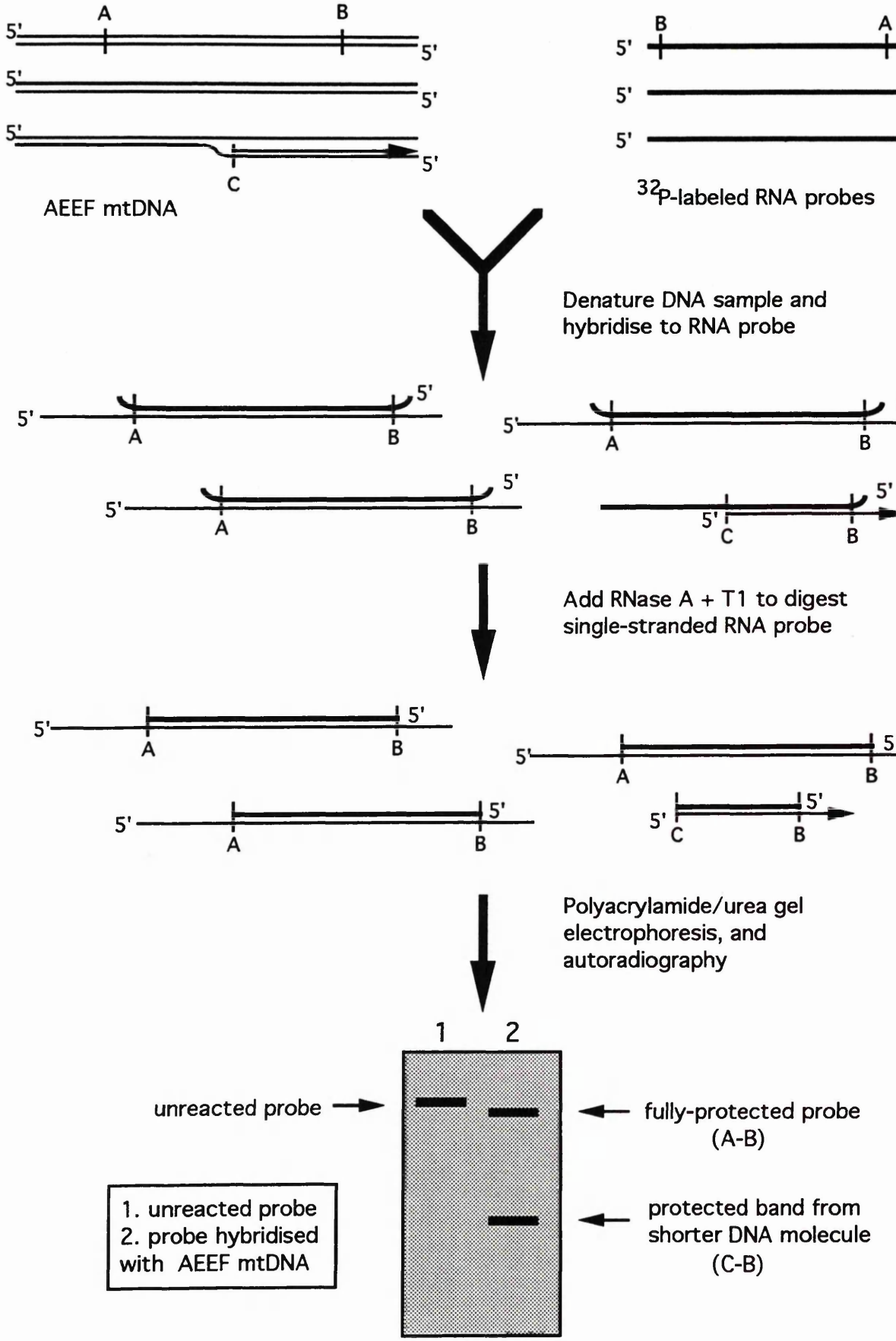


FIGURE 4.1: Schematic diagram of an RNase protection reaction involving an *in vitro* synthesised, ^{32}P-labelled RNA probe (thick line) and AEEF mtDNA (thin line), with the arrow representing a newly synthesised strand of mtDNA, initiating within the region covered by the RNA probe. The complementary regions of probe and DNA sample occur between sites A and B; site C represents the 5' end of a nascent DNA strand.

(facing page 106)

possible occurrence of a termination site for lagging-strand synthesis at the leading-strand origin could be investigated.

4.2 RNase PROTECTION EXPERIMENTS.

The A6/COIII junction region and the non-coding region of the S. purpuratus mitochondrial genome were studied by RNase protection assays, using ^{32}P-labelled RNA probes synthesised in vitro. The 2 regions of the mitochondrial genome were subcloned into plasmids containing the promoter for SP6 RNA polymerase, with the inserts introduced in both orientations so that RNA probes complementary to either strand could be synthesised. Solution hybridisation of these probes to mitochondrial DNA or RNA samples, isolated from sea urchin AEEFs, was followed by treatment of the reaction mixtures with RNase A and RNase T1 to digest away any single-stranded RNA. Electrophoresis of the denatured reaction products, followed by autoradiography, then revealed the lengths of the protected probe molecules (Fig. 4.1). The use of probes of different lengths enabled the ends of the molecules which did not protect the probe fully to be mapped within the genome. An internal control for the RNase digestion step is provided by the polylinker sequences present at the ends of the probe molecules. These sequences do not hybridise to the mitochondrial nucleic acid samples, so the fully-protected probe molecules will give shorter bands than the unreacted (unhybridised and undigested) probe on gel electrophoresis (Fig. 4.1).

4.3 SUBCLONING OF REGIONS OF THE S. PURPURATUS MITOCHONDRIAL GENOME INTO SP6 VECTORS.

4.3.1 THE NON-CODING REGION.

MtDNA from the non-coding region, and some of the adjacent tRNA genes, was recloned into 2 vectors containing the SP6 RNA polymerase promoter upstream of the cloning site. The

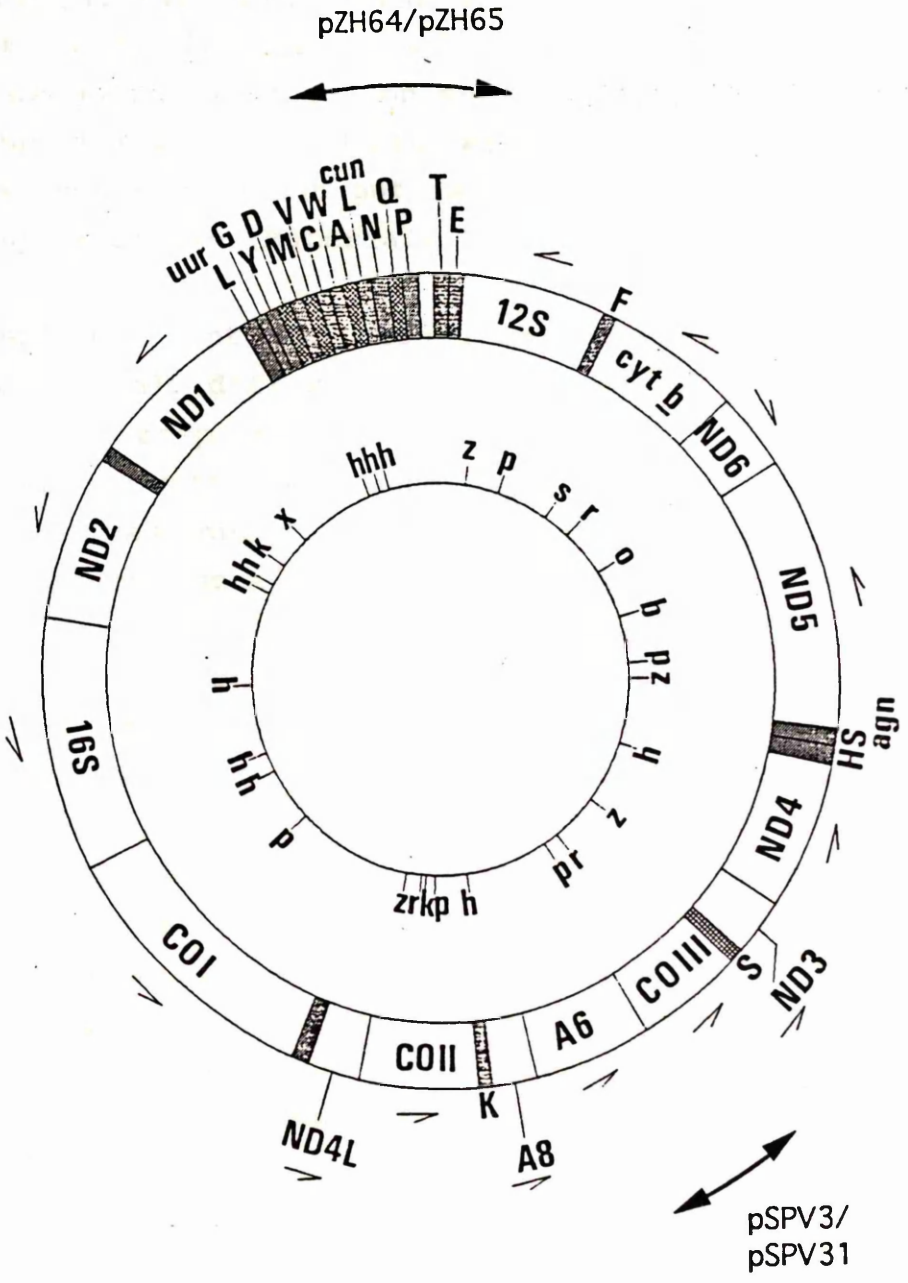


FIGURE 4.2: Map of the mitochondrial genome of *Strongylocentrotus purpuratus*, showing the locations of the various mtDNA subclones used as probes in the hybridisation experiments described in this chapter. The map also shows the positions of the restriction enzyme sites for *Bgl* II (b), *Eco*RI (r), *Hin*dIII (h), *Kpn*I (k), *Pst*I (p), *Sal*I (s), *Sst* I (z), *Xba*I (x), *Xho*I (o). For the abbreviations of the names of the various mitochondrial genes, see the legend to Figure 1.7.

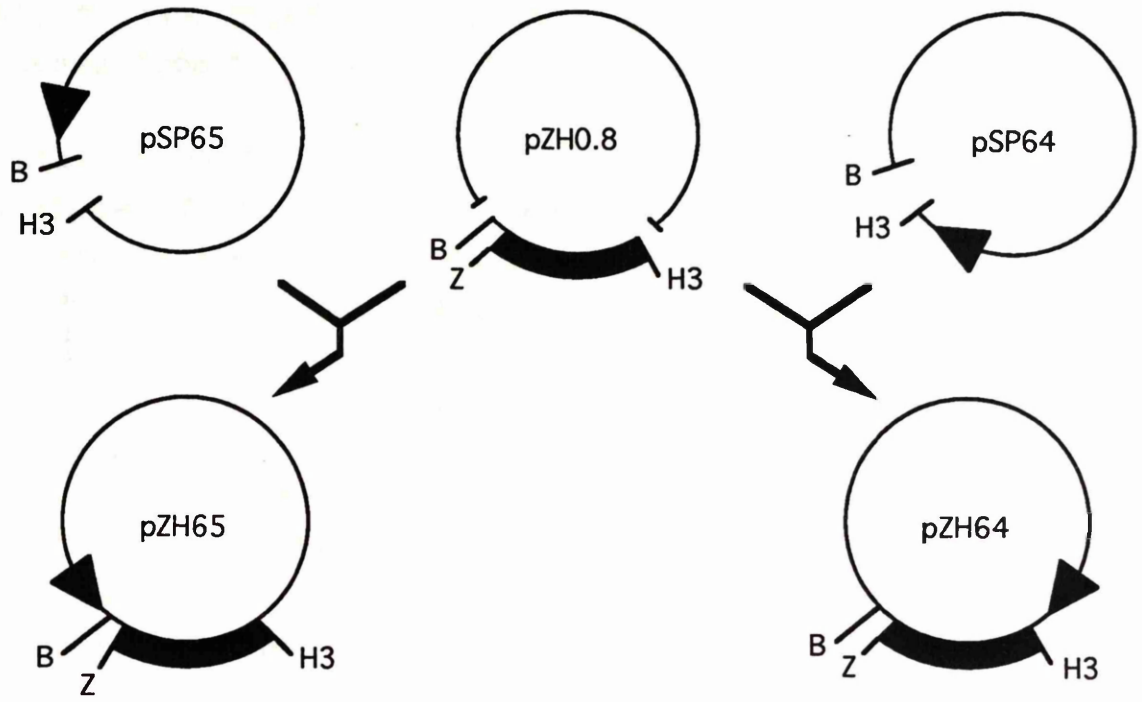
mtDNA insert from pZH0.8 (Fig. 4.2), which stretches from the SstI site (nt 840) within the 12S rRNA gene, to the HindIII site (nt 1603) in the $tRNA^{trp}$ gene, was cut out of the plasmid using the restriction enzymes BamHI and HindIII (Fig. 4.3). The digestion products were separated on an agarose gel, the insert band cut out, and the DNA recovered from the gel using Geneclean (Materials & Methods, section 2.15.1).

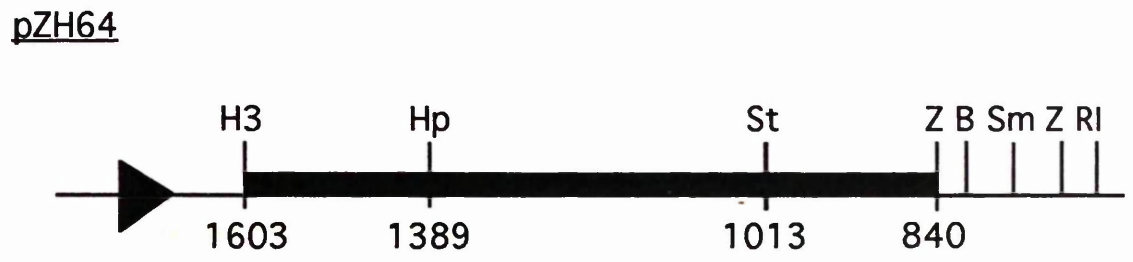
The insert fragment was cloned into the vectors pSP64 and pSP65, which only differ in the orientation of the polylinker sequence with respect to the SP6 promoter. The vectors were both digested with BamHI and HindIII (Fig. 4.3), and the vector DNA was purified away from the small region of polylinker sequence as outlined above for the insert DNA.

The vector and insert DNA fragments were ligated together, transformed into E. coli DS941, and some of the resulting ampicillin-resistant clones screened for the presence of the desired plasmid, by digestion of the DNA from small-scale plasmid (STET) preparations. Although the vectors pSP64 and pSP65 are derived from the pUC plasmids, the majority of the lacZ coding sequence was deleted in their construction (Melton et al, 1984), so the X-gal colour screening test for recombinants cannot be used: hence the reliance on the analysis of plasmid DNA. Plasmid DNA was prepared from 6 clones from each of the 2 transformation reactions, and analysed by BamHI digestion. From the size of the DNA band after agarose gel electrophoresis all the plasmids contained the insert, so a large-scale CsCl gradient plasmid preparation was carried out on one of each of the clones. The maps of the 2 plasmids (named pZH64, and pZH65 for the vectors pSP64, and pSP65, respectively) are shown in Figure 4.3.

4.3.2 THE A6/COIII GENE BOUNDARY.

A 730 bp region of S. purpuratus mtDNA covering the A6/COIII gene boundary, stretching from the PvuII site within A6 (nt





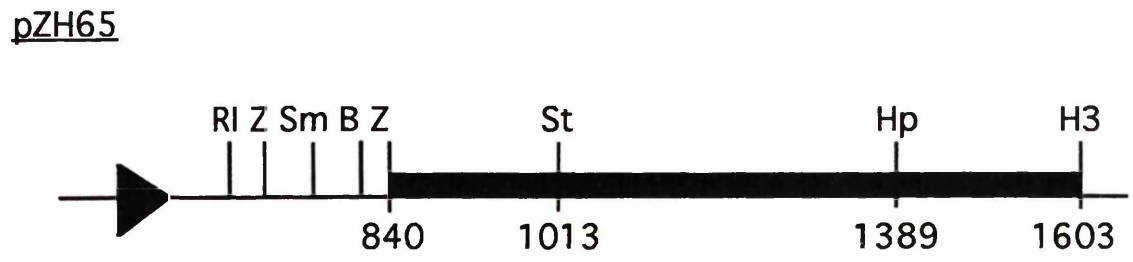


FIGURE 4.3: Schematic diagram showing the cloning strategy used for the construction of the plasmids pZH64 and pZH65, and maps of the mtDNA insert and the adjacent restriction enzyme sites. The coordinates of the mtDNA insert are given according to Jacobs *et al*, 1988. Filled box represents mtDNA insert, and the arrow head, the SP6 promoter, showing the direction of transcription. Restriction enzyme abbreviations: B - *Bam* HI; H3 - *Hin* dIII; Hp - *Hpa* I; RI - *Eco* RI; Sm - *Sma* I; St - *Stu* I; Z - *Sst* I. Diagrams not to scale.

9148), to the EcoRI site in COIII (nt 9778; Fig. 4.2), was subcloned from the plasmid pR3 into the vector pSP64.

The plasmid pSPV3 was created by subcloning the gel-purified PvuII/EcoRI fragment of pR3 containing the mtDNA sequences, into EcoRI/SmaI-digested DNA from the vector pSP64. Digestion with the enzymes PvuII and SmaI both produce blunt-ended molecules, so the 2 sites will ligate together (Fig. 4.4b). After transformation into E. coli DS941, and selection on ampicillin, 12 of the transformants were picked, and plasmid DNA was prepared from them. HindIII digests revealed that 5 of the clones contained the insert, with the others being pSP64 religated to the polylinker fragment. Further restriction enzyme digests (EcoRI, EcoRI/BamHI, BamHI, and SstI) confirmed the presence of the insert (Fig. 4.4a). Plasmid DNA was then prepared from one of these clones by CsCl-gradient centrifugation. In the plasmid pSPV3, transcription through the mtDNA insert from the SP6 promoter occurs in the same direction as transcription through this region of the genome in vivo.

pSPV31, in which the mtDNA insert is inverted in its orientation compared to pSPV3, was produced following the same procedure as for pSPV3, except that the insert was removed from pR3 by a HindIII/PvuII double-digest. This DNA fragment, which also contains the pUC8 polylinker sequences, was subcloned into HindIII/SmaI-cut pSP64 (Fig. 4.4c). HindIII digests of 12 transformants revealed 10 containing the insert; 2 of these were further analysed by 4 different enzyme digests (EcoRI, SstI, SstI/HindIII, and XbaI), confirming that the insert had been subcloned as desired (Fig.4.4a). The maps of the plasmids pSPV3 and pSPV31 are shown in Figures 4.4b and 4.4c respectively.

With the RNase protection technique, co-terminal probes of different lengths are needed to map the end points of any molecules present to a specific location within the genome. For the 2 constructs, pZH64 and pZH65, which cover the origin region, the presence of HpaI and StuI restriction sites

a)

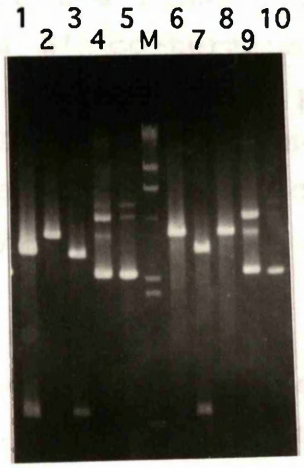
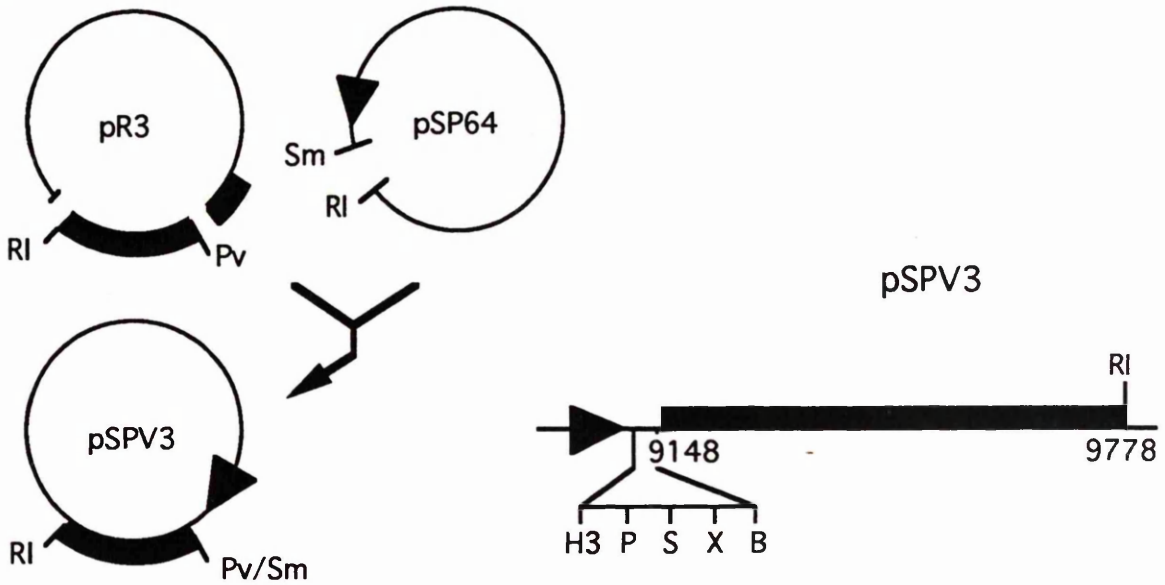
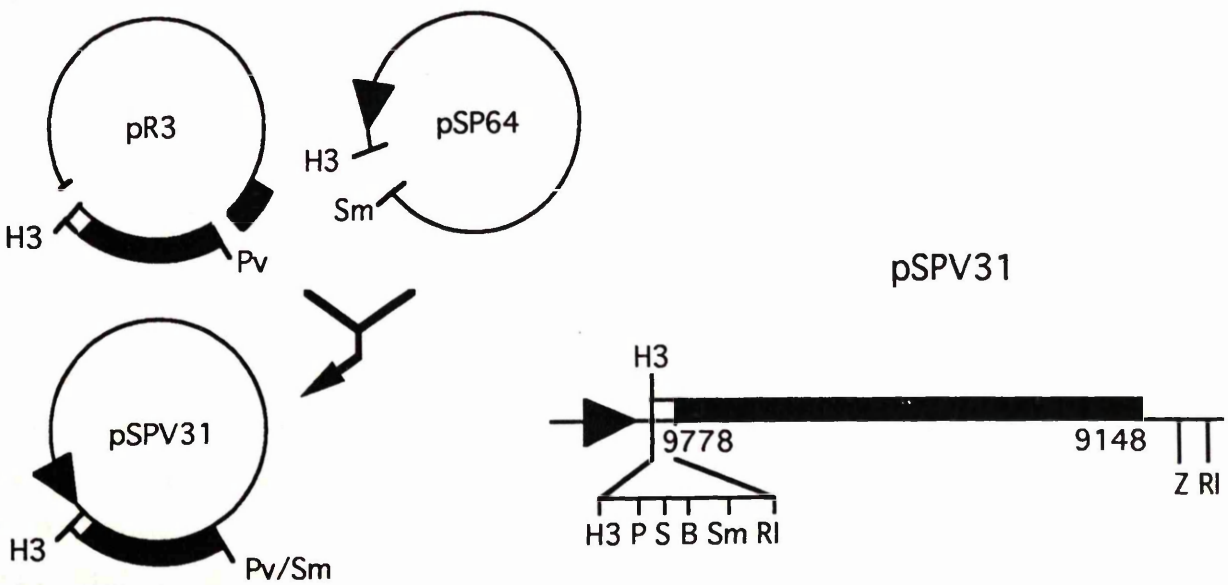


FIGURE 4.4: (a) EtBr-stained agarose gel showing electrophoresis of plasmid DNA isolated from colonies obtained from the subcloning reactions of the mtDNA insert of pR3 into the vector pSP64. Lanes 1-5: pSPV31: 1 - RI; 2 - Z; 3 - Z/H3; 4 - X; 5 - uncut. Lanes 6-10: pSPV3: 6 - B; 7 - B/RI; 8 - RI; 9 - Z; 10 - uncut. M - H3-digested λ DNA (size marker).(b,c) Schematic digrams (not to scale) showing the cloning strategy used for the construction of the plasmids (b) pSPV3 and (c) pSPV31, with maps of the mtDNA insert and the adjacent restriction enzyme sites. See legend to Figure 4.3 for explanation of symbols, and restriction enzyme abbreviations. Further restriction enzyme abbreviations: K-*Kpn* I; P -*Pst*I; S - *Sal* I; X - *Xba* I.

b)



c)



within the mtDNA insert (Fig. 4.3) allows the production of three 5′ co-terminal probes of different lengths. However, for the A6/COIII pause site region, an exonuclease III (exoIII) deletion series was made in order to generate the series of shorter clones required for the RNase protection analysis.

The starting point for the exoIII deletion series was the plasmid pSPV31. This was first digested with PstI, and then BamHI to generate a linear plasmid with one 3′ and one 5′ protruding end. Exonuclease III cannot attack DNA with a 3′ protruding end, so this confined the action of exoIII to the mtDNA insert. ExoIII digestion was carried out at 30°C, and aliquots were removed from the reaction at 90 second intervals for 6 time points (up to and including 9 mins). After treatment with mung bean nuclease, to digest away the single-stranded DNA left by the exoIII, aliquots from each time point were analysed by agarose gel electrophoresis to check the extent of the deletions (Fig. 4.5a). The gel showed that about 500 nt had been deleted at the final time point (lane 6), leaving about 200 nt of the mtDNA insert intact.

The DNA from the different time points was self-ligated and transformed into E. coli DS941. Six colonies from each of the time points 3 to 6 were picked, and their plasmid DNA analysed by restriction enzyme digestion. Double-digests with HindIII and SstI cut out the inserts, revealing a degree of heterogeneity in the extent of the deletions, especially for time points 4 and 6 (Fig. 4.5b). From these clones, 5 were chosen (designated pAM35/43/55/62/65) which gave an approximately uniform spread of insert sizes, and CsCl-gradient purified plasmid DNA was prepared from each of them.

The mtDNA insert from each of these plasmids was then recloned into the vector pSP65, to change its orientation with respect to the SP6 promoter. For each member of the deletion series, the insert DNA was excised from the vector by a HindIII/SstI double-digest, purified on an LMP agarose gel, and ligated to HindIII/SstI-digested pSP65 DNA. Plasmid

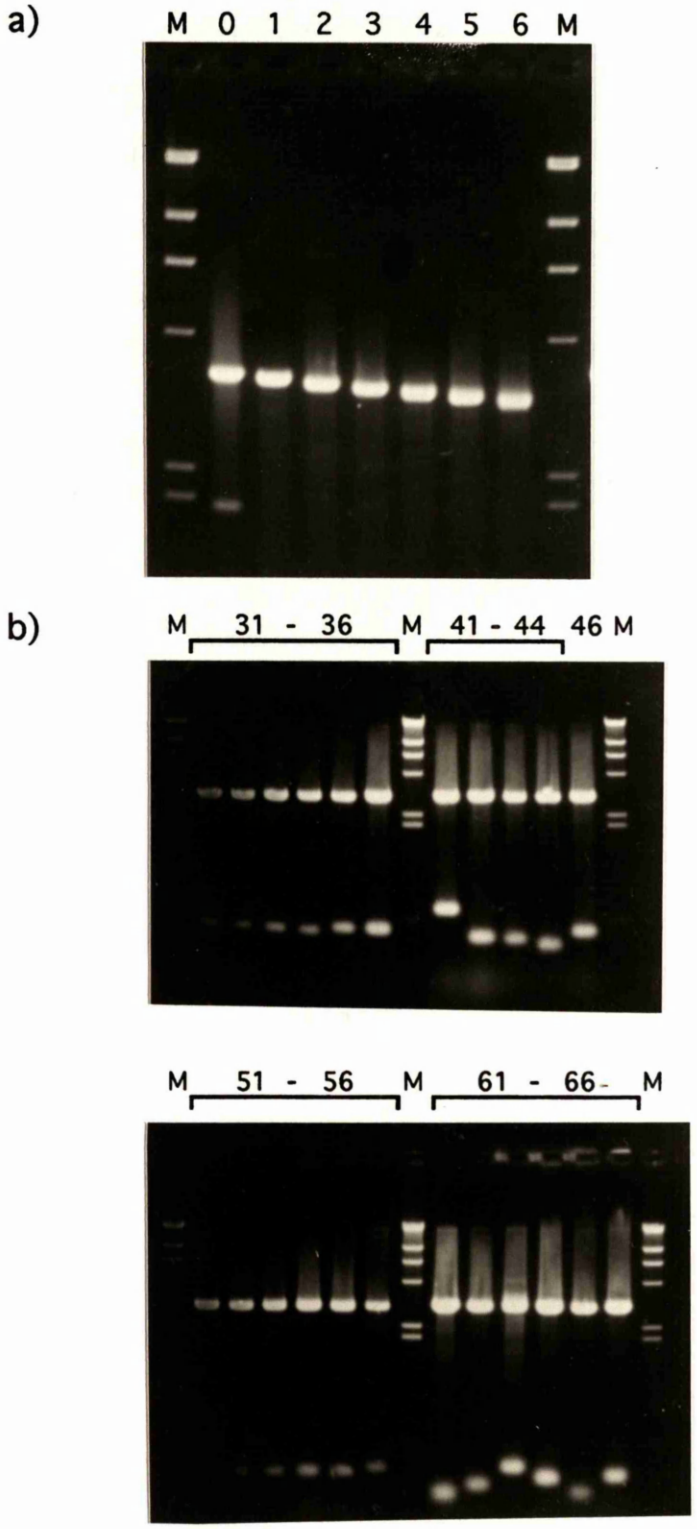


FIGURE 4.5: (a) EtBr-stained agarose gel showing electrophoresis of aliquots from each of the exoIII deletion time-point reactions (lanes 1-6), linearised pSPV31 DNA (lane 0), and *Hin* dIII-digested λ DNA, used as size markers (M). (b) EtBr-stained agarose gel showing *Hin* dIII/*Sst* I-digested plasmid DNA from colonies obtained following the transformation of self-ligated DNA from the exoIII reactions into *E. coli* DS941. DNA from 6 colonies from each of the time-points 3-6 was analysed, and the samples were designated 31-36, 41-46, 51-56, and 61-66. *Hin* dIII-digested λ DNA was used as a size marker (M).

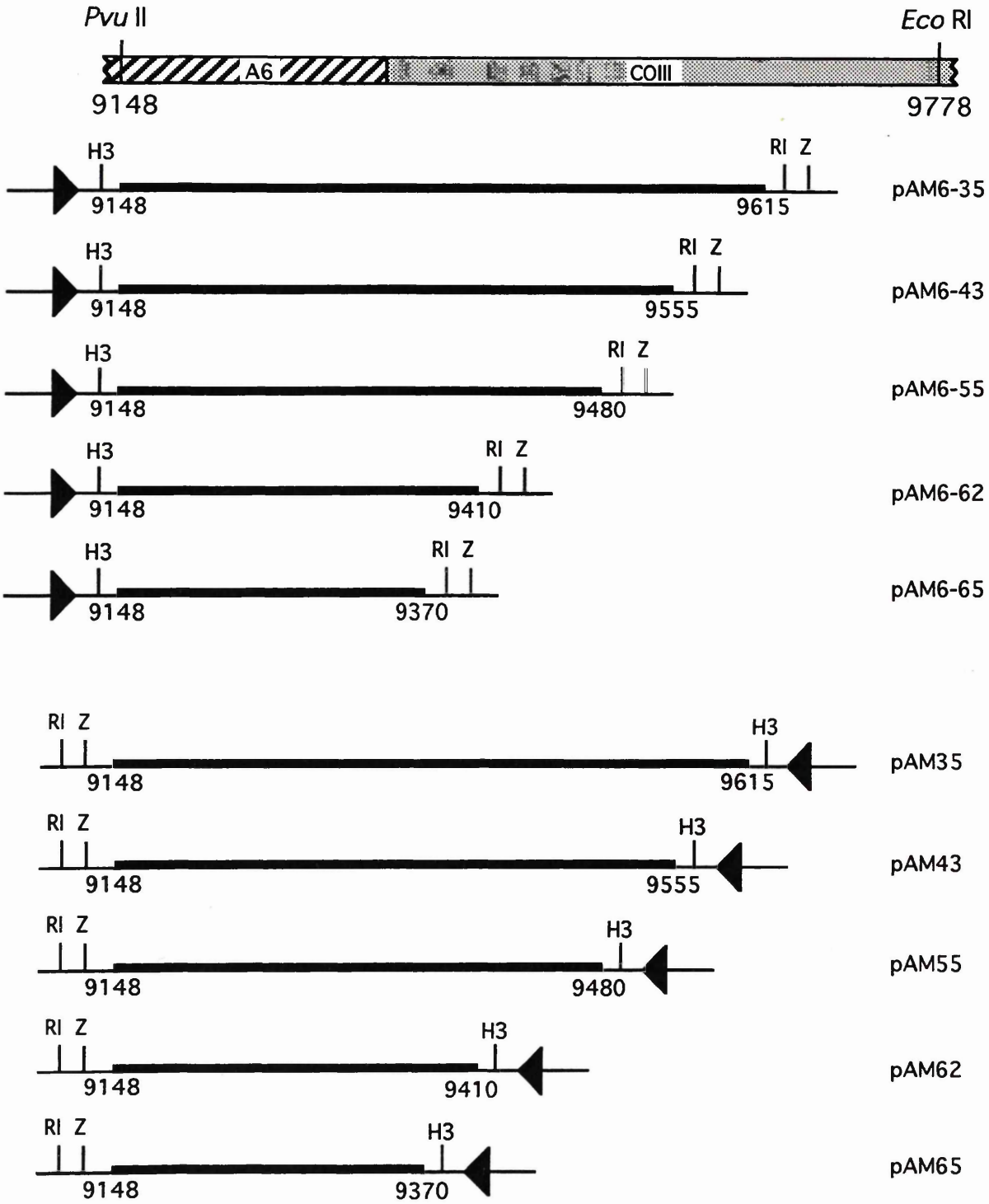


FIGURE 4.6: Plasmid maps of the members of the pAM35 and pAM6-35 exoIII deletion series, showing the location and direction of transcription of the SP6 promoter (filled arrow head), the adjacent restriction enzyme sites, and the extent of the insert (thick line). Diagrams not to scale. Coordinates of the end of the mtDNA inserts are ± 5 nt.

(facing page 114)

DNA from the transformants was analysed by HindIII digestion, which linearised the plasmids, therefore allowing the presence of the insert DNA to be detected. These plasmids were named pAM6-35, through to pAM6-65.

The approximate sizes of the mtDNA inserts present in the exoIII deletion clones were deduced by analysis of the RNA products generated by SP6 transcription of linearised plasmids. These transcripts were analysed by denaturing polyacrylamide gel electrophoresis, taking into account the lengths of the polylinker sequences when calculating the sizes of the mtDNA inserts. The plasmid maps, indicating the extents of the deletions, are shown in Figure 4.6.

4.4 RNase PROTECTION ANALYSIS OF THE NON-CODING REGION OF mtDNA.

4.4.1 MAPPING THE ENDS OF DNA MOLECULES.

RNA probes complementary to the leading- and lagging-strands of mtDNA at the non-coding region were prepared by *in vitro* transcription of HindIII-digested pZH65 and BamHI-digested pZH64 DNA, respectively (Fig. 4.3). Both probes spanned tRNA genes on either side of the non-coding region, stretching from the SstI site in the 12S rRNA gene (nt 840) to the HindIII site in the gene for $tRNA^{trp}$ (nt 1603; Fig. 4.2).

The 2 probes were hybridised to RNase-treated AEEF mtDNA, and to plasmid DNA containing the same mtDNA insert as the probe. The latter reactions were carried out to investigate whether the formation of an RNA/DNA hybrid by the region of mtDNA under study would produce structures containing single-stranded regions of probe. Such regions would be open to RNase digestion, generating bands on electrophoresis of the reaction products which did not correspond to the ends of DNA molecules. Any bands in common between the reactions involving plasmid DNA and those with AEEF mtDNA could therefore be discounted as being artifactual.

Hybridisation of the pZH64 and pZH65 probes to AEEF mtDNA in each case produced a band of 310 ± 10 nt (Fig. 4.7a). However, hybridisation of the pZH64 probe to the plasmid pZH65 produced a major band of 500 ± 10 nt, rather than one at 769 nt (the size expected from complete protection of the probe molecules from RNase digestion). A faint "full-length" band of 769 nt was present, as was a smaller one of 380 nt. The 500 nt band was not, however, simply due to the presence and hybridisation of large amounts of non-full-length probe, since the unreacted probe was mainly full-length (Fig. 4.7a; lane 1).

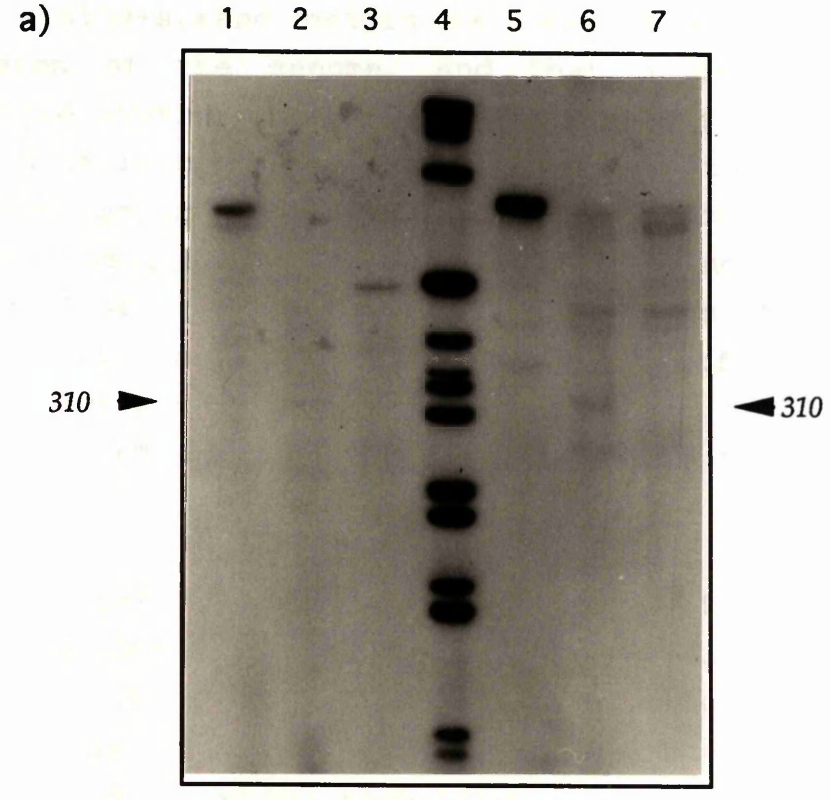
The presence of this major 500 nt reaction product also cannot be explained as being due to a single unpaired site in the DNA/RNA hybrid. Although RNase digestion, in such a situation, could generate a 500 nt product, the reciprocal portion of the probe (a 270 nt molecule) would be present in equal amounts, which was not seen. The absence of a reciprocal band suggests that there are many stem-loop structures and/or (just conceivably) small deletions present, generating several smaller bands, which are not detected. Alternatively, the formation of a large, single-stranded loop of probe by mispairing could leave a large part of the probe susceptible to RNase digestion, resulting in the formation of one major RNA product.

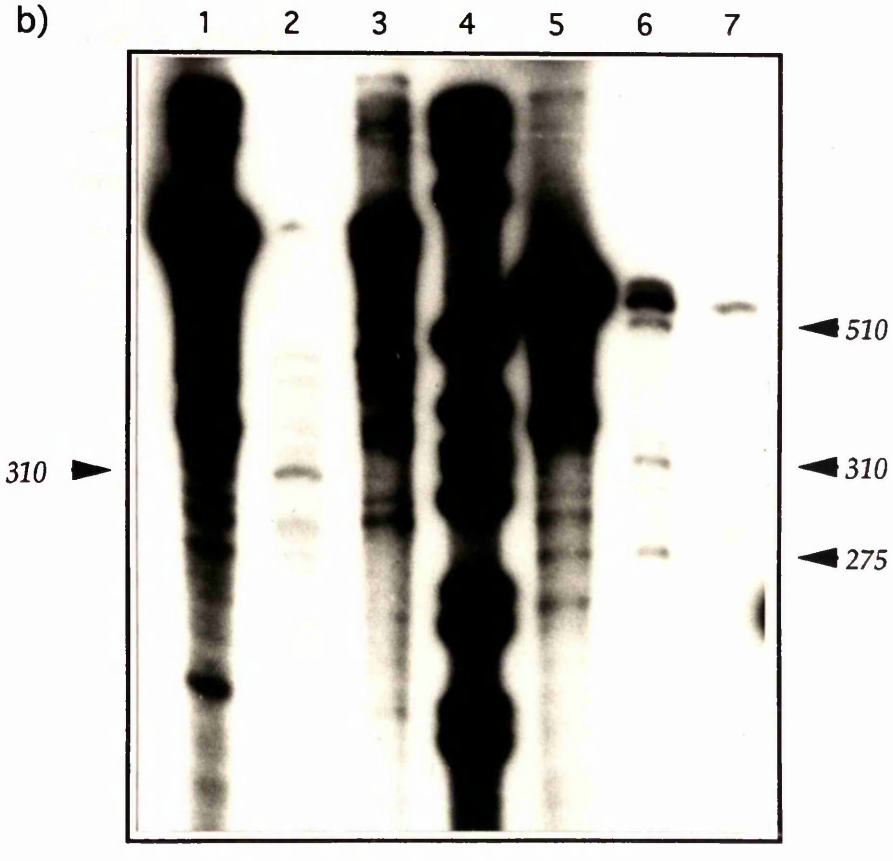
With the pZH65 probe, the situation is complicated by the presence of several bands observed in the products of the reactions with both the mtDNA and plasmid (Fig. 4.7a; lanes 6,7). All of them, except the 310 nt band, were present in both lanes, which suggests that only the 310 nt band represents the end of a DNA molecule found <u>in vivo</u> within this region of the genome. The other bands are presumably due to the nicking of the probe molecules in regions of unpaired secondary structure. The presence of bands around 480 nt and 270 nt supports this idea, since together their sizes are approximately equal to that of the full-length probe.

The gel electrophoretic data presented in Chapter 3 implied

FIGURE 4.7: Autoradiograph showing the results of polyacrlyamide/urea gel electrophoresis of RNase protection reactions of AEEF mtDNA. (a) Probes synthesised from *Hin*dIII-digested pZH65 DNA and *Bam* HI-digested pZH64 DNA: lane 1 - unreacted pZH64 probe RNA; 2 - pZH64 RNA hybridised to AEEF mtDNA; 3 - pZH64 RNA hybridised to pZH64 plasmid DNA; 4 - BRL 1 kb ladder; 5 - unreacted pZH65 probe RNA; 6 - pZH65 RNA hybridised to AEEF mtDNA; 7 - pZH65 RNA hybridised to pZH64 plasmid DNA. (b) Probes synthesised from *Hin* dIII-digested and *Hpa* I-digested pZH65 DNA: 1 - unreacted *Hin* dIII probe RNA; 2 - *Hin* dIII probe hybridised to AEEF mtDNA; 3 - *Hin* dIII probe hybridised to pZH64 plasmid DNA; 4 - BRL 1 kb ladder; 5 - unreacted *Hpa* I probe RNA; 6 - *Hpa* I probe hybridised to AEEF mtDNA; 7 - *Hpa* I probe hybridised to pZH64 plasmid DNA. The sizes and positions of the RNA bands, produced as a result of protection of the RNA probes by the AEEF mtDNA sample, are indicated on each side of the autoradiographs.

(facing page 116)


a)
1 2 3 4 5 6 7
310
310


b)
1 2 3 4 5 6 7
510
310
310
275

that the leading-strand origin is located within the non-coding region of the genome, and that a pause site for lagging-strand synthesis also occurs at this site. The RNase protection data is consistent with these results, in that the 310 nt bands, generated in the RNase protection reactions with the 2 probes, would be produced by DNA molecules whose ends mapped close to nt 1150, 310 nt away from the SstI site at nt 840. This locates the 5′ and 3′ ends of the leading- and lagging-strand molecules, respectively, to just 5′ of the end of the run of consecutive G residues in the non-coding region.

In order to confirm the position of the 5′ end of the leading-strand DNA molecule, RNase protection experiments were carried out using 2 shorter, 5′ co-terminal probes synthesised from pZH65 DNA. The plasmid was digested with either StuI or HpaI, which each have one recognition site in the mtDNA insert (Fig. 4.3), and RNA probes were synthesised by *in vitro* transcription.

As with the probe synthesised from HindIII-digested pZH65 (the HindIII probe), minor RNA products, both shorter and longer than the probe, were observed when samples of the unreacted probe were subjected to gel electrophoresis. The longer transcripts were probably due to the presence of a small amount of uncut plasmid in the transcription reaction. These plasmid molecules would also be transcribed, but they do not have a cut end to terminate transcription, so longer transcripts could be produced. The generation of discrete bands may be due to the presence of regions of secondary structure or SP6-like termination sequences within the plasmid.

The shorter probe bands may represent premature termination products caused by stalling of the polymerase on the template DNA. The most prominent of these seen with the HindIII probe (380 nt) was also seen with the probe from HpaI-digested pZH65 DNA. This suggests that termination is occurring at the same sequence in the mtDNA insert, implying that some

structural or sequence characteristic induces SP6 RNA polymerase to stop at this site. The presence of shorter RNA probes does complicate the analysis, since their complete or partial protection on hybridisation to the mtDNA samples leads to the formation of spurious shorter bands. However, the relative amounts of these shorter probe molecules is small compared to the full length product, and the 310 nt band specific to the mtDNA lane is much more intense than any of the other bands, so its generation from a prematurely-terminated probe can probably be discounted.

The hybridisation of the RNA probe synthesised from HpaI-digested pZH65 DNA (referred to below as the HpaI probe) to AEEF mtDNA produced 3 bands smaller than the fully-protected probe, which were all more intense than the background of spurious bands (also seen in the control hybridisation to plasmid DNA), and absent from the hybridisation reaction between the probe and HindIII-digested plasmid DNA (Fig. 4.7b; lanes 5,6). The bands were 235 ± 10, 310 ± 10, and 510 ± 10 nt in size. In contrast, a preliminary experiment using the StuI probe only produced 2 bands on hybridisation to the AEEF mtDNA sample (data not shown). These were 165 ± 10 and 210 ± 10 nt in length, which corresponds to the sizes of the fully-protected probe molecule, and unreacted probe, respectively. However, the presence of a band which probably represents undigested probe suggests that the RNase digestion reaction was not totally efficient.

The results with the HpaI and StuI probes support the interpretation of the data obtained using the probe generated from HindIII-digested pZH65 DNA. Nascent DNA molecules with their 5′ ends mapping to nt 1150 would generate a 310 nt band with the HpaI probe (Fig. 4.8a), and protect the StuI probe completely, generating a 165 nt band. This interpretation is also consistent with the mapping of the 5′ ends of D-strand molecules in S. purpuratus egg mtDNA to nt 1159 ± 5 (Jacobs et al, 1989).

It is, however, theoretically possible that the 310 nt band

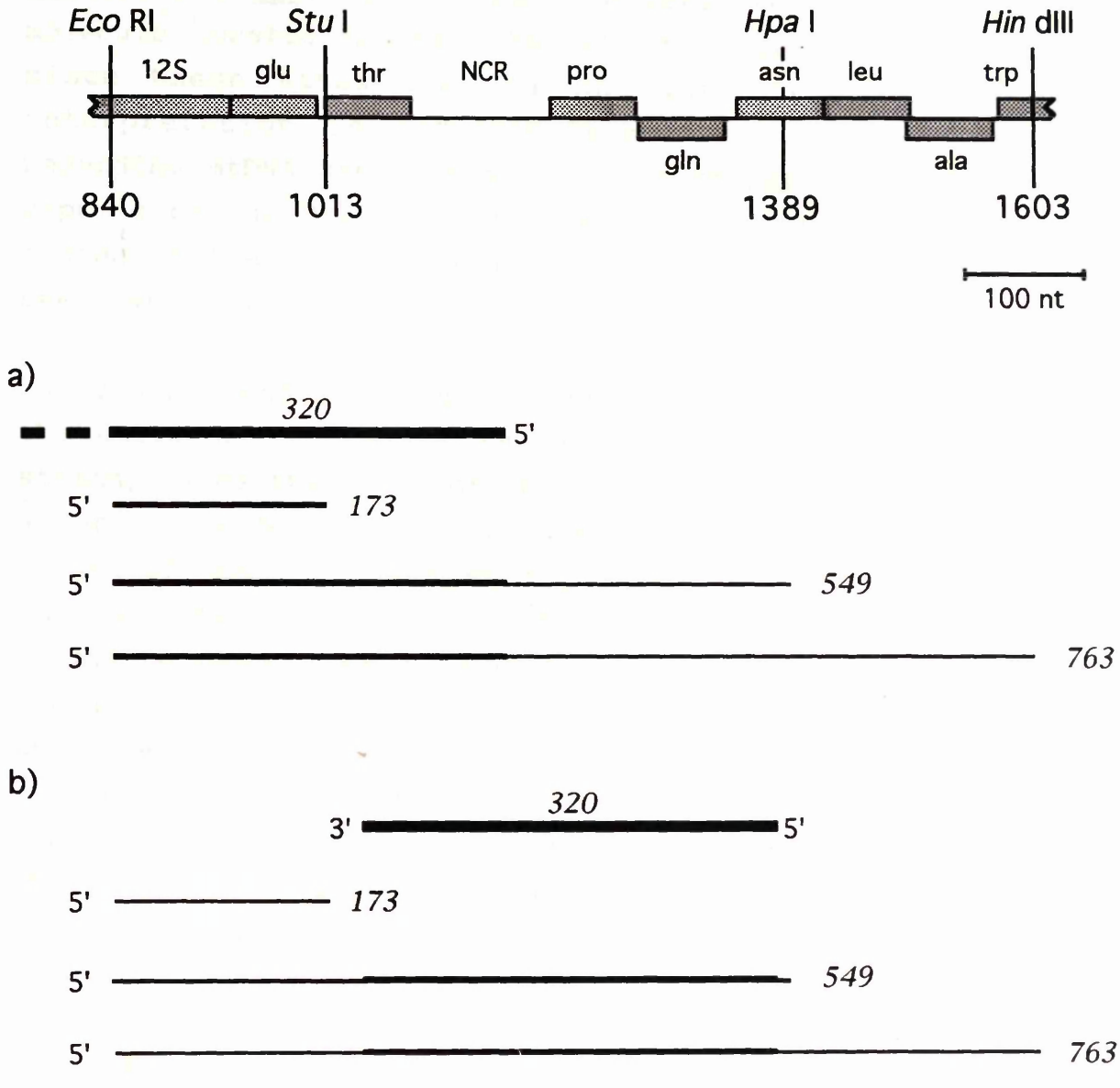


FIGURE 4.8: Diagram showing 2 possible interpretations (referred to as (a) and (b) in the text) of the RNase protection results obtained with the 3 different length probes synthesised from pZH65 DNA. The map of the genome shows the relevant enzyme sites, with coordinates; the tRNA genes are identified by the 3 letter amino acid code; genes encoded by the major coding-strand are shown above, and the minor coding-strand, below. NCR - non-coding region. The thick lines represent DNA molecules; thin lines, RNA probes (excluding polylinker sequences); and intermediate lines show the proposed regions of protection from RNase T1 digestion by hybridisation of the RNA to the DNA. The sizes of the different RNA probes, and the DNA molecules are given in nt. See text for explanation.

observed in the RNase protection experiments with both the HindIII and HpaI probes could be generated by a 310 nt DNA molecule located entirely between the HpaI and StuI sites, since these sites are 374 nt apart (Fig. 4.8b). This interpretation is not consistent with the other data regarding mtDNA replication in S. purpuratus, including the mapping of the D-loop in egg mtDNA entirely within the non-coding of the genome (Jacobs et al, 1989), and consequently seems most unlikely.

The 235 nt band observed with the HpaI probe could represent the unligated 3′ end of the completely replicated leading-strand, since the sizes of the 2 bands taken to represent the 5′ and 3′ ends of the DNA molecules are about equal to the length of the mtDNA portion of the probe (310 + 235 = 545 nt; approximately equal to 547 nt, the size of the probe excluding polylinker sequences). However, the 3′ end of such a DNA strand would be expected to give a band of 453 nt with the HindIII probe, but one was not observed, although the presence of spurious bands in this region of the gel may have obscured it. Alternatively, the 235 nt band could represent the 5′ end of another, shorter, leading-strand DNA molecule, although again it should have been detected by the HindIII probe, which was not the case. As no spurious bands were observed in the region of the gel around 235 nt, I favour the former interpretation of this result. The 510 ± 10 nt product generated by the reaction with the HpaI probe is discussed below (section 4.6.1).

4.4.2 TRANSCRIPT MAPPING.

When the riboprobes produced from BamHI-digested pZH64 and HindIII-digested pZH65 were hybridised to total nucleic acid isolated from AEEFs, the results were rather different from those obtained with RNase-treated AEEF mtDNA. Both reactions generated product bands smaller than the full-length probe (Fig. 4.9), but they did not correspond to those observed in the hybridisations with the mtDNA sample, described above.

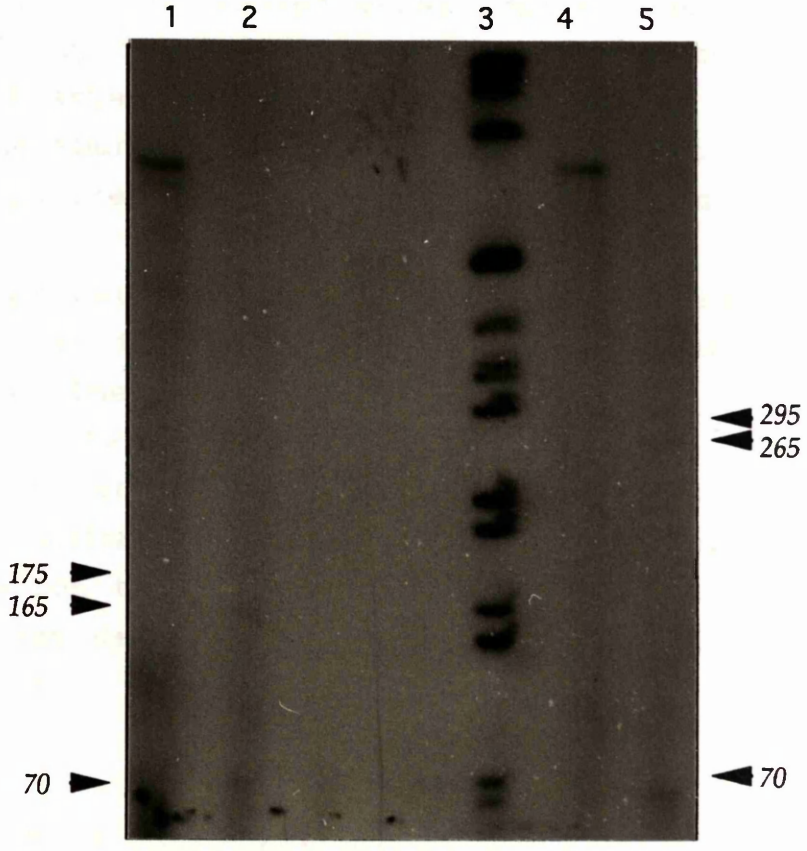


FIGURE 4.9: Autoradiograph showing the results of polyacrlyamide/urea gel electrophoresis of RNase protection reactions between AEEF nucleic acid and RNA probes covering the non-coding region of the genome: lane 1 - unreacted pZH64 probe RNA; 2 - pZH64 RNA hybridised to AEEF nucleic acid; 3 - BRL 1 kb ladder; 4 - unreacted pZH65 probe RNA; 5 - pZH65 RNA hybridised to AEEF nucleic acid. The sizes and positions of the bands generated by protection of the RNA probes by the AEEF nucleic acid samples are indicated on each side of the autoradiograph.

The RNA probe from pZH64, which should detect the DNA lagging-strand, and the transcripts from the 12S rRNA gene, and most of the tRNAs encoded in this region, revealed bands of 175 ± 10 and 160 ± 10 nt, and a faint one of 70 ± 5 nt. The complementary RNA probe from the plasmid pZH65, which should hybridise to the DNA leading-strand, and the tRNAs for alanine, and glutamine, gave bands of 295, 265, and 70 nt. The absence of these bands in products obtained from the reactions with RNase-treated mtDNA implies that these bands are due to hybridisation to RNA, and not DNA molecules. The failure to observe bands from the DNA molecules is probably a result of the abundance of the RNA, compared to the DNA, and the short exposure times for the autoradiographs.

The 70 nt products are almost certainly the result of hybridisation to tRNA molecules, which are transcribed from both strands. The D-loop molecules detected in egg mtDNA (Jacobs et al, 1989), were also 65-75 nt long, so they may also have been detected by the pZH65 probe, but the products were not resolved under gel electrophoresis from those generated by hybridisation to the tRNAs. The possible identity of the other molecules detected by the 2 probes is discussed below (section 4.6.2).

4.5 RNase PROTECTION ANALYSIS OF A PAUSE SITE REGION IN mtDNA.

The pause site region of the mitochondrial genome, covering the A6/COIII gene boundary (Fig. 4.2) was studied using RNA probes produced from the 2 families of plasmids generated by exoIII deletion of the plasmid pSPV31 (section 4.3.2). Since the probes spanned 2 genes, both the RNA transcripts, and the replicating DNA molecules were studied.

4.5.1 TRANSCRIPT MAPPING.

Initially, total nucleic acid isolated from S. purpuratus

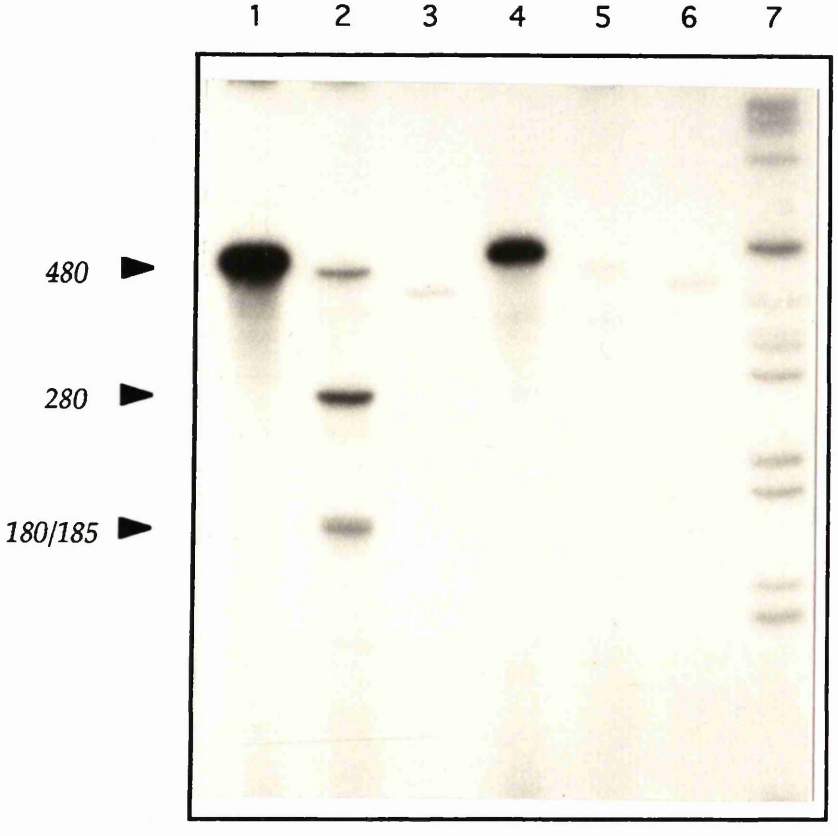


FIGURE 4.10: Autoradiograph showing the results of polyacrlyamide/urea gel electrophoresis of RNase protection reactions between AEEF nucleic acid and RNA probes covering the A6/COIII gene junction: lane 1 - unreacted pAM35 probe RNA; 2 - pAM35 RNA hybridised to AEEF nucleic acid; 3 - pAM35 RNA hybridised to pAM43 plasmid DNA; 7 - BRL 1 kb ladder; 4 - unreacted pAM6-35 probe RNA; 5 - pAM6-35 RNA hybridised to AEEF nucleic acid; 6 - pAM35 RNA hybridised to pAM43 plasmid DNA. The arrows show the positions and sizes (in nt) of the bands discussed in the text.

(facing page 123)

AEEFs was hybridised to 2 complementary RNA probes transcribed from EcoRI-digested pAM35, and HindIII-digested pAM6-35. These probes are complementary, repectively, to the sense, and antisense transcripts of the A6 and COIII genes. The resulting autoradiograph from the RNase protection reactions (Fig. 4.10) clearly showed 3 major bands with the pAM35 probe, and only one with pAM6-35.

The RNA probe from pAM35 gave bands of 480 ± 10, 280 ± 10, and a doublet of one major and one minor band of 180/185 ± 10 nt, respectively, as well as some faint minor bands. The 480 nt band probably represents fully-protected probe, and the smaller bands were generated by hybridisation to 2 molecules with termini lying within the region. From their sizes, it is probable that together these molecules span the region, since their total length is approximately equal to that of the probe (280 + 180 = 460). The RNA probe from pAM6-35, which would detect the antisense RNA transcripts, and the DNA leading-strand, only gave a band of 480 ± 10 nt (fully-protected probe).

Two questions arise from these results; firstly, are the protecting molecules RNA or DNA? And secondly, where within this region of the genome are the ends of the molecules located?

To answer both of these questions, probes were made from the plasmids pAM55, and pAM6-55, members of the pSPV31 deletion series with smaller mtDNA inserts. They were hybridised to total AEEF nucleic acid, and to samples which had been pre-treated with either RNase A, or DNase, prior to the hybridisation reaction.

The results again showed 3 bands when total nucleic acid was hybridised to the lagging-strand/sense probe, pAM55 (Fig. 4.11). The fully-protected probe gave a band of 330 ± 10 nt, and the 2 smaller bands were 180 ± 10, and 140 ± 10 nt in size. These 2 bands were absent in the lane containing the RNase A-treated sample, but were unaffected by DNase pre-

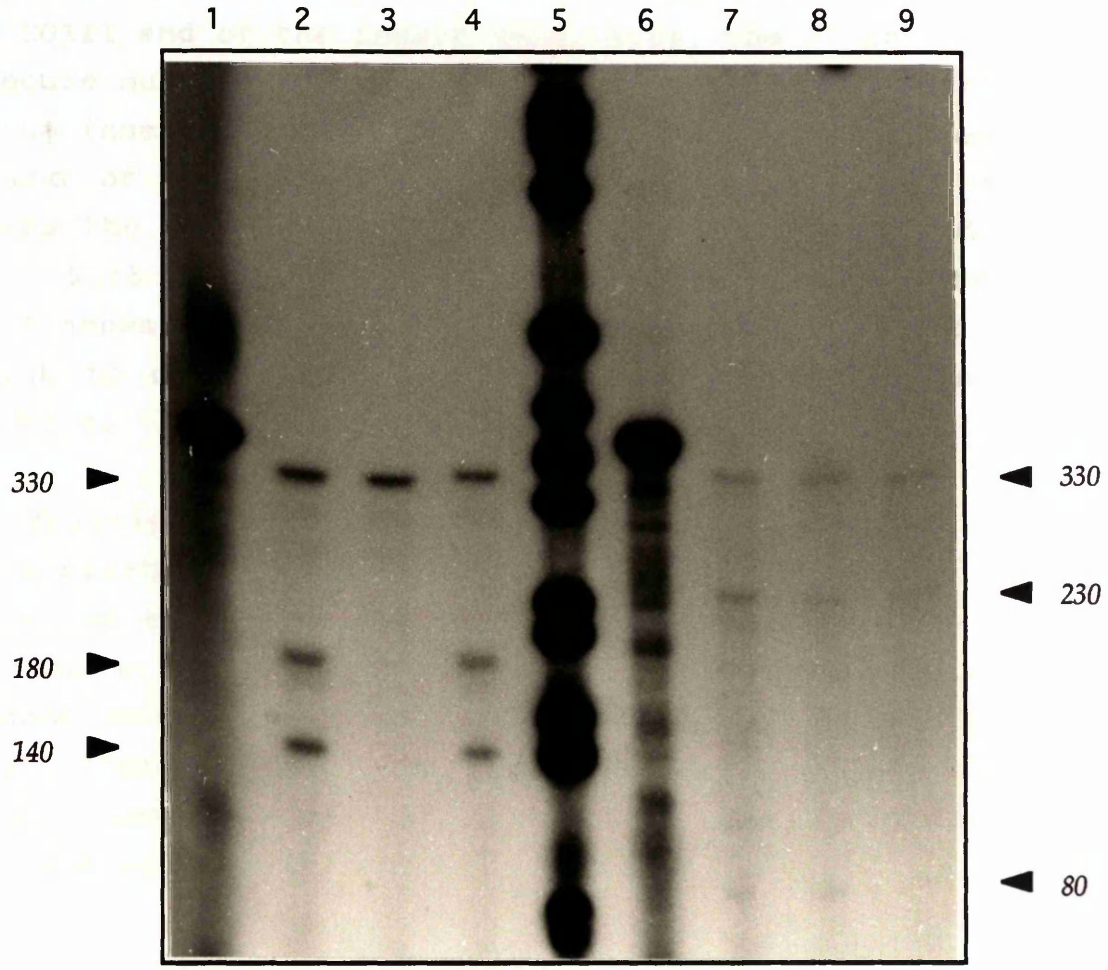


FIGURE 4.11: Autoradiograph showing the results of polyacrlyamide/urea gel electrophoresis of RNase protection reactions between AEEF nucleic acid and RNA probes covering the A6/COIII gene junction: lane 1 - unreacted pAM55 probe RNA; 2 - pAM55 RNA hybridised to AEEF nucleic acid; 3 - pAM55 RNA hybridised to RNase A pretreated AEEF nucleic acid; 4 - pAM55 RNA hybridised to DNase I pretreated AEEF nucleic acid; 5 - BRL 1 kb ladder; 6 - unreacted pAM6-55 probe RNA; 7 - pAM6-55 RNA hybridised to AEEF nucleic acid; 8 - pAM6-55 RNA hybridised to RNase A pretreated AEEF nucleic acid; 9 - pAM6-55 RNA hybridised to DNase I pretreated AEEF nucleic acid.

treatment (Fig. 4.11; lanes 3,4), clearly demonstrating that they were the result of hybridisation of the probe to RNA. Comparison of the band sizes with those detected by the pAM35 probe revealed that the 180 nt product band is common to the reactions with both probes. Since the deletions extend from the COIII end of the insert (Fig. 4.6), the 3′ end of the RNA molecule must map 180 nt from the PvuII site at the other end of the insert, which locates it at nt 9328 ± 10, close to the 3′ end of the A6 coding sequence. The other RNA molecule covers the COIII end of the insert, with its 5′ end mapping to nt 9335 ± 10. This maps the 2 transcripts to the A6 and COIII genes, respectively, although the data are not accurate enough to determine whether or not the transcripts overlap slightly.

The full-length product observed in the reaction with the DNase pre-treated sample (Fig. 4.11; lane 4) could have been generated by a precursor RNA molecule which completely spans the region covered by the probe. Alternatively, the 480 nt product could have been the result of hybridisation to the mtDNA, indicating that the DNase treatment was ineffective: a view supported by the results obtained using the complementary probe pAM6-55 (see below).

The RNA probe from the pAM6-55 plasmid produces a fully-protected probe band in the hybridisation reactions (Fig. 4.11), as well as 2 shorter bands of 230 ± 10, and 80 ± 10 nt. Their presence in both the RNase-pretreated and untreated samples, and the fact that the sum of their lengths is approximately equal to that of the probe (230 + 80 = 310), suggests that the bands are formed as the result of the structure of the RNA/DNA hybrid allowing RNase digestion of the probe. The production of the same 3 bands on the hybridisation of the pAM6-55 probe to its plasmid DNA (data not shown) strongly supports this interpretation. The occurrence of all 3 reaction products with the DNase pre-treated sample (Fig. 4.11; lane 9), when no product bands would have been predicted, implies that the DNase treatment was ineffective.

Two other pairs of probes from the same deletion series were also used to map the transcripts, and the results obtained were consistent with those described above.

4.5.2 MAPPING THE ENDS OF DNA MOLECULES.

The analysis of replicating mtDNA by agarose gel electrophoresis, discussed in Chapter 3, points to the presence of a replication pause site, and a prominent lagging-strand replication origin near the boundary of the A6 and COIII genes. RNase protection experiments were therefore carried out on RNase-treated AEEF mtDNA, using probes pAM35, and pAM6-35 to map any DNA ends present within this region.

The probes pAM35, and pAM6-35 did not, however, detect any bands, except for the fully-protected probe molecule, above a background of faint bands possibly due to nicking of the RNA probes, and the hybridisation of shorter probe products. Three different interpretations could be placed on this result. Firstly, it could indicate that the lagging-strand origin and the pause site lie outside the region covered by the probe, which would imply that the sizing from the 2D gels is less accurate than originally thought. I consider this unlikely, since it would put the origin and pause site either upstream of nt 9148, or downstream of nt 9620, more than 150 or 320 nt away from the predicted site, respectively. This contradicts the evidence from the 1D gel analyses of AEEF mtDNA, which mapped the pause site to nt 9300 $\pm$ 50. The 2D gel system was also used to map the leading-strand replication origin, in conjunction with the 1D gels of native and denatured mtDNA, and again the results obtained were consistent with the position of the origin implied from the RNase protection studies of the non-coding region described above (section 4.4.1).

Assuming, therefore, that the experiments described in Chapter 3 are reliable, a second interpretation of the results is that the technique used is not sensitive enough to

map the position of the pause site or the lagging-strand origin, due to the low abundance of molecules with termini at these positions. A third interpretation could, however, be applied to the failure to detect discrete ends of DNA molecules in this region, in that it could be due to a large degree of heterogeneity at the nucleotide level in the positions of the pause site and the second-strand origin. Although the ends of DNA molecules were detected in the non-coding region, their intensity was not particularly great (less than 10-fold more intense than background). The 5′ DNA end in the non-coding region represents the initiation point of leading-strand DNA synthesis of all the replicating molecules in the sample, whereas, on the basis of the 2D gel analysis, the 3′ ends of nascent DNA strands are probably present at the pause site in at most 10% of these molecules. Molecules initiating at this lagging-strand origin may also be present in only about 10% of all molecules, which would make their detection above background improbable using this method, unless a much greater amount of DNA could be used without any corresponding increase in the background signal.

4.6 DISCUSSION.

4.6.1 MAPPING THE ENDS OF DNA MOLECULES IN THE NON-CODING REGION.

Three 5′ co-terminal RNA probes of different lengths were used to map the 5′ ends of the leading-strand DNA molecules. The probes were produced by *in vitro* transcription of HindIII-, HpaI- and StuI-digested pZH65 DNA (Fig 4.3), and were referred to as the HindIII, HpaI and StuI probes above. On hybridisation to AEEF mtDNA, the HindIII and HpaI probes both produced bands of 310 ± 10 nt, in addition to a larger band representing fully-protected probe molecules, whereas the StuI probe only gave a band corresponding to the fully-protected probe. These results imply that the 5′ ends of the nascent DNA molecules lie 310 ± 10 nt away from the SstI site (nt 840) at nt 1150 ± 10. This is consistent with both the

gel electrophoresis data described in Chapter 3, which maps the leading-strand origin to the non-coding region of the genome, and the mapping of the 5′ ends of D-strand molecules from S. purpuratus egg mtDNA to nt 1159 ± 5 (Jacobs et al, 1989).

The RNase protection data map the leading-strand replication origin to approximately 40 nt downstream of a TTATATATAA-like (putative promoter) sequence, and just 3′ of a run of G residues (the G-string). This gives a similar organisation to the non-coding region as that of vertebrate mtDNA, in which a G-rich CSB II sequence occurs near the site of the transition from RNA to DNA synthesis.

These mapping results are not, however, completely unambiguous, since the distance between the HpaI and the StuI restriction enzyme sites is 374 nt (Fig. 4.8b). But a 310 nt DNA molecule located between these enzyme sites would not be consistent with the size and position of the D-strand molecules mapped in sea urchin egg mtDNA (Jacobs et al, 1989), or the gel electrophoretic data described in Chapter 3 which mapped the leading-strand origin to within the non-coding region.

The probe for lagging-strand DNA molecules, synthesised from BamHI-digested pZH64 DNA, gave a band of 310 ± 10 nt when hybridised to AEEF mtDNA, suggesting that a termination or pause site for lagging-strand synthesis occurs at the leading-strand replication origin. This result is consistent with the 1D gel electrophoretic data (section 3.4.2), and could account for the completely double-stranded (Cairns form) replication intermediates seen by electron microscopy (Matsumoto et al, 1974), assuming that leading-strand synthesis pauses at the lagging-strand replication origin. A pause in lagging-strand DNA synthesis also occurs at the leading-strand replication origin (O_H) in vertebrate mtDNA (Bogenhagen et al, 1978), and a halt to lagging-strand synthesis at this point in the genome has implications for the control of mtDNA replication (section 6.2).

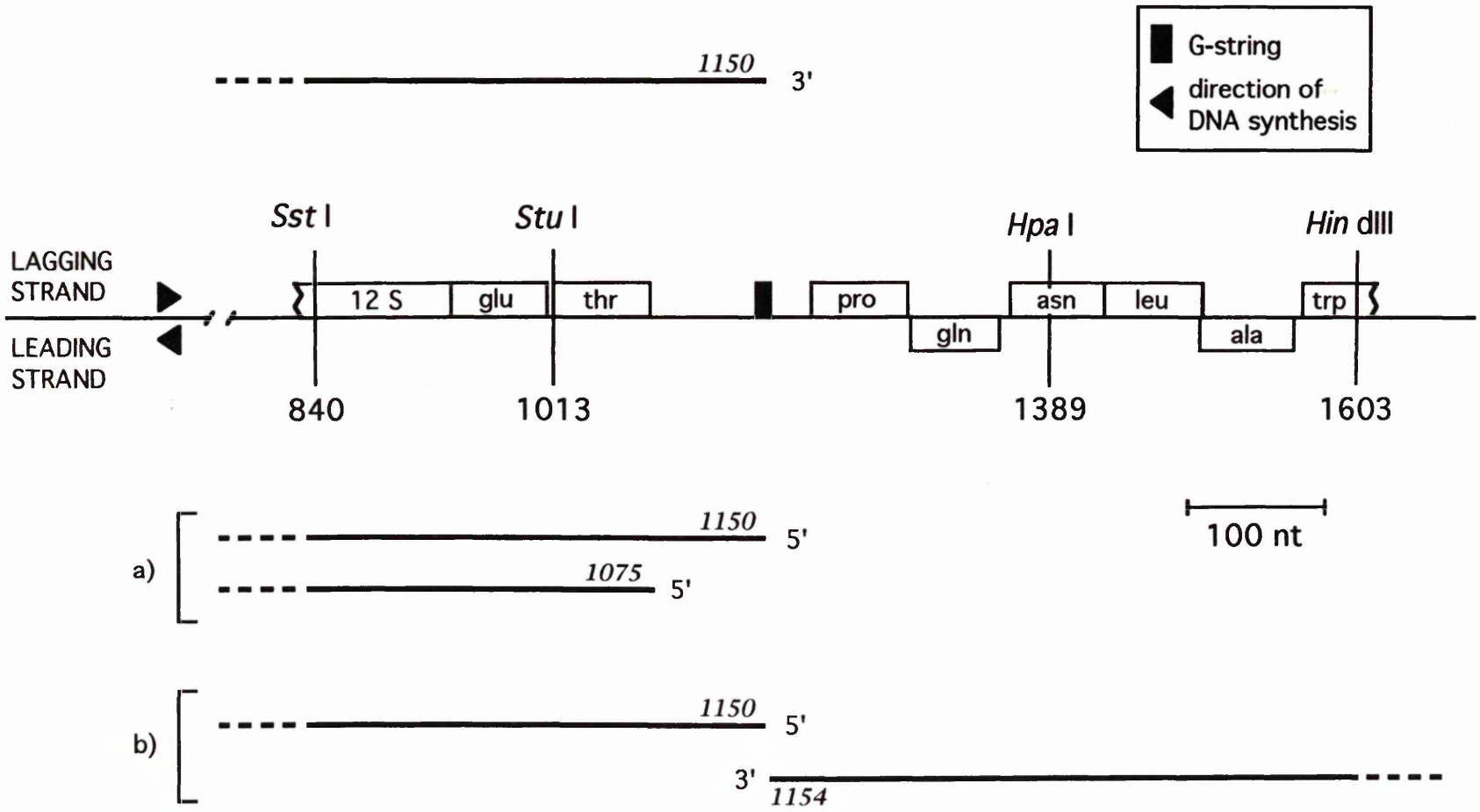


FIGURE 4.12: Diagram showing possible interpretations of the results of the RNase protection experiments performed on AEEF mtDNA, using probes for the non-coding region of the genome. The region of the genome is shown, with the genes above the line being encoded by the major-coding strand, and those below, by the minor-coding strand. The lines represent the postions of the ends of the DNA molecules, with the 2 possibilities given below for the probe complementary to the leading-strand (pZH64; a,b), and one for the probe complementary to the lagging-strand (pZH65) shown above. The numbers in *italics* give the coordinates of the ends of the DNA molecules.

The 510 nt band seen with the HpaI probe remains unexplained, since it could only have been generated by a leading-strand molecule almost completely covering the region spanned by the probe. The most likely explanation is that it is an artifact, caused by the generation of a single-stranded region of probe on the formation of the RNA/DNA hybrid. The short fragment of the probe, generated by RNase digestion at this single-stranded region, would be too small to be detected on the gel. The absence of the 510 nt band after the hybridisation reaction with the HindIII-digested plasmid is surprising, but may be due to a reduction in the stability of the structure formed by the RNA/DNA hybrid with the cut plasmid, compared to the full-length mtDNA molecule.

The 235 nt band produced in RNase protection experiments using the HpaI probe could either represent the 5′ end of a leading-strand molecule mapping to nt 1075 ± 10 (Fig. 4.12a), or the unligated 3′ end of the leading-strand at nt 1154 ± 10, the site of the leading-strand replication origin (Fig. 4.12b). A termination site for leading-strand synthesis seems the most likely explanation for this 235 nt band. In the replication of animal mtDNA, the daughter molecules separate before the gap in the newly synthesised leading-strand is ligated (Robberson *et al*, 1972), so a similar situation in *S. purpuratus* would lead to the detection of a 3′ end at the leading-strand replication origin. In kinetoplast minicircle DNA, the ligation of the gaps in the newly-synthesised daughter strands is under cell cycle control, occurring in G2 phase when the kDNA network divides (Englund, 1978), whereas kDNA replication occurs in S phase (Cosgrove & Skeen, 1970).

To resolve the question of a possible second start site for leading-strand synthesis, a primer extension reaction, using a primer located downstream from the 2 potential start sites, could be performed. This experiment would also confirm the position of the leading-strand origin, and eliminate the problem of spurious bands due to the structure of the RNA/DNA hybrid formed in the RNase protection experiments. A thermal cycling reaction could also be carried out to amplify the

signal, making the DNA molecules easier to detect. This technique can only be used to map the 5′ ends of molecules, so would not help in mapping the potential lagging-strand pause site, or the putative leading-strand replication terminator. However, further RNase protection experiments, using probes of different lengths, could be used to map the lagging-strand pause site within this region of the genome.

4.6.2 TRANSCRIPT MAPPING AT THE NON-CODING REGION.

The probes from HindIII-cut pZH65 and BamHI-cut pZH64 both gave bands of 70 ± 5 nt in RNase protection reactions with total AEEF nucleic acid (section 4.4.2). These bands were almost certainly due to mature tRNA molecules, and the leading-strand probe may also have hybridised to D-strand molecules, producing a band which would also migrate at about 70 nt.

The pZH64 RNA probe, which is complementary to the transcripts from genes encoded by the lagging-strand, gave, in addition to the 70 nt band, 2 bands of 165 ± 10 and 175 ± 10 nt on hybridisation to AEEF nucleic acid. Based on the genome organisation of S. purpuratus mtDNA, these bands could represent the 3′ ends of transcripts containing the 12S rRNA gene, and the adjacent $tRNA^{glu}$ gene, with 2 transcripts of different sizes possibly reflecting alternative termination or processing sites. In P.lividus, the 3′ ends of the transcripts covering the 12S rRNA gene have been mapped to the 3′ end of the gene for $tRNA^{glu}$, but Elliott (1990) mapped the 3′ end of the 12S rRNA transcripts in S. purpuratus to 7-13 nt inside the 5′ end of the $tRNA^{glu}$ gene. The heterogeneity in the position of the 3′ ends, stretching over 7 nt, may indicate that the 3′ end of 12S rRNA is generated by transcriptional attenuation, rather than RNA processing.

Another possibility is that the bands represent processing intermediates produced in the formation of mature tRNA molecules from a polycistronic transcript initiating from the

putative promoter sequence in the non-coding region. The polycistronic transcript extending from the promoter to beyond the HindIII site would be expected to produce a band of around 400 nt, which is not seen, though this may indicate that some of the processing events are rapid. S1 nuclease protection experiments on S. purpuratus mtRNA using a probe covering the same region of the genome as the pZH64 probe (Elliott, 1990), also gave a band of 176 nt. One possible scheme for the generation of the size of RNA molecules observed involves processing at or near the 5′ ends of tRNA genes (Fig. 4.13).

With the probe for the leading-strand, 2 bands of 265 ± 10 and 295 ± 10 nt were observed, in addition to that of 70 nt, and again it seems probable that they represent RNA intermediates in the process of tRNA formation. Again cleavage of a polycistronic RNA precursor, this time initiating from the TTATATAA-like putative promoter sequence located between the genes for $tRNA^{val}$ and $tRNA^{met}$ (nt 1786-95), could generate RNA intermediates of the observed sizes. The cleavage sites are again postulated as being at the 5′ ends of the 2 tRNA genes (gln & ala), with a termination or processing site at the 3′ end of the $tRNA^{gln}$ gene (Fig. 4.13). Further RNase protection experiments using probes of different lengths would be needed to test the validity of these ideas regarding the identity of the protected bands.

4.6.3 RNase PROTECTION MAPPING AT THE A6/COIII REGION.

RNase protection mapping using probes for the region of the mitochondrial genome containing the 3′ end of the A6 gene and the 5′ end of the COIII gene (nt 9148-9778) did not detect any discrete ends of DNA molecules. This implies that they are below the level of detection of the method used. The location of the lagging-strand origin in this region could be further investigated by using primer extension analysis. Another possibility would be to gel-purify the full-length RNA probe molecules before the hybridisation reaction, which

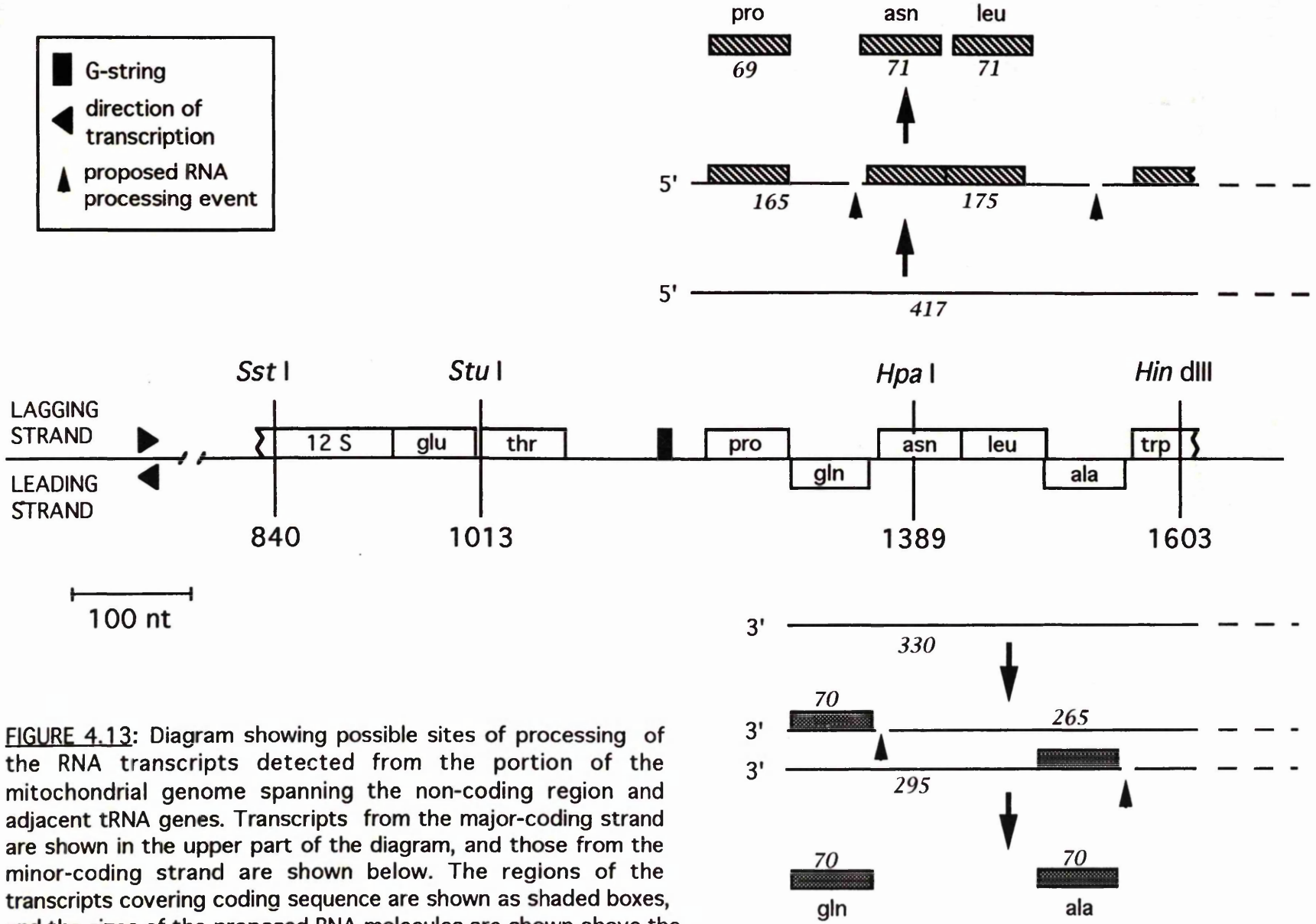


FIGURE 4.13: Diagram showing possible sites of processing of the RNA transcripts detected from the portion of the mitochondrial genome spanning the non-coding region and adjacent tRNA genes. Transcripts from the major-coding strand are shown in the upper part of the diagram, and those from the minor-coding strand are shown below. The regions of the transcripts covering coding sequence are shown as shaded boxes, and the sizes of the proposed RNA molecules are shown above the RNA molecules.

should eliminate, or at least reduce, any background bands caused by the hybridisation of any shorter RNA molecules present.

The mapping of the ends of the 2 transcripts from the genes for A6 and COIII showed that the 2 ends were close together, but did not clearly indicate whether or not they overlapped. The ends could either be generated by the processing of a single, polycistronic transcript containing sequences from the 2 genes (though mutually exclusive pathways for the generation of the 2 mRNAs would have to exist if the transcripts overlapped), or the termination of transcription upstream or within the COIII gene to create the A6 mRNA in a proportion of transcripts. COIII mRNA would then be generated by the processing of long transcripts in which readthrough of the termination site had occurred. The termination of transcription within COIII, rather than at the 3′ end of A6 would require the transcript to be processed to generate the mature A6 mRNA mapped by the RNase protection experiments (section 4.5.1).

The detection of one major and one minor 3′ end for the A6 transcript suggests that RNA processing, rather than transcriptional attenuation (which generates multiple 3′ ends), is responsible for the generation of the 3′ end of the A6 mRNA. In the transcription studies on S. purpuratus mtRNA, the multiple ends characteristic of transcriptional termination were not prominent (Elliott & Jacobs, 1989), suggesting that processing of the polycistronic messages was the major mechanism generating mature mRNAs. In P. lividus, however, some of the transcripts did show multiple 3′ ends (Cantatore et al, 1990), suggesting termination as a mechanism for generating the 3′ ends of certain mRNAs.

4.6.4 DNA BINDING PROTEINS ACTING AT THE A6/COIII GENE BOUNDARY.

The mapping of a lagging-strand replication origin, and a

replication pause site to the region of the genome around the A6/COIII gene boundary make this region of DNA a probable site for the action of DNA binding proteins, which could be involved in transcription and/or replication. To look for any such proteins, S.A.Qureshi (in this laboratory) carried out gel retardation experiments using mitochondrial proteins from sea urchin blastula, eluted from a heparin-sepharose column, and the mtDNA clones pAM62 and pAM65, which I constructed in the course of this study (section 4.3.2). This analysis revealed up to 6 different protein-binding sites (S.A. Qureshi & H.T.Jacobs, unpubl. data). DNase I footprinting experiments generated a large footprint at the 5′ end of the COIII gene (nt 9332-87) with the 0.4-1.0 M KCl eluate from the heparin-sepharose column, probably indicating multiple binding sites, and 2 footprints with the other protein fraction (the 0.2-0.4 M KCl eluate). One of these 2 footprints covered a 35 bp region of DNA near the 3′ end of the A6 gene (nt 9253-87), and the other lay within the 5′ end of the COIII gene (nt 9375-86). Further gel retardation experiments using oligonucleotides showed that the 2 protein fractions bound to the same region of the 5′ end of COIII (nt 9351-80). Southwestern blots using oligonucleotides, and gel retardation experiments using proteins eluted from SDS-PAGE gel slices, identified one protein from each fraction responsible for the DNA binding. These were a 25 kD from the 0.2-0.4 M KCl eluate, and a 22 kD protein from the 0.4-1.0 M KCl eluate, which were named mtPBP I and II (for pause-site binding protein) respectively (S.A.Qureshi & H.T.Jacobs, unpubl. data).

The location of the binding sites of these proteins is consistent with a role in transcription and/or replication, though the position of a binding site for mtPBP I within A6 argues against a role for this protein as a transcription termination factor. The position of a binding site for mtPBP II within the 5′ end of COIII would, however, be the predicted location for a terminator of A6 transcription on the evidence from vertebrate mitochondrial rRNA transcription. However, as mentioned above, the RNase

protection assays did not show the multiple 3′ ends characteristic of termination in vertebrate mtDNA.

Alternatively these proteins could be involved in the formation of the correct DNA secondary structure for the initiation of lagging-strand DNA synthesis. This process could involve the stalling of the leading-strand replication fork, resulting in the appearance of replication pause sites. Dissociation of the leading-strand replication complex at these pause sites may occur, which could conceivably provide the components for the replication complex to initiate synthesis on the lagging-strand.

4.6.5 G-RICH SEQUENCES AND THE CONTROL OF mtDNA REPLICATION.

A striking feature of the non-coding region of mtDNA in the sea urchin species S. purpuratus and P. lividus is the extensive run of consecutive G residues (20 and 26, respectively). These sequences are superfically similar to the CSB II sequence of vertebrate mtDNA (12/16 G residues in mouse; Chang & Clayton, 1989), which is involved in the transition from RNA to DNA synthesis. G-rich regions of DNA are also found in the ori sequences of S. cerevisiae mtDNA (Baldacci et al, 1984; de Zamaroczy et al, 1984), with the initiation of both leading- and lagging-strand synthesis occurring within a 17 nt G-rich region of DNA (box C; Fig. 1.6). The mapping of the 5′ ends of the nascent leading-strand DNA molecules, and D-strand molecules (Jacobs et al, 1989) to the 3′ end of this sequence suggests that the string of G residues present in sea urchin mtDNA may be acting as a signal for the transition from primer to DNA synthesis, possibly through interactions with the sea urchin equivalent of RNase MRP. It must, however, be kept in mind that the non-coding region of Drosophila mtDNA is 90-95% A+T (Clary & Wolstenholme, 1985), so G-rich DNA does not appear to be involved in the control of mtDNA replication in all organisms.

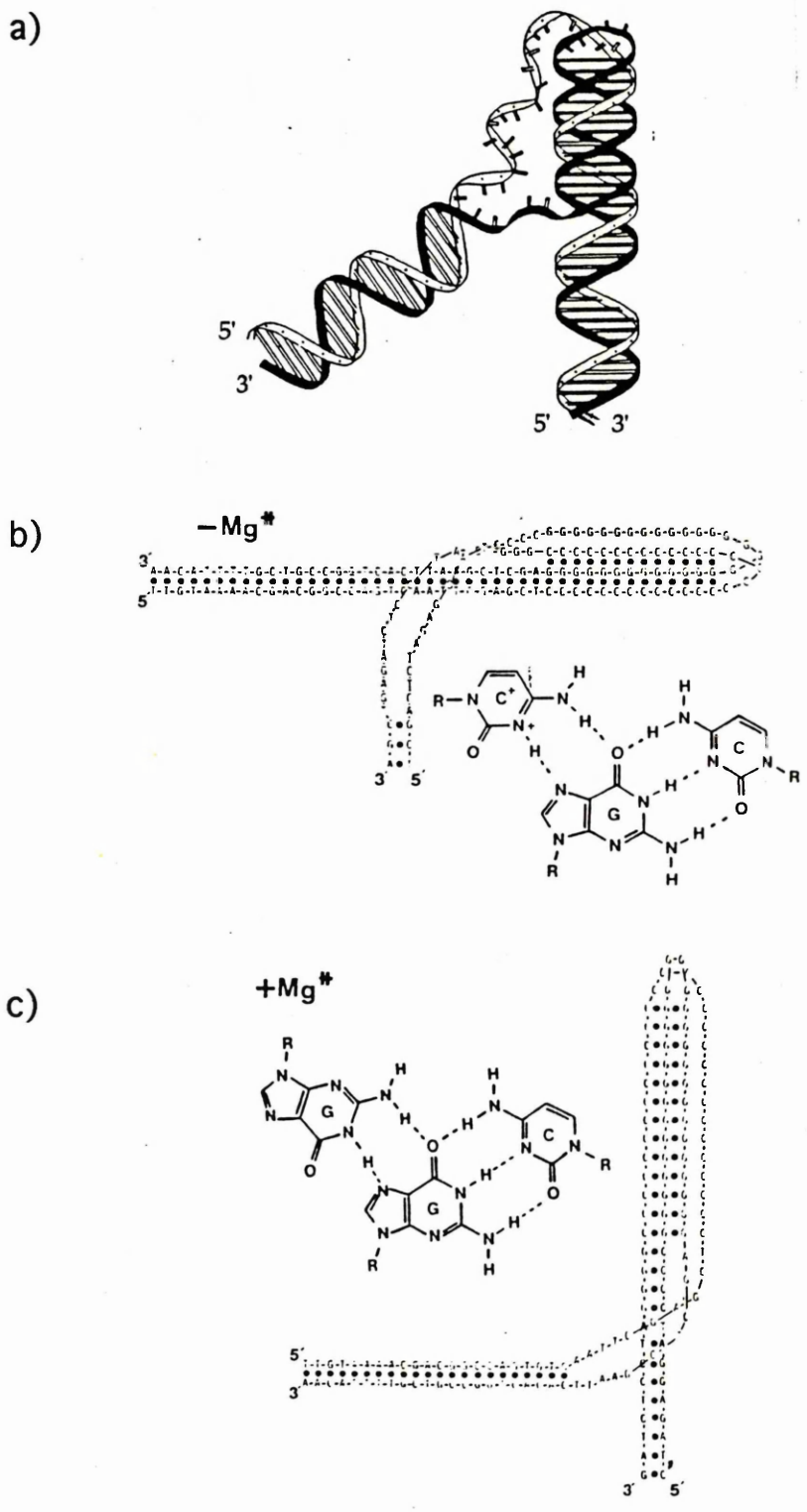


FIGURE 4.14: The proposed structure of triplex DNA regions formed by homoguanine.homocytosine stretches of DNA. (a) Pyrimidine-rich (homocytosine) sequences are shown shaded, and purine-rich (homoguanine) sequences are shown unshaded. The homocytosine sequence occupies the major groove of the duplex (as in b), associated by Hoogsteen base-pairing. From Hanvey *et al*, 1988. (b,c) Base-pairing in poly(dC).poly(dG) tract in (b) the absence of Mg ions, showing C-G-C pairing, and (c) the presence of 2 mM Mg ions, showing G-G-C pairing. From Kohwi & Kohwi-Shigematsu, 1988.

(facing page 137)

In the plasmid ColE1, a run of 6 G residues is also involved in the initiation of DNA synthesis by ensuring that the primer RNA molecule can be correctly processed. This run of G residues occurs 13-20 nt upstream of the processing site for the primer RNA molecule (RNA II), and promotes the association between the primer and the DNA template (Masukata & Tomizawa, 1990). This association allows the RNA II molecule to be processed by RNase H at the replication origin, so DNA replication can occur. The G-string in sea urchins may therefore be acting to promote primer/template association, rather than as a recognition site for processing of the RNA primer. As the mtDNA samples were treated with RNase A, not RNase H before the RNase protection assays, it is not possible to say whether the 5′ ends detected at the non-coding region represent the site of RNA to DNA transition, or the start of base-pairing between the primer RNA and the template.

Another possible role for the G-string in sea urchin mtDNA could be in the control of DNA replication, with nuclear factors (RNPs) acting at this site to block primer formation. Stretches of homopurine/homopyrimidine DNA have been shown to form unusual triple-stranded structures *in vitro* when present in supercoiled plasmid DNA (Hanvey *et al*, 1988; Kohwi & Kohwi-Shigematsu, 1988; Fig. 4.14). These structures would not be expected to form in sea urchin egg mtDNA *in vivo*, since the DNA appears to be predominantly in a relaxed form, but *in vitro* studies have shown that triple-stranded structures can be formed by complementary oligonucleotides binding to homopurine/homopyrimidine sequences. A G-rich oligonucleotide binding upstream of the human c-*myc* gene has been shown to inhibit transcription of this gene *in vitro* through the formation of triple-stranded DNA (Cooney *et al*, 1988). This suggests a possible mechanism for the regulation of transcription through the use of oligonucleotides, probably in the form of RNPs, interacting with sequences downstream of the promoter. The location of the G-string in sea urchin mtDNA also raises the possibility that a similar mechanism could be involved in the control of mtDNA

replication. An oligonucleotide or RNP binding to the G-string could block DNA synthesis by preventing the synthesis of an active RNA primer. Such a mechanism could either act in the normal cell cycle control of mtDNA replication, and decrease the rate of both D-loop synthesis and productive DNA replication, or (more attractively) only act in early embryogenesis to block replication of the mitochondrial genome.

CHAPTER

FIVE

IN VITRO TRANSCRIPTION OF THE A6/COIII REGION

5.1 INTRODUCTION.

A series of experiments using an *in vitro* transcription system was carried out in order to address the question of whether there is a link between transcriptional control and replication in *S. purpuratus* mtDNA. In *S. purpuratus* mtDNA, some of the tRNAs appear to act as the processing sites for the generation of mature mRNAs from polycistronic transcripts (Elliott & Jacobs, 1989). However, some of the protein-coding mitochondrial genes of *S. purpuratus* lack tRNA genes at their 3′ ends, so a different mechanism is needed for the generation of mature transcripts from these genes. The convergent transcription of the ND5 and ND6 genes suggests that transcriptional termination may be occurring to prevent the generation of antisense RNAs for these 2 genes. Termination could also be responsible for creating the 3′ ends of other mRNAs, as has been proposed by Cantatore *et al* (1990) for several of the mitochondrial genes (*e.g.* 16S rRNA) of the sea urchin *P. lividus*. Some of the pause sites for DNA replication observed in *S. purpuratus* mtDNA occur in the region of the 3′ ends of genes which lack a downstream tRNA gene, so it is possible that the factors involved in the production of mature mitochondrial transcripts might also affect DNA replication.

The pausing of replication fork movement at a transcription termination site has been observed in the nuclear rRNA gene cluster of yeast (*S. cerevisiae*), and it has been postulated that this pausing is due to direct confrontation between the RNA and DNA polymerases moving in opposite directions along this highly transcribed region of DNA (Brewer & Fangman, 1988). In the *E. coli* chromosome, most of the highly transcribed genes are arranged so that the direction of transcription and replication fork movement is the same, thus minimising interference between the 2 polymerases (Brewer, 1988). However, the mitochondrial genome of *S. purpuratus* is organised such that DNA polymerase movement for leading-strand replication (section 3.7.2) is in the opposite direction to that of the RNA polymerase transcribing all the

protein coding and rRNA genes except for ND6 (Jacobs et al, 1988), so that the potential for polymerase interference does exist.

The possible involvement of transcription terminators in the production of mature mRNAs from the A6/COIII region of the mitochondrial genome was investigated by in vitro transcription experiments using this region of the genome. This stretch of mtDNA was chosen because it contains the most prominent replication pause site, observed by 2D gel electrophoresis (section 3.6). Mitochondrial RNA polymerase had not been successfully isolated from S. purpuratus at the time these experiments were performed: the bacteriophage polymerases from SP6 and T7 were therefore used, because they are single subunit enzymes, like the mitochondrial RNA polymerases so far characterised, and the phage enzymes show a considerable degree of similarity with the yeast mitochondrial RNA polymerase at the amino acid level (Masters et al, 1987). This similarity is mainly confined to the carboxy-terminal end of the polymerases, where the similarity between the distantly related T7/T3 and SP6 polymerases is mostly located (Masters et al, 1987). The carboxy-terminal end of the enzyme forms the catalytic portion of the molecule, and is also thought to be involved in promoter recognition, with the amino-terminal thought to have a role in stabilising the promoter-polymerase complex.

Transcription was analysed by run-off assays carried out in the absence and presence of mitochondrial protein extracts, to investigate the possible involvement of DNA binding proteins in transcription termination. Transcriptional analysis in the absence of added proteins indicates the occurrence of DNA sequences that may act as terminators in vivo. The products of the transcription reactions were analysed by polyacrylamide gel electrophoresis for the presence of prematurely-terminated transcripts. Both T7 and SP6 RNA polymerase were used, to test whether any of the shorter products were polymerase-specific. A lack of specificity would seem to indicate a general termination site

for this type of RNA polymerase within the region studied.

5.2 SUBCLONING OF THE A6/COIII REGION OF mtDNA INTO SP6 AND T7 TRANSCRIPTION VECTORS.

In addition to the SP6 constructs described in section 4.3.2 (pSPV3/31, and the pAM35 and pAM6-35 exoIII deletion families), several other plasmids were also used in the *in vitro* transcription studies. Two plasmids, pSP3 and pSP31, were produced by subcloning part of the mtDNA insert from the plasmid pR3 into the vector pSP64. These constructs contained additional mtDNA sequences compared to pSPV3 and pSPV31, so that the insert contained the complete A6 and A8 genes, and part of the genes for $tRNA^{lys}$ and COIII (Fig. 5.1c,d).

The region of mtDNA from the HindIII site within the $tRNA^{lys}$ gene (nt 8424) to the EcoRI site within COIII (nt 9778) was cut out of the plasmid pR3 using HindIII, which meant that the pUC8 polylinker was attached to the mtDNA insert. This DNA was ligated to HindIII-digested pSP64, and transformed into *E. coli* DS941. Plasmid DNA isolated from a number of transformants was screened by EcoRI digestion and agarose gel electrophoresis for the presence of the insert DNA. This analysis revealed one clone containing the insert, which was orientated such that transcription from the SP6 promoter through the insert would occur in the same direction as transcription of the A8, A6, and COIII genes *in vivo*. Further digests using SstI, EcoRI and HindIII confirmed the identity of the clone (Fig. 5.1a), which was designated pSP3. To obtain the insert in the opposite orientation, DNA from this clone was digested with HindIII, and then religated to itself, and transformed into *E. coli* DS941. Plasmid DNA from some of the transformants was again isolated and analysed by EcoRI digestion and agarose gel electrophoresis. A clone containing the mtDNA insert in the opposite orientation to pSP3 was identified. Further enzyme digests using HindIII, EcoRI, and SstI confirmed this (Fig. 5.1b), and the plasmid was designated pSP31. Maps of the 2 plasmids are shown in

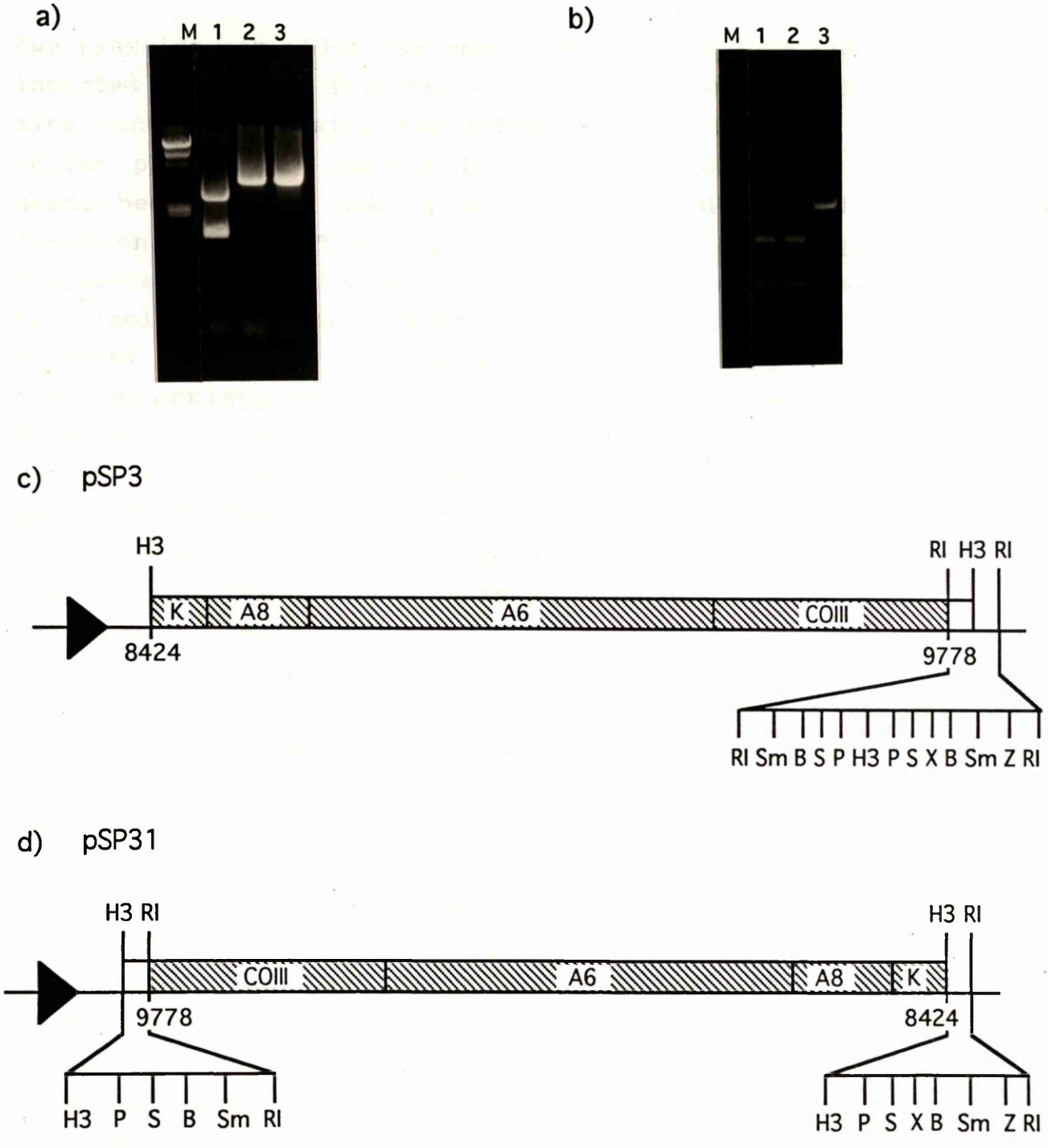


FIGURE 5.1: (a,b) EtBr-stained agarose gels showing digests of plasmid DNA to confirm the identity of the subclones pSP3 and pSP31. (a) pSP3 and (b) pSP31 DNA digested with: lane 1 - *Hin* dIII; 2 - *Eco* RI; 3 - *Sst* I. *Hin* dIII-digested λ DNA was used as a size marker (M). (c,d) Maps of the 2 plasmids, showing the extent of the mtDNA insert (shaded box) and the genes it contains (K - $tRNA^{lys}$), the SP6 promoter (filled arrow head), and the restriction enzyme sites (see legends to Figures 4.3 & 4.4 for abbreviations). The open box represents the pUC8 polylinker sequence derived from pR3. Diagrams not to scale.

Figure 5.1c,d.

Two plasmids, in which the sequences derived from mtDNA were inserted downstream from the T7 RNA polymerase promoter, were also constructed using the mtDNA insert from pR3, and the vector pTZ19R. The method followed was identical to that described for the making of plasmids pSPV3, and pSPV31 (section 4.3.2). The PvuII/HindIII and PvuII/EcoRI DNA fragments from pR3, containing the desired mtDNA sequences, were isolated from an agarose gel using Geneclean. pTZ19R was digested with SmaI, and either HindIII, or EcoRI, and then the appropriate mtDNA and vector sequences were ligated together (Fig. 5.2b,c). The DNA was transformed into E. coli DS941, and plasmid DNA was isolated from some of the resulting transformants and digested with HindIII. Analysis of the digested DNA by agarose gel electrophoresis revealed that most of the clones contained the insert. Those that did not were probably the result of religation by the small proportion of pTZ19R DNA in which SmaI-digestion had not occurred. One clone containing insert DNA was chosen from each of the screened transformants, and named pTZV3 and pTZV31 for the PvuII/EcoRI and the PvuII/HindIII inserts, respectively.

Further restriction enzyme digestions of the 2 different clones, using EcoRI, PstI, SstI, and SalI for pTZV31, and EcoRI, EcoRI/ HindIII, PstI, and SstI for pTZV3, gave the predicted band sizes, confirming the maps of the clones (Fig. 5.2a). Plasmid DNA was then isolated from each clone by CsCl gradient centrifugation. The restriction maps of the 2 plasmids are shown in Figure 5.2b,c.

Another exonuclease III (exoIII) deletion series covering the A6/COIII junction region was constructed from the plasmid pSPV3, with the deletions of the mtDNA sequences starting from the PvuII site at nt 9148. The plasmid was digested sequentially with PstI, and BamHI to generate the necessary 3′ and 5′ overhangs for the exoIII reaction. The deletion reaction was carried out at 30^{o}C, with 8 time points taken at

a)

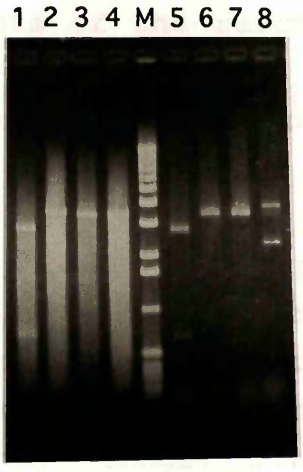
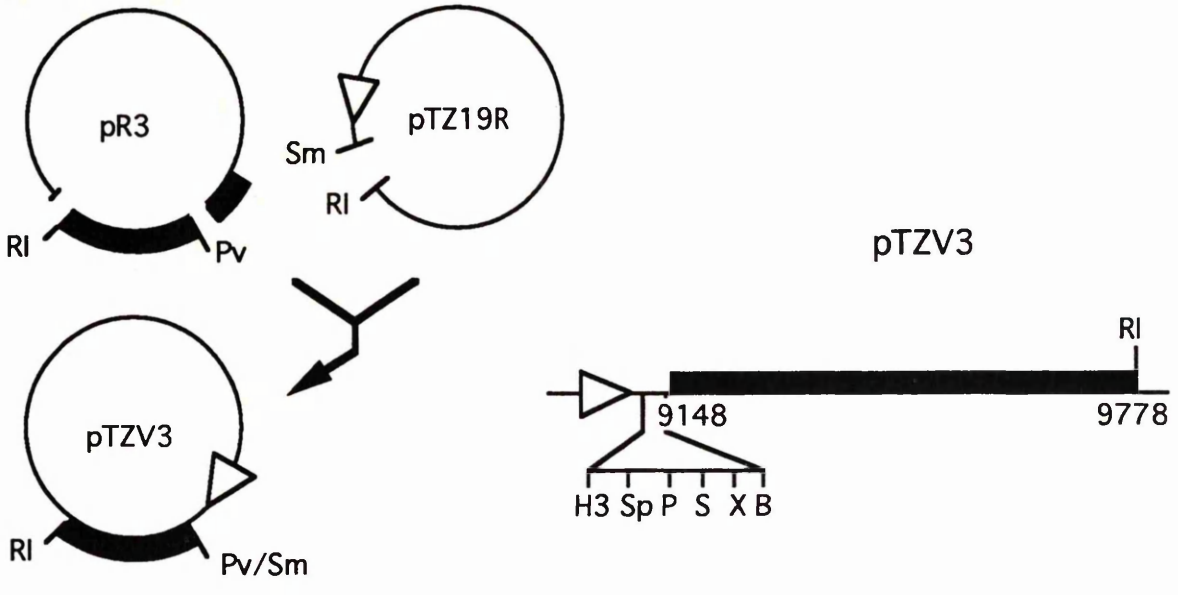
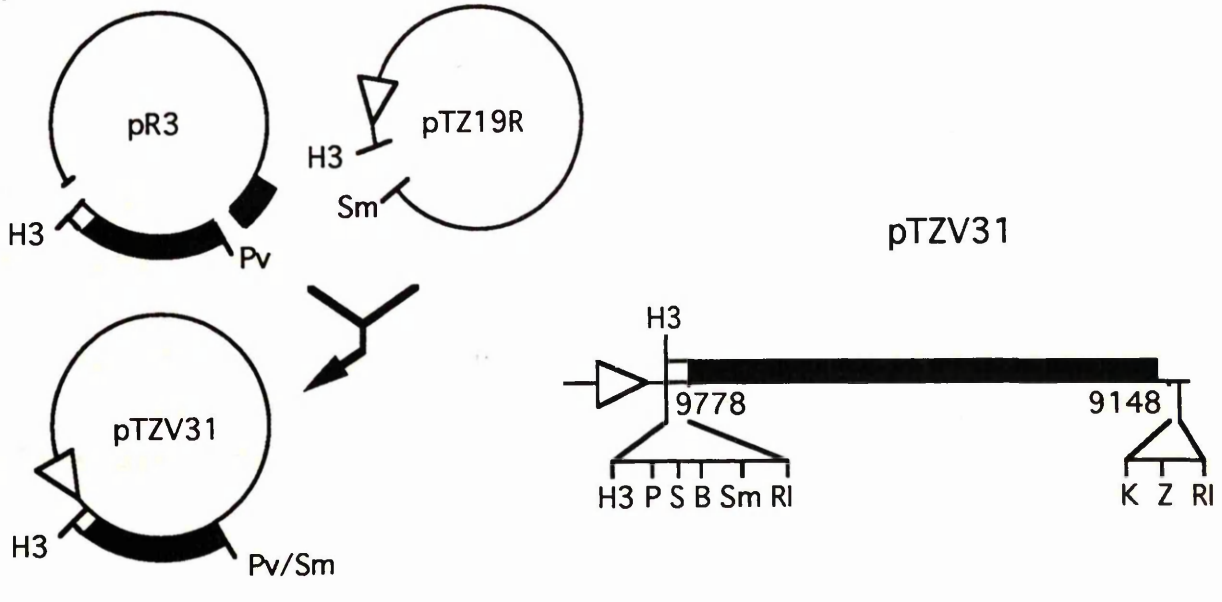


FIGURE 5.2: (a) EtBr-stained agarose gel showing restriction enzyme digestions of pTZV3 and pTZV31 plasmid DNA carried out to confirm the identity of the subclones. Lanes 1-4, pTZV31 DNA, digested with: 1 - *Eco* RI; 2 - *Pst* I; 3 - *Sst* I; 4 - *Sal* I. Lanes 5-8, pTZV3 DNA, digested with: 5 - *Eco* RI/*Hin* dIII; 6 - *Eco* RI; 7 - *Pst* I; 8 - *Sst* I. BRL 1 kb ladder was used as a size marker (M). (b,c) Schematic diagrams showing the cloning strategies used in the construction of the plasmids (b) pTZV3 and (c) pTZV31, and maps of the respective plasmids. The mtDNA insert is shown as a filled box; the T7 promoter as an open arrow head; and the pUC8 polylinker sequence derived from pR3 as an open box.

b)



c)



60 second intervals. After treatment with mung bean nuclease, aliquots of the reaction mixtures from time points 2, 4, 6, and 8 were subjected to agarose gel electrophoresis to check the extent of the exoIII activity.

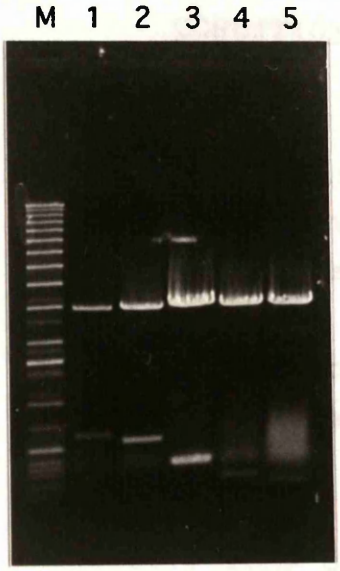
Self-ligation of the deleted DNA, followed by transformation of *E. coli* DS941 produced few colonies. Analysis of the plasmid DNA from 12 clones (4 from each of the time points 1, 2, and 3) by restriction enzyme digestion revealed that in only 4 of them had the DNA undergone deletion, and only 2 different insert sizes were produced. The lack of viable transformants probably indicates that exoIII deletion was occurring into the vector, as well as into the mtDNA insert, so that most of the religated plasmids were no longer viable.

A further deletion series was therefore prepared and analysed as above, and 3 more clones were isolated. Together with the 2 from the previous exoIII deletion reaction, the 5 clones had mtDNA inserts of decreasing size (Fig. 5.3a). Maps of the plasmids (designated pAMV1/2/3/4/5) indicating the sizes of the inserts, and their coordinates, as judged by run-off transcription assays, are shown in Figure 5.3b.

5.3 IN VITRO TRANSCRIPTION REACTIONS.

The various plasmids described in sections 4.3.2, and 5.2 were used to examine the effects of transcription *in vitro* through the cloned mtDNA region to see whether the RNA polymerase would terminate within this area, or simply produce full-length run-off products. Pairs of plasmids, in which the inserts were transcribed in opposite directions, were examined to see whether any of the effects were orientation-dependent. The use of plasmids containing the same mtDNA insert with either the T7 or SP6 promoter, also made it possible to investigate whether any effects were polymerase-dependent.

a)



b)

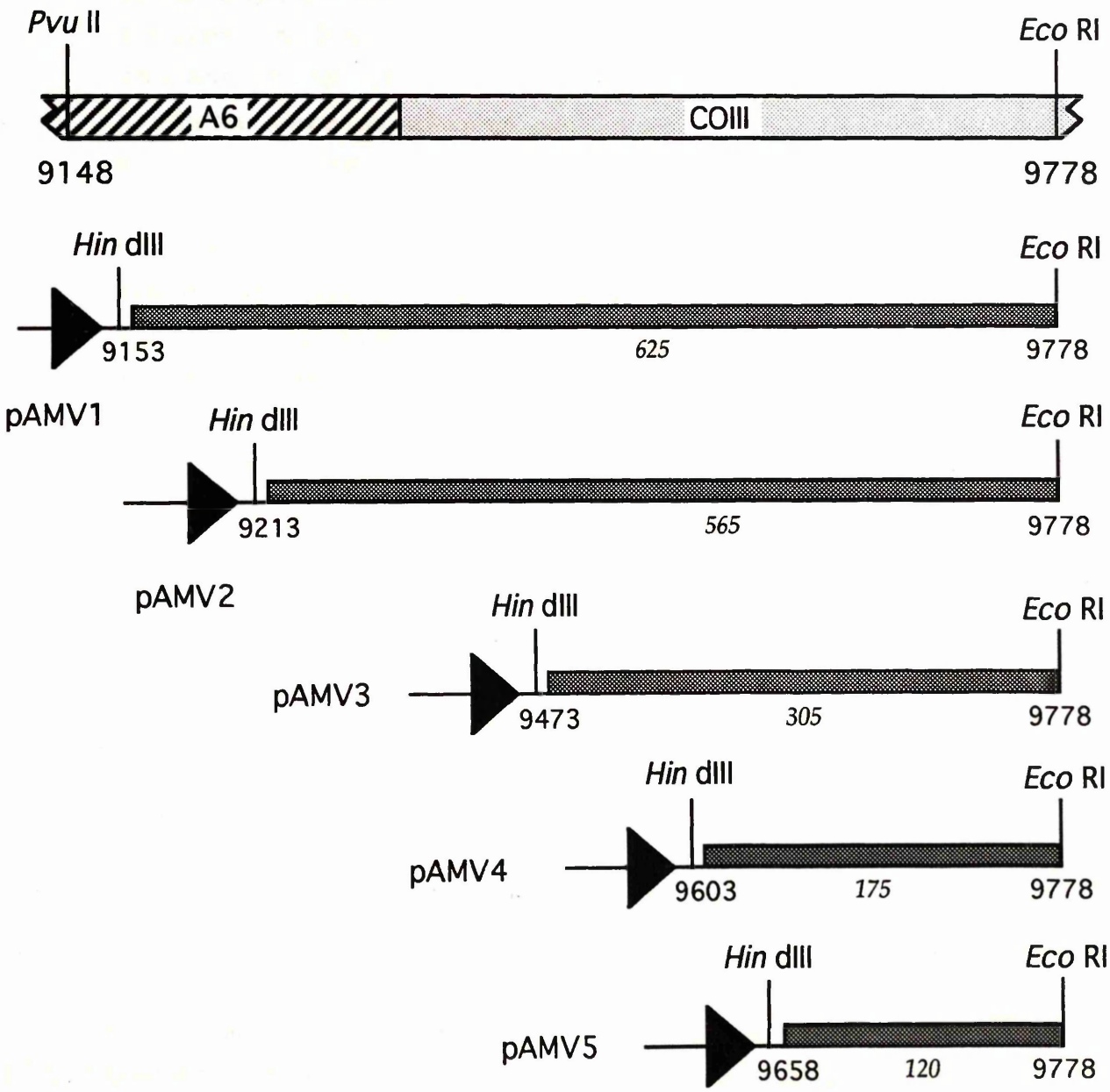


FIGURE 5.3: (a) EtBr-stained agarose gel showing *Eco* RI/*Hin* dIII-digested plasmid DNA from the 5 members of the pAMV exoIII deletion series. Lanes 1-5: pAMV1-pAMV5. BRL 1 kb ladder was used as a size marker (M). (b) Maps of the pAMV plasmids, showing the sizes (*italics*) of the mtDNA inserts (shaded boxes), their coordinates in the mitochondrial genome, the position of the SP6 promoter (filled arrow head), and the flanking restriction enzyme sites. Above the maps of the clones is a diagram of the mitochondrial genome, showing the position of the boundary between the A6 and COIII genes.

5.3.1 OPTIMISATION OF IN VITRO TRANSCRIPTION REACTION CONDITIONS.

Initially the *in vitro* transcription conditions were optimised to minimise the generation of non-full-length RNA products. This was to ensure that if any of these RNAs were observed it was due to the sequences present within the insert, and not merely to a limiting amount of nucleotides, and/or the effects of temperature leading to the dissociation of the RNA polymerase from the DNA.

The effects of different temperatures, and the increasing concentrations of unlabelled GTP (the limiting nucleotide) on transcription were studied. Reactions were carried out for 30 minutes at both 30°C and 37°C, in the presence of a final concentration of either 30, 60, or 120 uM cold GTP. The products were separated on a 3.5% polyacrylamide/8M urea gel, and visualised by autoradiography. The results, using *Sst*I-digested pSPV31 DNA as the template, showed that increasing the amount of unlabelled GTP in the reaction decreased the number of non-full-length bands observed (Fig. 5.4), as did a decrease in the reaction temperature from 37°C to 30°C, although the effect of this was less marked. Based on these results, I decided to carry out all further transcription reactions at 30°C, with unlabelled GTP at a final concentration of 60 uM.

5.3.2 PRODUCTION OF NON-FULL-LENGTH TRANSCRIPTS.

Transcription through the mtDNA region between the *Pvu*II and *Eco*RI sites at nt 9148 and 9778 respectively produced the same major products with T7 and SP6 RNA polymerases. *In vitro* transcription reactions were carried out on *Eco*RI-digested pSPV3 and pTZV3 DNA, and on *Sst*I-digested pSPV31 and pTZV31 DNA, and the products were analysed by polyacrylamide/urea gel electrophoresis. All 4 reactions produced a band of approximately 670 nt, which represents the full-length RNA product (Fig. 5.5), as well as one or more shorter products,

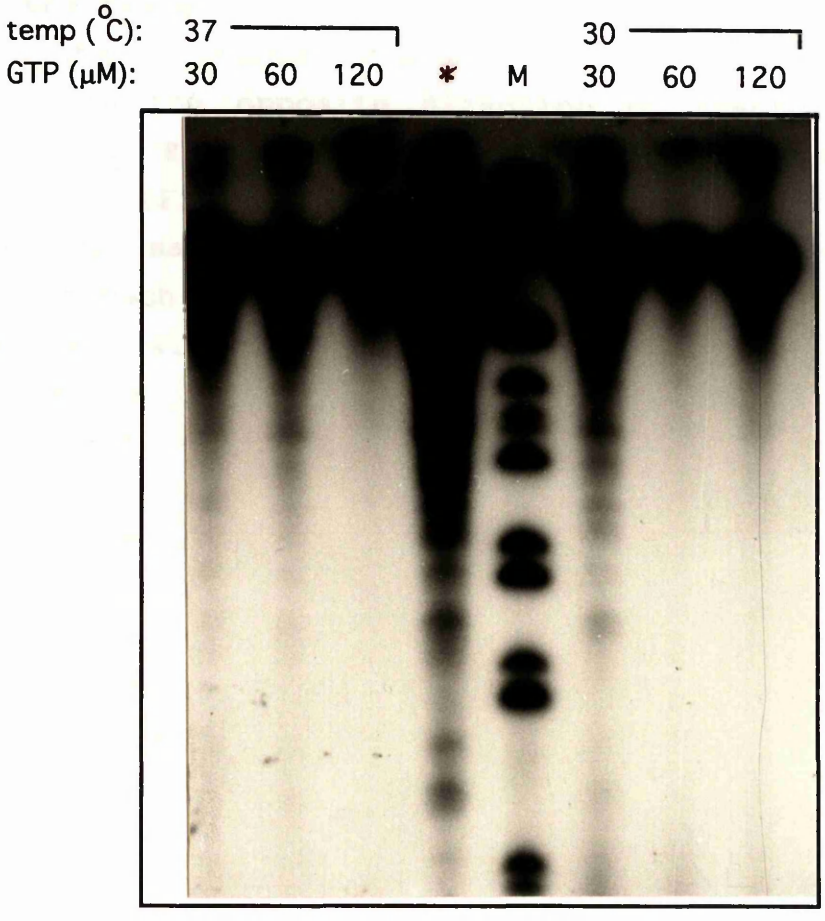


FIGURE 5.4: Autoradiograph of a polyacrylamide/urea gel on which the products of *in vitro* transcription reactions were analysed for the effects of altered temperature and GTP concentrations. In each reaction *Sst* I-digested pSPV31 DNA was used as the template. The different temperatures (30 or 37°C) and GTP concentrations (30, 60, or 120 μM) are indicated above each lane. BRL 1 kb ladder was used as a size marker (M). Lane marked with an asterisk (*) is not relevant to this figure.

which could be the result of premature termination, or of RNA processing events.

Several higher molecular weight bands of more than 1 kb were also present in all 4 lanes, though in differing amounts compared to the run-off product. These were probably the result of the transcription of small amounts of uncut plasmid present in the reactions. Transcription of both pTZV31 and pSPV31 DNA (in the opposite direction to transcription *in vivo*: *i.e.* nt 9778 to 9148) generated an extra, smaller band of 430 ± 10 nt (Fig. 5.5; lanes 5,6). This 430 nt band was present at the same intensity relative to the full-length transcript in each reaction. A shorter product band of 510 ± 10 nt was seen with pTZV3 and pSPV3 (transcription from nt 9148 to 9778, as *in vivo*), and pSPV3 generated 2 additional bands of 380 ± 10 and 360 ± 10 nt.

The reactions were repeated using the same 4 templates, and in addition pSPV3 and pSPV31 DNA cut with PvuII, which has a recognition site within the vector, 230 nt downstream from the cloning site. Again the bands observed with the T7 and SP6 polymerases were the same in each case, and the 510 nt band produced by the transcription of EcoRI-digested pSPV3 was also observed with the PvuII-cut plasmid (Fig. 5.5; lane 3). However, the same was not true when the 2 different digests of the pSPV31 samples were compared. In this case only the run-off transcript was produced by the PvuII-digested DNA, with the 430 nt band seen using SstI-cut pSPV31 definitely absent (Fig. 5.5; compare lanes 6 & 7). There are several possible explanations for the difference between the 2 digests. Firstly, the extra 230 nt of DNA transcribed in the reaction containing PvuII-digested DNA could allow the transcribed region to form a different secondary structure from the DNA in SstI-cut pSPV31, favouring termination in one conformation (SstI), but not the other (PvuII). Secondly, the minor band may be produced by processing of the RNA transcript, which could also be affected by the RNA adopting different secondary structures due to the differences in length. The latter explanation does seem less likely, since

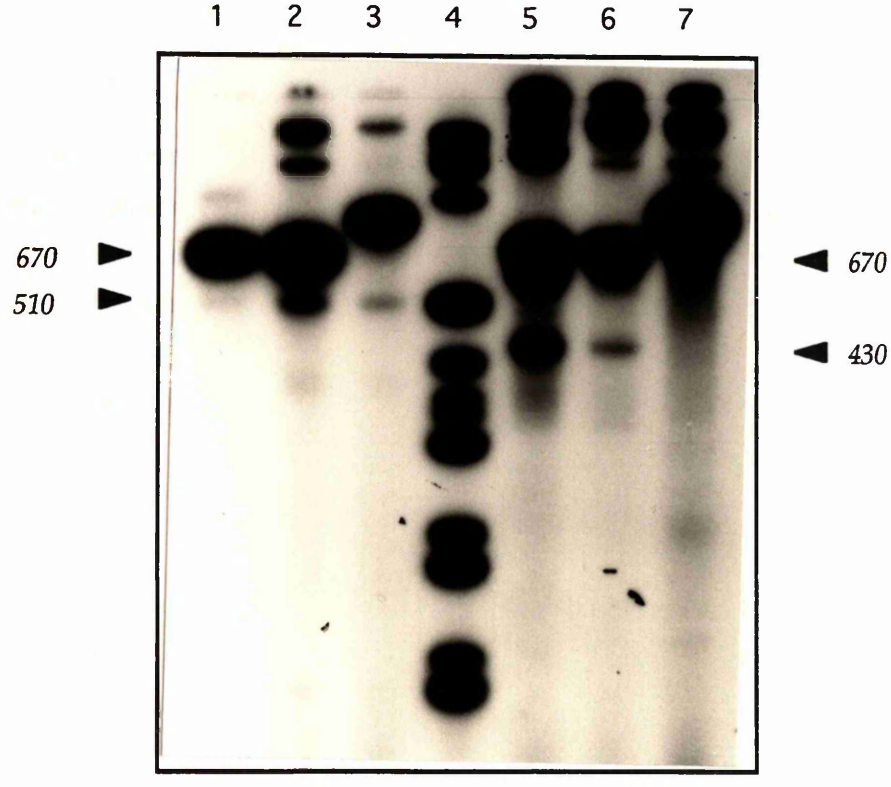


FIGURE 5.5: Autoradiograph showing the products of *in vitro* transcription reactions analysed by polyacrylamide/urea gel electrophoresis. Lane 1 - *Eco* RI-digested pTZV3; 2 - *Eco* RI-digested pSPV3; 3 - *Pvu* II-digested pSPV3; 4 - BRL 1 kb ladder; 5 - *Sst* I-digested pTZV31; 6 - *Sst* I-digested pSPV31; 7 - *Pvu* II-digested pSPV31. The arrows indicate the positions and sizes (in nt) of the bands discussed in the text.

only one smaller band is generated, whereas cleavage of the RNA at a single site would be expected to produce 2 product molecules. However, it is possible that the processing could either generate several small RNAs, which were not detected on the gel, or an unstable RNA product, which was degraded, or even a lariat structure, as is produced in the group I intron self-splicing reactions (Cech, 1990), which would migrate abnormally through the gel.

A third possibility is that the minor band was generated by spurious transcription of the DNA template, occurring from the 3′ overhang generated by SstI digestion of the template. This has been reported to occur with SP6 polymerase (Promega Protocols & Applications Guide), and means that transcription would occur in the opposite direction through the mtDNA insert compared to transcription from the phage promoter. The size of the minor product band generated by transcription of pSPV31 would, however, imply that the polymerase is not terminating at the same site in the mtDNA as that recognised in the pSPV3 reactions. Transcription from the 3′ end could also produce longer transcripts if termination did not occur within the insert, or if transcription initiates from the 3′ end of the vector.

Two different approaches were taken to test whether the minor RNA species observed with SstI were merely a consequence of the particular enzyme used, rather than the structure of the template, or the resulting RNA transcript.

The first approach was to carry out *in vitro* transcription reactions using the members of the exoIII deletion series derived from pSPV31 (pAM35 *etc.*). The exoIII deletion reaction removed the EcoRI site in pSPV31, which is located within the polylinker adjacent to the SP6 promoter, so EcoRI digestion could be used to linearise the plasmids at the EcoRI site at the other end of the mtDNA insert. SstI digestion and EcoRI digestion of the plasmids generated linearised DNA templates for *in vitro* transcription in which the lengths of the transcribed region only differed by 6 nt.

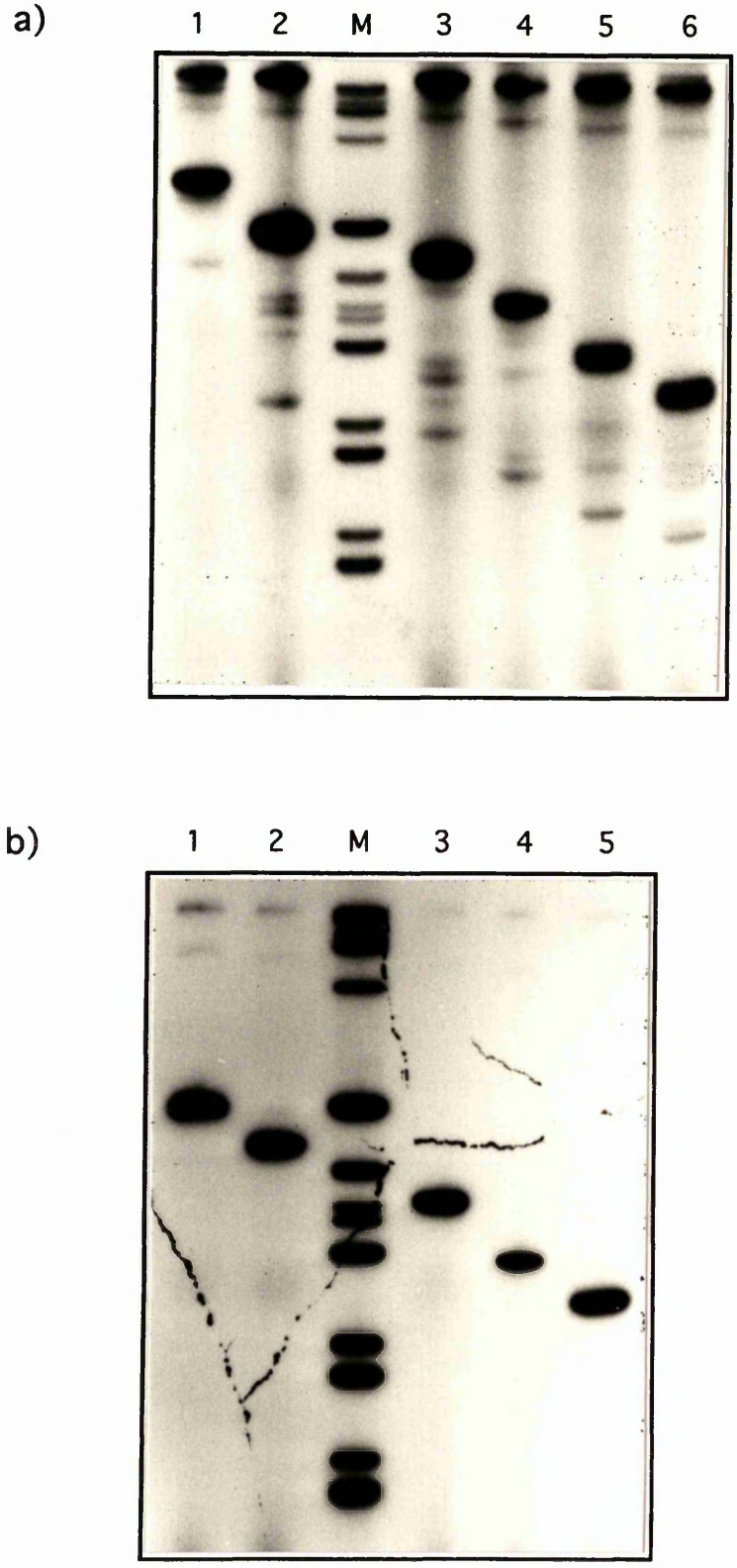


FIGURE 5.6: Autoradiograph showing the products of *in vitro* transcription reactions analysed by polyacrylamide/urea gel electrophoresis. (a) *Sst* I-digested plasmid DNA. Lane 1 - pSPV31; 2 - pAM35; 3 - pAM43; 4 - pAM55; 5 - pAM62; 6 - pAM65. (b) *Eco* RI-digested plasmid DNA. Lane 1- pAM35; 2 - pAM43; 3 - pAM55; 4 - pAM62; 5 - pAM65. BRL 1 kb ladder was used as a size marker (M) for each gel.

(facing page 152)

This therefore effectively excluded any effects that the extra DNA sequence might have had on the transcription of PvuII-digested pSPV31, compared to SstI-digested pSPV31.

The results of the transcription reactions showed that the extra, shorter bands were only produced with SstI-digested template (Fig. 5.6a), and not with plasmids cut with EcoRI (Fig. 5.6b). This strongly suggests that it is the use of SstI *per se* that results in the generation of the shorter RNA products, rather than the structure of the template DNA, or the resulting RNA transcript.

The second method used was to convert the 3′ overhang generated by SstI digestion to a blunt end by making use of the 3′ to 5′ exonuclease activity of Klenow DNA polymerase, as recommended by Promega (Protocols & Applications Guide). Four different plasmids were used (pSPV31, pAM35, pAM43, and pSP3), and comparisons were made of the products of the *in vitro* transcription reactions carried out on SstI-digested DNA, which either had or had not undergone Klenow pretreatment. Klenow pretreatment was effective in reducing the number and intensity of the minor bands produced with the plasmids pSPV31, pAM35 and pAM43 (Fig. 5.7), although it did not eliminate these bands completely. However, with pSP3 there was no effect on the 2 prominent RNA products of 480 ± 10, and 280 ± 10 nt (Fig. 5.7; lanes 7,8). These bands were also observed in reactions in which the template DNA was linearised by different restriction enzymes, so were not thought to be the result of SstI digestion, and therefore should not have been affected by pretreatment with Klenow DNA polymerase.

These 2 sets of results strongly suggest that SstI digestion can affect the products generated from the transcription reaction, and that the non-full-length transcripts observed with pSPV31 and its derivatives were not due to the structure adopted by the template DNA. It therefore appears that *in vitro* transcription of the region of the mtDNA spanning the

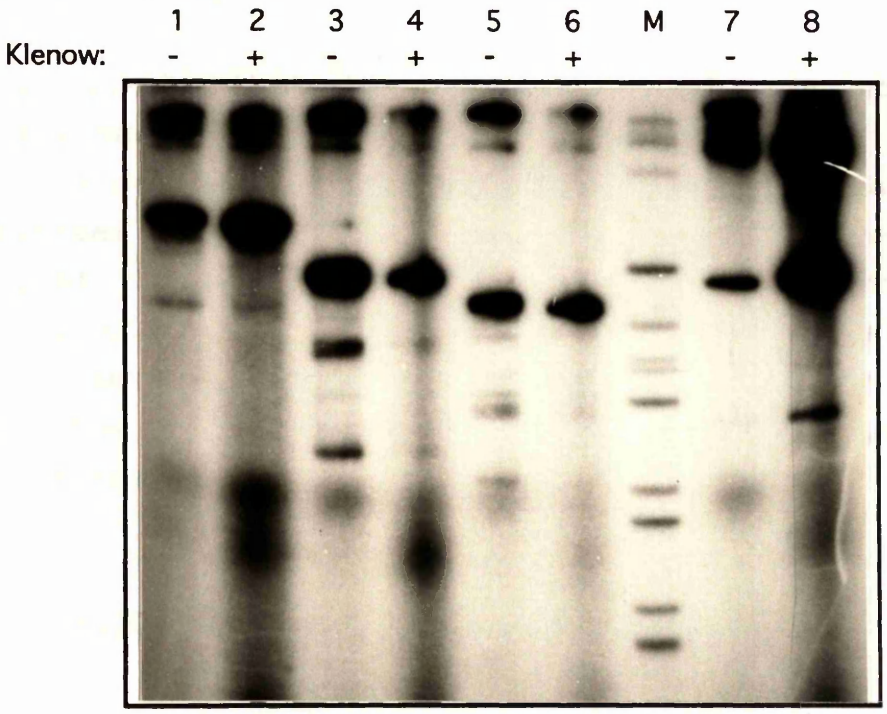


FIGURE 5.7: Autoradiograph showing the products of *in vitro* transcription reactions analysed by polyacrylamide/urea gel electrophoresis, demonstrating the effects of treatment of the template with Klenow DNA polymerase before transcription. The use (+) or absence of Klenow (-) is indicated above each lane. Lane 1,2 - *Sst* I-digested pSPV31; 3,4 - *Sst* I-digested pAM35; 5,6 - *Sst* I-digested pAM43; 7,8 - *Sst*I-digested pSP3. BRL 1 kb ladder was used as a size marker (M).

A6/COIII gene boundary only generates shorter RNA molecules when transcription is occurring in the same direction as *in vivo*. The generation of these shorter RNA products did not appear to be polymerase-dependent, since both T7 and SP6 polymerases gave the same pattern of bands. These results imply either that there is a termination sequence present in this region of the mtDNA, which can be recognised by the 2 polymerases, and therefore may also be recognised by the mitochondrial RNA polymerase *in vivo*, or that the RNA transcripts produced form a particular structure, which undergoes processing.

It is most likely that the non-full-length RNA molecules produced are the result of premature termination by the RNA polymerase, since the absence of proteins (other than phage RNA polymerase) in the transcription reactions implies self-processing of the RNA transcripts. If premature termination does occur, it is not clear from the above experiments whether the termination sites are determined soley by the *sequence* of the DNA template, or are related to the *length* of the template transcribed.

5.3.3 DOES PROCESSING OF THE TRANSCRIPTS OCCUR?

Two different approaches were taken in order to determine whether processing of the *in vitro* transcripts was occurring. The absence of any proteins in the reaction mixture, except for the RNA polymerase, does argue against it, although the self-splicing of some class I and II introns can occur *in vitro* (Cech, 1990; Jacquier, 1990). Both of the approaches taken tackle the question in a rather indirect way, in that they examine whether the ratio of the full-length RNA to the smaller product varies during the time course of the reaction. A direct way of addressing the question would have been to isolate the full-length RNA transcript from a polyacrylamide gel, and then incubate that RNA under various conditions to determine whether any shorter RNA products were generated. Although this approach seems attractive, there

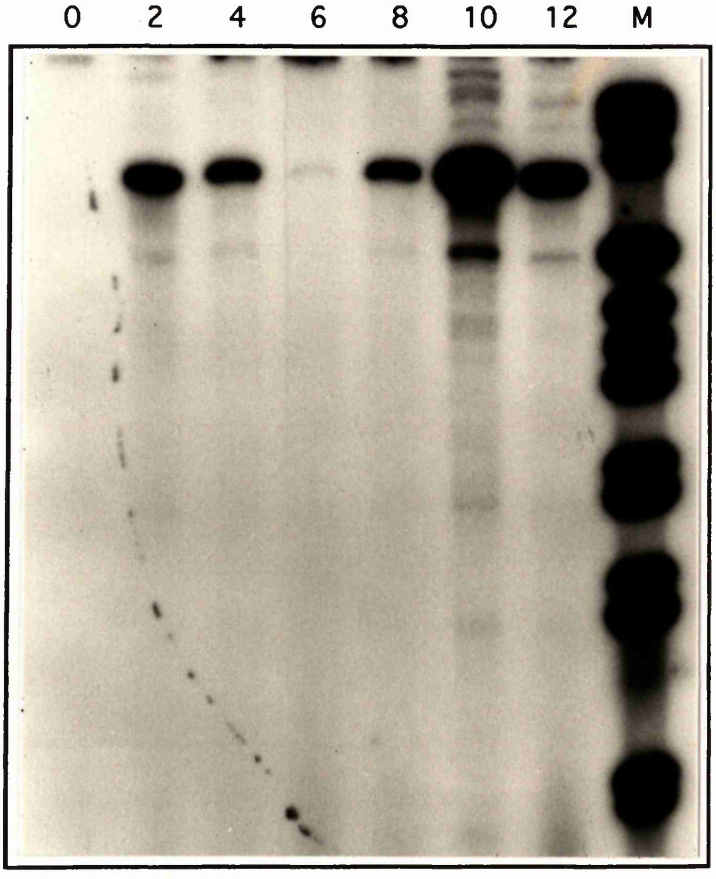


FIGURE 5.8: Autoradiograph showing the products of *in vitro* transcription reactions analysed by polyacrylamide/urea gel electrophoresis. Aliquots of the reaction mixture were taken at 2 minute intervals for 12 minutes (indicated above each lane), as well as at time zero (lane 0). *Pvu* II-digested pSPV3 DNA was used as the template, and BRL 1 kb ladder was used as a size marker (M).

(facing page 155)

would be the possibility that refolding of the RNA after its denaturation during electrophoresis could alter its conformation, leading to the stimulation, or inhibition of any potential processing reaction.

The first approach was to do a time course reaction for *in vitro* transcription and determine whether the ratio of full-length product to the shorter RNA products altered over time. Aliquots were removed from the reaction at several time intervals, and the reaction in each stopped by phenol extraction, followed by EtOH precipitation. The products were then separated by polyacrylamide/urea gel electrophoresis, and visualised by autoradiography.

*Pvu*II-digested pSPV3 was the DNA tempate used in these reactions, and aliquots were taken every 2 minutes for 12 minutes. Unfortunately, the subsequent manipulations of the samples meant that the amounts loaded onto the gel for each time point varied considerably, with no apparent increase in the amount of product with time (Fig 5.8). However, the ratio of the run-off product to the smaller (510 nt) RNA molecule appeared to remain constant at each time point (compare lanes 2 and 7, Fig. 5.8). Unless the processing reaction was very rapid, and an equilibrium level was reached at which no net processing occurred, an increase in the the amount of shorter RNA compared to full-length product would be expected to take place with time, which was not the case.

The second approach was to carry out an *in vitro* transcription reaction, and then aliquot out equal amounts of the RNA product and investigate whether incubation under different conditions would promote an increase in the ratio of minor RNA product to full-length transcript.

*Pvu*II-digested pSPV3 DNA was used as the template. As controls, 5 ul of RNA from the transcription reaction was added to 45 ul of dH_2O, and 15 units of the RNase inhibitor RNasin, and incubated either at room temperature or 30°C. In the other reactions, the RNA from the transcription reaction

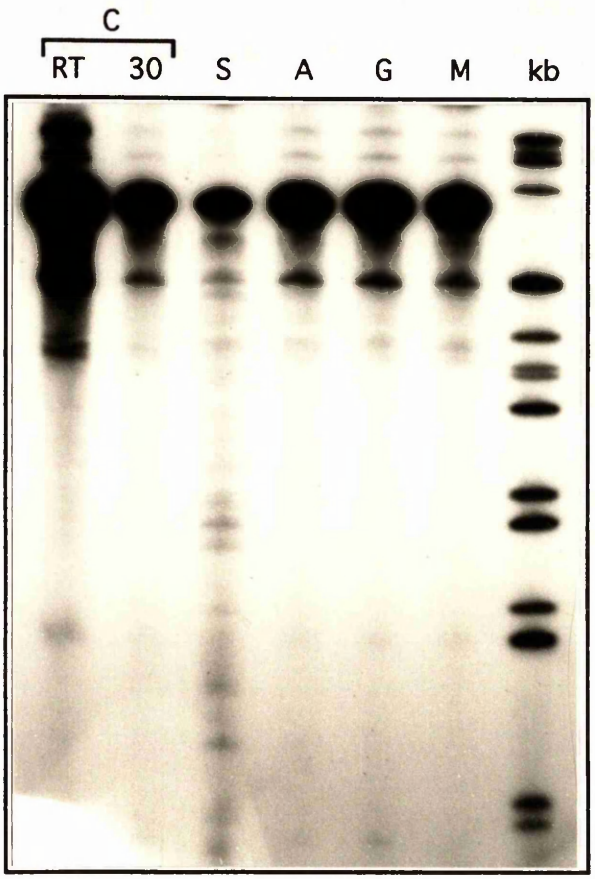


FIGURE 5.9: Autoradiograph showing the products of *in vitro* transcription reactions analysed by polyacrylamide/urea gel electrophoresis. *Pvu* II-digested pSPV3 DNA was used as the template, and following the transcription reaction, aliquots of the reaction mixture were incubated under various conditions. These are shown above the lanes: S - 1 x SP6 buffer; A - 500 μM ATP; G - 500 μM GTP; M - 100 mM $MgCl_2$. Incubation mixtures A, G, & M also contained 1 x SP6 buffer. The control incubations were carried out in dH_2O at 30^oC or room temperature (RT), as indicated. BRL 1 kb ladder was used as a size marker (kb).

was incubated at 30°C in 1 x SP6 transcription buffer only, or with a final concentration of either 500 uM GTP, 500 uM ATP, or 100 mM $MgCl_2$. After a 30 min incubation the RNA was precipitated, and then subjected to polyacrylamide/urea gel electrophoresis.

The only reaction which showed any alteration in the ratio of run-off product to shorter RNAs was that containing SP6 reaction buffer (Fig. 5.9; lane 5). This gave a number of extra RNA bands, which were less than 220 nt in length. It does, however, seem unlikely that this is the result of processing, since it did not lead to the production of a specific band, and the reaction was blocked by the addition of GTP, ATP and $MgCl_2$. The production of these extra RNA bands was probably the result of nuclease activity nicking the RNA at certain points, which occurred either during the incubation period, or the subsequent manipulations. The results of these 2 experiments do not support the proposal that a specific self-processing reaction generates the prominent minor RNA bands observed in the pSPV3 transcription reactions.

5.3.4 SEQUENCE SPECIFICITY OF TERMINATION SITES.

On the basis of the 2 experiments described in the proceeding section, it appears that the non-full-length RNA molecules seen on transcription of pSPV3 were due to premature termination by the RNA polymerase. It was therefore necessary to determine whether termination was in response to specific sequences within the mtDNA, or dependent on the actual distance from the promoter, or to the end of the template. In order to address this question, the 2 families of plasmids pAM6-35 (section 4.3.2; Fig. 4.6b), and pAMV1 (section 5.2; Fig. 5.3b), produced by exoIII deletions from opposite ends of the mtDNA insert, were used. If the position of the termination site(s) for *in vitro* transcription was determined by the distance from either the promoter, or the end of the template, then the pattern of products observed with both

deletion series would be the same. If, however, termination was sequence dependent, then the pattern would differ for the 2 deletion series.

The pAMV1 family of plasmids has deletions of the mtDNA insert from the PvuII site within A6 (nt 9148). In vitro transcription reactions were carried out with both EcoRI- and PvuII-digested plasmid DNA, and both gave the same results. Two bands, representing the run-off and the terminated products of transcription were observed with plasmids pAMV1, pAMV2, and pAMV3, the first 3 members of the deletion series (Fig. 5.10a). However, with pAMV4, and pAMV5, only the run-off transcript was seen. The size of the non-full-length RNA molecules decreased from 510 ± 10 nt with pSPV3 template, to 470 ± 10, 410 ± 10, and 150 ± 10 nt for pAMV1, pAMV2, and pAMV3 respctively. In each case the deletion decreased the length of the plasmid template by 30, 90, and 350 nt respectively. The size decrease in the run-off product therefore mirrored that seen for the shorter RNA molecules, and when the difference in the length between the transcribed polylinker sequence in pSPV3, and its derivatives is taken into account (40 nt compared to 15 nt), the stop site maps to the same DNA sequence (nt 9618). Of course, the fact that the stop site might simply be a fixed distance from the end of the template cannot be ruled out at this stage. The shorter RNA molecules produced by pAMV4 would only be 30 nt long, and would be too small to appear on the gel.

The less prominent termination products, which gave the 360/380 nt doublet with EcoRI-cut pSPV3, are also affected by the deletions (Fig. 5.10a). The products are clearly visible in the EcoRI-digested pAMV2 reaction, and are 240 and 260 nt in size, and so show almost the same reduction in size as the larger termination product.

The pAM6-35 deletion series was used to confirm that termination was sequence-specific. If the termination sites were merely determined by the distance from the 3′ end of the mtDNA template, then a similar pattern to that seen with the

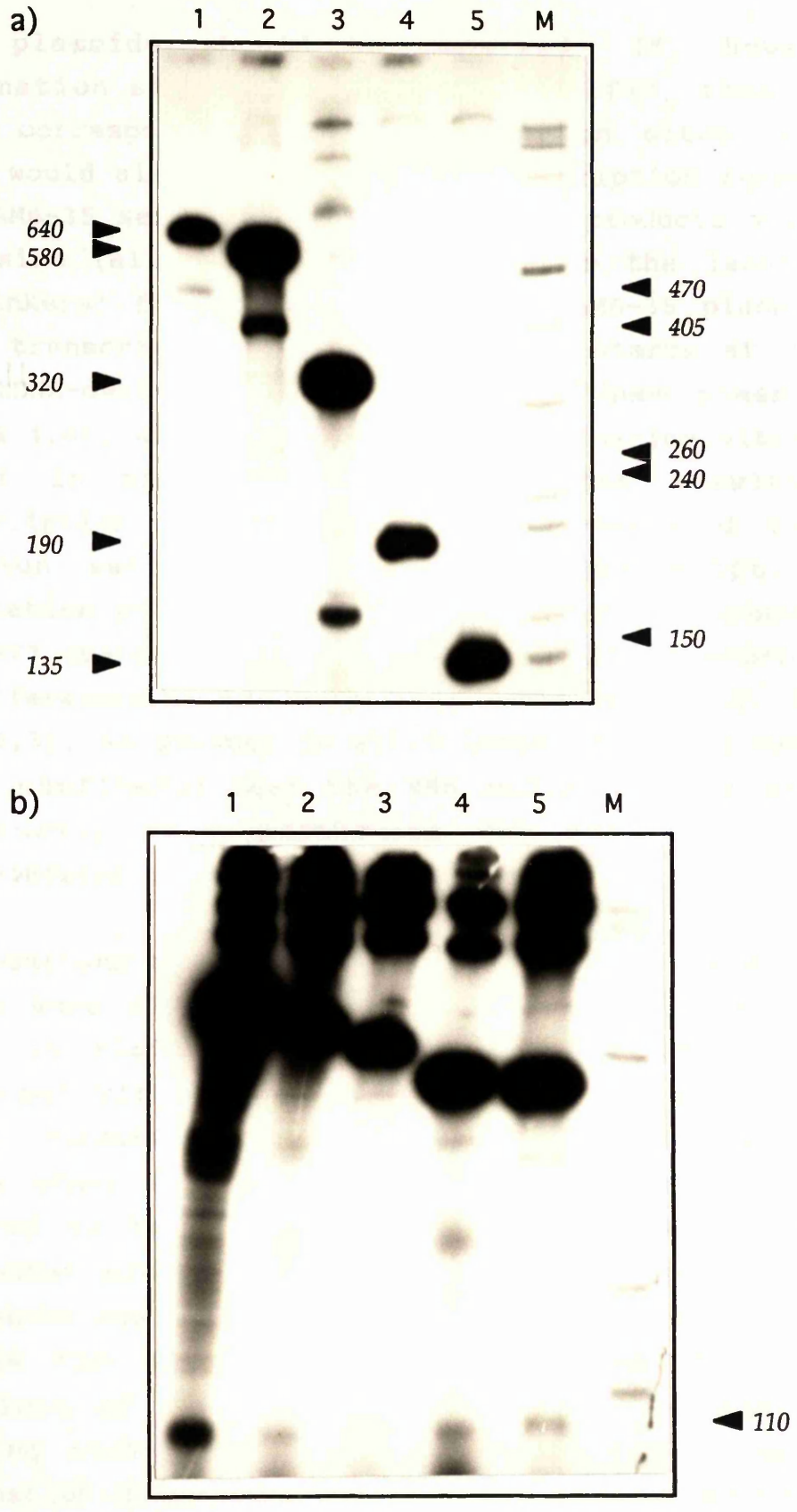
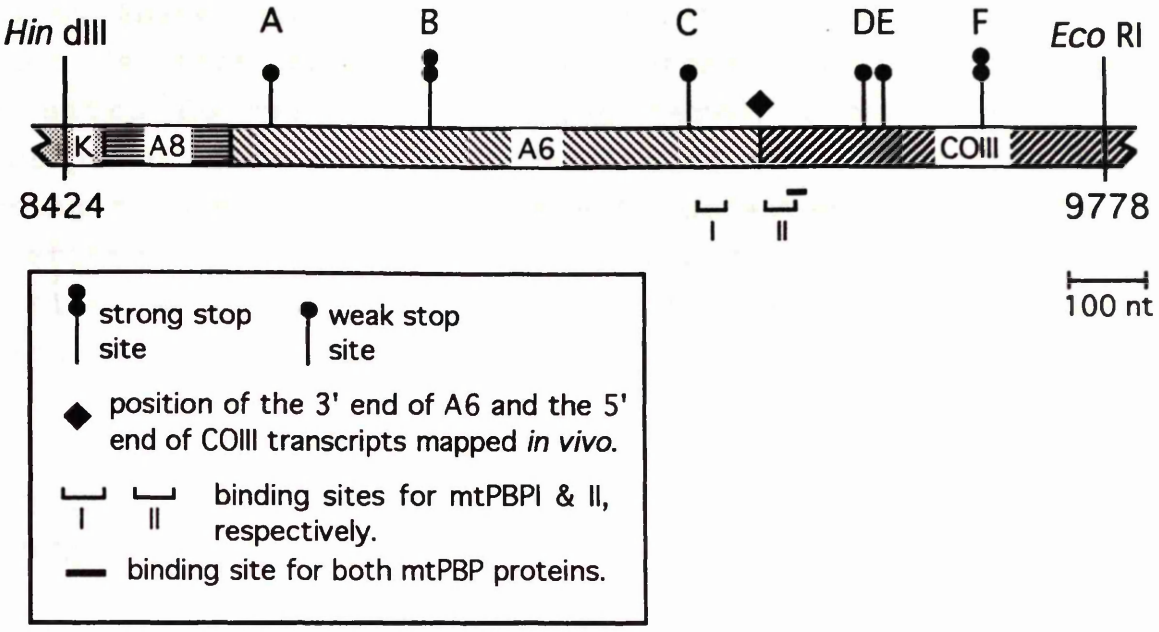


FIGURE 5.10: Autoradiograph showing the products of *in vitro* transcription reactions analysed by polyacrylamide/urea gel electrophoresis. (a) *Eco* RI-digested plasmid DNA: lane 1 - pSPV3; 2-6 - pAMV1-pAMV5. (b) *Pvu* II-digested pAM6-35 exoIII deletion series plasmids: 1 - pAM6-35; 2 - pAM6-43; 3 - pAM6-55; 4 - pAM6-62; 5 - pAM6-65. BRL 1 kb ladder was used as a size marker (M). The arrows show the positions and sizes (in nt) of the bands discussed in the text: those corresponding to the run-off products are shown on the left; those for termination products on the right.

pAMV plasmids should be observed. If, however, the termination sites were sequence-specific, then the minor bands corresponding to the termination sites observed in pSPV3 would also be detected in transcription reactions with the pAM6-35 series of plasmids. These products would be the same size (allowing for differences in the lengths of the polylinkers) for both pSPV3 and the pAM6-35 plasmid series, since transcription of the insert DNA starts at nt 9148 of the mtDNA-derived sequence in all of these plasmids (Figs. 4.4b & 4.6), although some of the termination sites would be absent in the shorter plasmids. The results of the transcription reactions with the 5 members of the pAM6-35 deletion series are shown in Figure 5.10b. A small termination product of 110 $\pm$ 10 nt, which is sometimes seen in pSPV3 transcription reactions as a 135 nt product because of differences in the polylinker sequences (Figs. 5.9 & 5.5; lane 2,3), is present in all 5 lanes, and does not alter in size, confirming that the RNA polymerase is terminating prematurely at a particular DNA sequence within the mitochondrial inserts.

The locations of these DNA sequences within the mitochondrial genome were calculated on the basis of of the experiments shown in Figures 5.5 and 5.10, and other experiments performed with the constructs pSP3 and pSP31 (data not shown). Termination at 2 of the sites was much stronger than at the other 4 (based on the levels of non-full-length RNA compared to the full-length transcript) so the sites were designated either strong or weak sites accordingly. Figure 5.11 shows the positions of the termination sites in relation to the the gene boundaries, and gives the nucleotide positions of the sites. A region of the mtDNA sequence covering each site is also shown, and for 5 out of the 6 termination sites, the end of a run of 5 or more T residues is present within 5 nt of the site. The implications of these T runs (which are characteristic of phage T7 transcriptional terminators) are discussed below (section 5.5).

None of the 6 termination sites were found to map precisely



SITE	NUCLEOTIDE POSITION	SEQUENCE
A	8693	TTTTTTATTCCAATGAATGTATT
B	8903	ACGTCTTGGGCCTTTTTCCCCC
C	9242	TTTTTATTTTTATTATTTGTGC
D	9468	TTTAGTAGGCTTTTTATTATT
E	9493	ACAAAAATGGTTAACTGGTGA
F	9618	TTTTTTTTCGCCTTTTTTTGGGC

FIGURE 5.11: Positions of the 6 termination sites for SP6 transcription identified by the *in vitro* transcription reactions. The upper portion of the figure shows the region of the mitochondrial genome studied, indicating the genes, the positions of the termination sites and the coordinates of the enzyme sites flanking the region. The position of the ends of the A6 and COIII transcripts mapped *in vivo* (section 4.5.1) and the binding sites for the proteins mtPBPI & II (section 4.6.3) are also given. Below, the table identifies the nucleotide positions (± 10 nt) of the stop sites, which are shown underlined on the portion of DNA sequence.

to the gene boundaries between A6 and A8, or A6 and COIII, or to correspond to the ends of the transcripts seen *in vivo*, as mapped by RNase protection experiments to the A6/COIII gene junction (section 4.5.1). This therefore suggests that if these sites do represent *in vivo* termination sites for transcription, then further processing of the transcripts is occurring to generate the stable RNAs detected *in vivo*. The implications of these termination sites for transcriptional regulation, and DNA replication are discussed further below (section 5.5).

5.4 EFFECTS OF MITOCHONDRIAL PROTEIN EXTRACTS ON TRANSCRIPTION.

Two DNA binding proteins have been identified (S.A.Qureshi & H.T.Jacobs, unpubl. data) which interact with the mitochondrial genome close to the boundary between the genes for A6 and COIII (section 4.6.4). To investigate whether one of these proteins (mtPBPI) could affect transcription by its interaction with the DNA, *in vitro* transcription reactions were performed in the presence of this protein. A parallel reaction was also carried out using an equal volume of the buffer solution in which the protein sample was stored, to control for any effects the buffer solution might have on transcription. *Eco*RI-digested pSPV3 DNA, and *Pvu*II-digested pSPV31 DNA were used as the templates for the transcription reactions, but in each case the addition of the protein sample produced no effect on the transcription products observed (data not shown). The protein sample used was the 0.2-0.4 M KCl fraction from a heparin-sepharose column (prepared by S.A.Qureshi), which was then heated to destroy the activity of the other proteins present in the fraction. Gel retardation assays performed using heat-treated aliquots of the same protein sample and an oligonucleotide containing the protein binding site generated the complex characteristic of mtPBPI binding, showing that the protein was active (S.A.Qureshi, pers. comm.).

The lack of any effect of mtPBPI on transcription *in vitro* suggests that the protein does not act as a transcription terminator *in vivo*, although a system would have to be developed using mitochondrial RNA polymerase to determine its effect on the mitochondrial enzyme. Recent gel retardation experiments using the protein fraction containing mtPBPI revealed the formation of a different, higher molecular weight complex when the protein sample was not heated, compared to the heated sample (S.A. Qureshi, pers. comm.). This result suggests that protein/protein interactions could also be important in the binding of mtPBPI to the DNA, raising the possibility that other proteins are required for the action of mtPBPI *in vivo*. The effects of these other factors on transcription *in vitro* could be tested by the addition of the unheated protein fraction to the reaction.

5.5 DISCUSSION.

The *in vitro* transcription analysis of the region of the mitochondrial genome covering the genes for A8, A6, and the 5′ half of COIII showed that certain RNA products, shorter than the full-length (run-off) RNA transcript were produced in the reactions. As mentioned above, the synthesis of some of these RNA molecules was shown to be dependent on the restriction enzyme used to linearise the DNA template for the reaction. This was the case for some *Sst*I-digested molecules, and may have resulted from the initiation of transcription at the 3′ overhanging end, rather than at the SP6 promoter.

Analysis also revealed 6 reproducible non-full-length RNA products, which were generated by *in vitro* transcription of the mtDNA insert in the same direction as the genes would be transcribed *in vivo*. The production of these RNAs was not dependent on the restriction enzyme used, and could either have been the result of transcriptional termination, or processing of RNA transcripts. Two different experiments failed to produce any evidence in favour of RNA processing, which was consistent with the view that the non-full-length

RNAs were the result of transcriptional termination. By using 2 different deletion series, the termination events were shown to be sequence-specific and 6 termination sites were mapped within this region of mtDNA. Of the 6 sites, 3 were located in the gene for A6, and 3 in COIII, with one strong site in each gene. Their locations within the genes (Fig. 5.11) give no obvious indication as to their role *in vivo*, and none of the stop sites located within the mtDNA region studied by RNase protection (sites C-F; nt 9148-9778) occurs at the position of the transcript ends mapped by RNase protection experiments (section 4.5.1). Two questions therefore have to be addressed when considering these results. Do these sites act *in vivo* to terminate mitochondrial transcription? And if so, what is their mechanism of action?

It is possible that the termination sites are simply the result of SP6-like (and T7-like) termination sequences present in the region of mtDNA studied, and that these sequences would not affect mitochondrial RNA polymerase, and are therefore irrelevant to mitochondrial transcription *in vivo*. *In vitro* transcription of the bovine preproparathyroid (PTH) gene, using both SP6 and T7 RNA polymerase (Mead *et al*, 1985; 1986) gave rise to truncated transcripts, with termination occurring at a T-rich region of the gene. Mutagenesis of the sequence from TGTTTTCTT to TGCGCGCTT abolished termination (Mead *et al*, 1986), implying that the T residues were involved in termination. Studies on T7 transcription termination have shown that runs of T residues are present at termination sites (Dunn & Studier, 1983; Christianson, 1988). Reductions in the number of T residues have also been shown to decrease termination efficiency (Jeng *et al*, 1990), although the termination site in the phage T7 genome (*e.g.* To) does allow some readthrough (Dunn & Studier, 1983). The T7 RNA polymerase also recognises some rho-independent termination sites in *E. coli* operons (*e.g.* *thra*, *rrnC*; Chen & Orozoco, 1988), stopping at or within a few nucleotides of the site recognised by the host polymerase. These sites are also recognised by spinach chloroplast RNA

polymerase *in vitro* (Chen & Orozoco, 1988).

Analysis of the DNA sequences at the termination sites inferred in this study showed that all except one of the sites (site E; nt 9493) occurred within 5 nt of the end of a run of 5 or more consecutive T residues. There are not, however, potential hairpin structures adjacent to these T runs, which are normally found in termination sites recognised by T7 RNA polymerase (Dunn & Studier, 1983; Christianson, 1988), but the studies with the PTH gene showed that hairpin structures are not required for termination (Mead *et al*, 1985; 1986). The termination sites seen in mtDNA could therefore be viewed in the same light as that present in the PTH gene, with termination not occurring at these sites *in vivo*.

Against this interpretation is the observation that these sites are only active when the genes are transcribed in the same direction as occurs *in vivo*. If termination is due to the presence of T runs in the mtDNA sequence, then runs of T residues on the opposite strand should affect transcription in the opposite direction. There are 5 runs of 5 or more T residues within the region studied (nt 8424-9778), but none of these has an effect on RNA polymerase, and one even occurs only 4 nt away fron the sole termination site on the other strand lacking a T run. However, 2 other runs of T residues are present on the major coding strand at which termination does not occur, suggesting that other sequences, or transcript structure may be involved as well.

The absence of termination sites on the minor-coding strand suggests that the termination sites seen *in vitro* could have a role in mitochondrial transcription *in vivo*, assuming that they are also recognised by the mtRNA polymerase. The similarity of the yeast mtRNA polymerase to those of phage SP6 and T7/T3, in terms of structure and amino acid sequence of the catalytic domains makes this possible (Masters *et al*, 1987). This is supported by the termination of both the SP6 and T7 RNA polymerases, which are not closely related, at the

T run in the PTH gene, and at the same sites within sea urchin mtDNA.

Based on these assumptions, it seems conceivable that these termination sites are involved in the processing of the RNA transcripts produced in this region. The 3 genes (A8, A6, and COIII) are not separated by tRNA genes, so the generation of mature message for each requires either the termination of transcription downstream of each gene, or the processing of a polycistronic message, or a combination of the 2 processes.

The positioning of the in vitro termination sites within the A6 and COIII genes would allow transcriptional termination to generate the different mRNAs, with readthrough of the termination sites allowing transcription of the downstream genes. Sites A and/or B (and possibly C) could be termination sites for A8 (though in other metazoa A6 and A8 are on a common transcript), and some or all of sites D-F could act as terminators for the A6 gene (Fig 5.11). Although this could generate transcripts with alternative 3′ ends, the function of which is unclear, genes have been found in other systems in which multiple 3′ (or 5′) transcription termini have been observed. These include c-myc (different promoters; Battey et al, 1983) and glutamine synthetase (different polyadenylation sites; Gibbs, 1988)*.

There are also examples where the primary sequence of the transcript (i.e. the use of different 5′ or 3′ ends) affects its processing pathway. One such case is the production of the mRNAs for membrane and secretory forms of immunoglobulin (Early et al, 1980; Rogers et al, 1980). The 2 forms of the protein differ at their carboxy terminus, which is reflected in the production of either long or short primary transcripts, which differ at the 3′ end. The longer transcript includes 2 membrane-specific exons which form the C-terminus of the membrane-bound protein. In the short form of the transcript, these exons are absent, so the secretory form of the protein is produced. The production of the 2 forms of the protein is therefore dependent on the cleavage

* ref listed on p219.

and polyadenylation site used to produce the primary transcript.

The *in vivo* mapping of A6 and COIII transcripts by RNase protection did not, however, detect multiple 3′ ends for A6 RNA. This may simply reflect the subsequent processing of the transcripts in the generation of stable mature message, with the intermediate forms too unstable to be detected. Although the generation of mature mRNAs for A6/A8 and COIII by transcription termination would appear to be wasteful in terms of the amount of RNA synthesised and discarded, the overlap of the 16S rRNA and the COI genes in the *S. purpuratus* mitochondrial genome (Jacobs *et al*, 1988) provides a precedent for this. The major part of the 16S rRNA gene transcript is redundant in the generation of COI mRNA, since the nearest likely promoter for COI is upstream of the 5′ end of the 16S rRNA gene, and the production pathways for the 2 RNAs are mutually exclusive (Elliott, 1990).

The transcriptional stop sites could, however, represent sites *in vivo* where RNA polymerase pauses, rather than terminates transcription. The stalling of RNA polymerase could be required to allow the transcripts produced to adopt the correct conformation for subsequent processing, or for translation to occur.

In the replication of the plasmid ColE1, the pausing of transcription of the RNA II molecule affects its structure, which influences its ability to act as a primer for DNA replication (Polisky *et al*, 1990). Mutations in RNA II, which affect plasmid copy number, have been shown to produce altered transcription termination patterns compared to wild-type RNA II *in vitro*. In phage lambda, pausing of the RNA polymerase complex is required for the action of antitermination proteins, which prevent the polymerase from recognising certain terminators (Yager & von Hippel, 1987). This occurs in the synthesis of both the delayed early and the late genes in the lytic phase of the phage lifecycle. The pausing of RNA polymerase is also needed in the regulation of

some amino acid biosynthetic operons in bacteria (e.g. trp operon of E. coli), to allow translation of the leader peptide to occur. Depending on the extent of the translation of this peptide, which is influenced by the cellular levels of the amino acid synthesised by the operon, the transcript either adopts a termination-competent or an antitermination structure, and so either halts transcription, or allows it to proceed through the operon (Landick & Yanofsky, 1987).

In the bacteriophages MS2 and QB, the expression of their 4 genes is controlled by the structure of the RNA molecule, which forms both the transcript and genome (Singer & Berg, 1991). Only the gene for the coat protein (CP), which is required in large amounts, has an initiation codon which is accessible to the ribosomes. The translation of the replicase (rep) gene, and the lys gene (for host cell lysis) relies on the initiation of translation of the CP gene, so the Rep and Lys proteins are produced in much smaller amounts than CP. Protein A, which is needed for entry of the progeny phage into the host cell, and is required in very small amounts, can only be produced as the 5′ end of the RNA molecule is synthesised.

It is possible that the transcriptional pause sites in S. purpuratus mtDNA could, if active in vivo, hinder the movement of DNA replication forks since the RNA polymerase would be attached to the DNA. Any transcriptional regulation mechanism which acted by the pausing of RNA polymerase at the stop sites detected in vitro could have a major effect on the movement of replication forks, and lead to the polymerase interference mentioned at the start of this chapter (section 5.1).

CHAPTER

SIX

CONCLUDING REMARKS

6.1 MtDNA REPLICATION MECHANISMS.

6.1.1 MtDNA REPLICATION IN S. PURPURATUS - A MODEL.

The analysis of replicating mtDNA molecules by a combination of gel electrophoretic techniques (described in Chapter 3) and RNase protection experiments (Chapter 4) has led me to formulate the following model for the replication of the Strongylocentrotus purpuratus mitochondrial genome. The leading-strand replication origin is located within the 121 bp non-coding region of the genome, mapping the 5′ end of the leading-strand to nt 1150 ± 10. This position is just 3′ to the G-string within the non-coding region, and is where the 5′ ends of the D-strand molecules were mapped in S. purpuratus egg mtDNA (Jacobs et al, 1989). DNA synthesis is initially unidirectional by D-loop expansion towards the 12S rRNA gene. The termination site for leading-strand synthesis is located within the same SstI restriction fragment (6.9 kb) as the leading-strand origin, probably at or just upstream of it (Fig. 3.12).

Pause sites for leading-strand DNA synthesis occur at several locations around the genome, the most prominent of which lies in the region of the A6/COIII gene boundary, where a prominent lagging-strand replication origin was located. The pause sites for leading-strand DNA synthesis could correspond to the sites of lagging-strand replication initiation, although not all may be active in any given mtDNA molecule. Alternatively, the replication pause sites may be the result of transcription of the mtDNA interfering with the replication process (see section 6.3).

I propose that a pause site for lagging-strand synthesis occurs at the leading-strand replication origin (as implied by the data presented in sections 3.4.2, 3.5 and 4.4.1), and it could be involved in the control of mtDNA replication by ensuring that a second round of mtDNA replication cannot initiate before the first has been completed (see section 6.2).

It must, however, be noted that the experiments described in this thesis were carried out using mtDNA isolated from activated enucleate egg fragments (AEEFs; section 3.2). MtDNA replication has been artificially stimulated in these egg fragments (Rinaldi *et al*, 1979a), so the question arises as to whether the results are applicable to mtDNA molecules replicating under physiological conditions *in vivo*.

The idea that the results are applicable to mtDNA replication under physiological conditions is supported by the fact that the results are consistent with what was already known about *S. purpuratus* mtDNA replication in oocytes (Matsumoto *et al*, 1974). However, removal of the nucleus may affect mtDNA replication by blocking the production of nuclear factors involved in the replication process. These nuclear factors also may not be present in the egg at the same levels as occur in, say, oocytes (in which mtDNA replication is extensive), which could account for the low levels of RIs observed in the 2D gel analyses.

In order to determine whether the replication scheme proposed for AEEF mtDNA is applicable to mtDNA replication under physiological conditions *in vivo*, the results would have to be compared with the results of similar experiments carried out with naturally replicating mtDNA. Oocytes would be one possible source of such material.

6.1.2 COMPARISONS BETWEEN mtDNA REPLICATION MECHANISMS IN DIFFERENT ORGANISMS.

The replication mechanism proposed for *S. purpuratus* mtDNA is broadly similar to that for vertebrates (Clayton, 1982) and *Drosophila* (Goddard & Wolstenholme, 1978; 1980), in that there is a single leading-strand replication origin, and replication of the leading strand is unidirectional, and by strand displacement. Where the *S. purpuratus* mechanism differs from that of vertebrates and *Drosophila* is in the presence of multiple lagging-strand origins, compared with

the one seen in vertebrate and Drosophila mtDNA. Cantatore et al (1989) also propose a single lagging-strand replication origin for the mtDNA of the sea urchin P. lividus, close to the A6/COIII gene boundary, though this is based solely on the potential of a region of the genome to adopt a stem-loop structure resembling that of the vertebrate O_L region. A similar argument was used by Jacobs et al (1988) to propose that a stem-loop structure in the non-coding region of S. purpuratus mtDNA could act as an O_L.

Although the non-coding regions of the sea urchin mitochondrial genomes of S. purpuratus (Jacobs et al, 1988) and P. lividus (Cantatore et al, 1989) are much smaller than those in vertebrate mtDNA (121 and 132 bp respectively, compared to 1122 bp in humans; Anderson et al, 1981), D-loop forms have been detected in egg mtDNA from S. purpuratus (Jacobs et al, 1989). The D-strand molecules lie within the non-coding region, and were 65-75 nt in length, much shorter than the D-strand molecules of vertebrates. The reduced size of the non-coding region and D-strand molecules in S. purpuratus compared to vertebrates may be a reflection of the need for a larger non-coding region in vertebrates because of its role in mitochondrial transcription (Chang & Clayton, 1984; Clayton, 1991). This could also explain the production of D-strand molecules of different, discrete lengths in vertebrates (Doda et al, 1981), since they may affect the secondary structure of the non-coding region in different ways, which could in turn influence transcription.

Although mitochondrial DNA replication in Drosophila is by a strand-displacement mechanism, no D-loop forms have been detected in mtDNA by electron microscopy (Goddard & Wolstenholme, 1978). This could be due either to their loss during the purification of the DNA samples, or to their absence in the mtDNA replication mechanism. The non-coding region of Drosophila mtDNA is very A+T-rich (90-95%; Clary & Wolstenholme, 1985), in marked contrast to the non-coding regions of vertebrate and sea urchin mtDNAs, which contain stretches of highly G+C-rich DNA. In mouse mtDNA, the CSB II

sequence involved in the switch from primer to DNA synthesis is 75% G+C (Chang & Clayton, 1987a; Bennett & Clayton, 1990), and in S. purpuratus mtDNA the non-coding region contains a run of 20 consecutive G residues, lying just upstream of the 5′ ends of the D-strands (Jacobs et al, 1989) and the nascent DNA molecules. This difference in sequence characteristics between the non-coding regions of Drosophila species on the one hand, and vertebrates and sea urchins on the other, could possibly reflect differences in the control of mtDNA replication between the 2 groups.

In Saccharomyces cerevisiae, studies using petite mutants have revealed 4 active origins of DNA replication in the mtDNA (Baldacci et al, 1984). The mapping of DNA and RNA ends at 2 of these origins (ori1 and ori5) has shown that only short RNA primers are used (5-15 nt), and that the ends of the leading- and lagging-strand DNA molecules are only 4 bp apart (Baldacci et al, 1984). Replication is therefore essentially bidirectional, but it is not clear whether it is semi-discontinuous (Okazaki fragments) or continuous on each strand (bidirectional strand displacement). Studies of replication intermediates from the yeast Torulopsis glabrata suggested a rolling circle mechanism of mtDNA replication (Maleszka et al, 1991), as also occurs in kDNA maxicircles (Hajduk et al, 1984). This is a similar mechanism to strand-displacement, except that a nick is introduced into the template strand at the replication origin (Gilbert & Dressler, 1968). Synthesis is continuous on each strand, and the appearance of single-stranded displaced tail regions indicates that the lagging-strand origin(s) are separate from that of the leading-strand, and act when single-stranded (Maleszka et al, 1991). It is possible that S. cerevisiae mtDNA replication could occur by the rolling circle replication mechanism, which would explain the failure to observe D-loop forms of mtDNA, and the high percentage of linear mtDNA molecules seen by PFGE and electron microscopy (Skelly & Maleszka, 1991; Bendich & Smith, 1990).

A rolling circle replication mechanism for S. purpuratus

mtDNA is not, however, consistent with the EM data of Matsumoto et al (1974), or the 2D gel electrophoresis data presented in Chapter 3. The pattern of RIs predicted for linearised (BglII-digested) mtDNA on the basis of a rolling circle mechanism would differ from that observed in Figure 3.4 (section 3.4.1), in that the curve of RIs would not contain a clear discontinuity, and would return towards the diagonal of linear DNA molecules once the genome was more than half-replicated.

In vertebrates, leading-strand mtDNA replication is primed by mtRNA polymerase (Schinkel & Tabak, 1989), whereas at the lagging-strand origin a primase is used (Wong & Clayton, 1986). The location of a putative promoter (TTATATAA-like) sequence within the non-coding region of S. purpuratus mtDNA, some 40 bp upstream of the mapped replication origin, makes it likely that RNA polymerase also primes sea urchin mtDNA replication at the leading-strand origin. This is supported by the detection of ribonucleotides within the D-strand molecules (Jacobs et al, 1989). In S. cerevisiae, a transcriptional promoter sequence is adjacent to the ori sequences (Christianson & Rabinowitz, 1983), but transcription is only detected from the promoter in the active replication origins (Baldacci & Bernardi, 1982). Mutations that disrupt the mitochondrial RNA polymerase gene (RPO41) show a loss of mtDNA, indicating that transcription of mtDNA is essential for its maintenance (Greenleaf et al, 1986). Both of these results point towards the priming of leading-strand DNA synthesis by mtRNA polymerase.

It therefore appears that the replication mechanism of mtDNA is broadly similar in different organisms, and that it differs markedly from the semi-discontinuous replication of prokaryotic and eukaryotic nuclear DNA (Kornberg, 1980). In mtDNA molecules, there is a single active leading-strand replication origin, and continuous DNA synthesis on the leading-strand, primed by RNA synthesised by mtRNA polymerase. One, or several lagging-strand origins exist, which appear to be active only when made single-stranded by

strand-displacement. DNA synthesis between these origins is continuous, and in mammals lagging-strand synthesis is primed by a primase enzyme.

6.1.3 EVOLUTION OF mtDNA REPLICATION MECHANISMS.

Given the overall similarity between the mtDNA replication mechanisms of different organisms, and the marked difference from the semi-discontinuous synthesis of prokaryotic and eukaryotic genomic DNA, it is interesting to consider why the mtDNA replication system is so different. Does the mtDNA replication system represent a relic of the replication mechanism of the ancestral bacterial endosymbiont, or has it arisen in response to the particular demands of the mitochondrial genome?

In both *E. coli* and the eukaryotic nucleus, more than one type of DNA polymerase is required to replicate the DNA, whereas only one (polymerase γ) is believed to be used in mtDNA replication (Baker & Kornberg, 1991). In *E. coli*, the continuous synthesis of the leading-strand, and the majority of lagging-strand synthesis is carried out by DNA polymerase III (pol III), but pol I is required to fill in the gaps on the lagging-strand, and remove the RNA primers. A pol II enzyme is also present in *E. coli*, but its role in replication (if any) is unclear, since mutants lacking this enzyme appear to grow normally. In the nuclei of higher eukaryotes, pol δ is responsible for continuous DNA synthesis on the leading-strand, whereas pol α synthesises the Okazaki fragments on the lagging-strand (Blow, 1989). Two other DNA polymerases (β and ϵ), as well as the mitochondrial polymerase (γ) are also found in higher eukaryotes (Hubscher & Thommes, 1992). In *S. cerevisiae*, 2 DNA polymerases (I and III) have been cloned which are structurally and biochemically homologous to pol α and δ, and have been shown to be involved in yeast chromosomal DNA replication. A third polymerase (II), which is homologous to mammalian pol ϵ, also appears to be involved in chromosomal replication (Araki *et*

al, 1992). Although its homologue (pol ϵ) was originally thought to be involved in DNA repair in human cells (Nishida et al, 1988), the conservation of structure and function between the yeast and mammalian polymerases suggests that pol ϵ may have a role in chromosomal replication (Araki et al, 1992).

The adoption of a replication mechanism in which the synthesis of both strands is completely or largely continuous could therefore mean that only one DNA polymerase is required for DNA replication. This would reduce the number of different proteins imported into the mitochondrion for DNA replication, and consequently fewer nuclear genes would have to be expressed to produce the active replication machinery. However, the large number of proteins needed for mitochondrial RNA processing and translation would seem to invalidate this argument. The similarities between the different mtDNA replication mechanisms imply either that they resemble the replication mechanism of the original endosymbiont, or the mechanism that evolved in the proto-mitochondrion, before the divergence of different eukaryotes. The possibility of convergent evolution, however slim, must also be considered.

The occurrence of semi-discontinuous DNA replication in prokaryotes and eukaryotes argues that it was the mode of DNA replication in the endosymbiotic ancestor of mitochondria. This in turn suggests that the present-day replication mechanism of mitochondrial DNA reflects the specific requirements of the genome. The generation of extensive regions of single-stranded DNA, as the result of a strand displacement replication mechanism would be predicted to increase the risk of damage to the DNA resulting in mutations. This may be tolerated in the mitochondria because multiple copies of the genome are present.

The rolling circle replication mechanism used by some phages (e.g. T4, M13; Kornberg, 1980) resembles the replication mechanism of some mtDNAs, as mentioned above. Other

similarities between the replication mechanisms of mitochondria and phages include the priming of DNA replication by RNA polymerase, rather than by a primase (Fuller et al, 1987), and the similarity at the amino acid level between the RNA polymerases of yeast mitochondria, and the different bacteriophages T7, T3 and SP6 (Masters et al, 1987). This has led to speculation that a phage replication and transcription machinery present in the original bacterial endosymbiont was hijacked for the use of the mitochondrial ancestor (Schinkel & Tabak, 1989; Jacobs, 1991). The transfer of the phage genes for the replication and transcription machinery, from a phage genome integrated into that of the proto-mitochondrion, to the nuclear genome of the "host" cell, could, in theory, provide a mechanism to ensure that the regulation of nuclear and mitochondrial DNA replication occurs by different, non-interfering mechanisms, because the 2 sets of genes are unrelated.

6.2 THE CONTROL OF mtDNA REPLICATION.

The replication of mtDNA in somatic cells has to be regulated so that the number of mtDNA molecules, and of mitochondria is doubled for each cell division. In early animal development, the situation is modified (section 1.3.6), so that in oogenesis there is extensive mtDNA replication but no nuclear DNA replication (Matsumoto & Piko, 1971; Webb & Camp, 1979; Webb & Smith, 1977). This establishes a pool of mitochondria, which are then partitioned in the early stages of development after fertilisation, when there is rapid nuclear DNA replication and division, but no mtDNA replication (Bresch, 1978; Chase & Dawid, 1972; Rinaldi & Giudice, 1985). This therefore requires a modification of the normal somatic cell regulation mechanism for mtDNA replication in early development.

In animals, there seem to be at least 4 possible steps at which mtDNA replication could be regulated: the start of RNA primer synthesis, RNA processing to form the primer molecule,

the initiation of DNA synthesis from the primer, and D-loop termination. In Drosophila mtDNA, however, no D-loops have been detected (Goddard & Wolstenholme, 1978), though this is probably the result of the isolation procedure for the mtDNA.

There is little known about the regulation of mtDNA in animals (the most intensively studied replication systems), though there is evidence to support the idea that the block to mtDNA replication in early development is caused by the presence of the nucleus in sea urchins and Xenopus (Rinaldi et al, 1979a; 1981). In mouse mtDNA, the rate of turn over of the 7S DNA molecules is higher than the frequency with which DNA replication is initiated (Bogenhagen & Clayton, 1978). This observation strongly suggests that the regulation of mtDNA replication is exercised at the level of the termination of D-loop synthesis, since only a proportion of the D-strand molecules are acting as replication primers. The detection of D-loop molecules in the eggs of S. purpuratus (Jacobs et al, 1989), mouse (Piko & Matsumoto, 1976) and Xenopus (Hallberg, 1974), in which mtDNA replication does not occur, also supports this view.

There are 2 possible mechanisms by which D-loop termination could be controlling DNA replication. D-strand synthesis could always terminate at a given stop site, and then DNA synthesis is either reinitiated, resulting in full genomic replication, or the D-strand molecule is degraded. Alternatively, the termination of D-strand synthesis is controlled, so that DNA synthesis either occurs through the termination site, leading to full genomic replication, or is blocked at this site, resulting in D-loop formation. In the sea urchin species P. lividus, a DNA binding protein has been identified which binds within the non-coding region (nt 1098-1126; Roberti et al, 1991), and in S. purpuratus, gel retardation assays using mitochondrial protein extracts and the non-coding region of the genome have revealed DNA-protein interactions (D.T.Segreto, pers. comm.). It therefore seems plausible that DNA-binding proteins are involved in the control of DNA replication at the level of D-loop

termination, although the precise mechanism is unclear.

Although, on the basis of the above observations, it appears that mtDNA replication in animals may be controlled by the regulation of D-loop expansion, there is the potential for an override control mechanism to block mtDNA replication in early development. This could be achieved by blocking the formation or processing of the RNA primer molecules, or the inhibition of DNA synthesis: all of which would result in the absence of D-loops in mtDNA molecules from early embryos. The presence of D-loops in egg mtDNA, in which mtDNA replication is also blocked, does, however, argue against the need for a second mechanism for the regulation of mtDNA replication in embryogenesis.

The detection of RNA transcripts covering the non-coding region of the genome in both rat (Sbisa et al, 1988; 1990) and P. lividus (Cantatore et al, 1990), which are complementary to the RNA primer, has raised the possibility of antisense regulation of mtDNA replication, as occurs in the ColE1 plasmid (Tomizawa, 1986). RNase protection experiments on S. purpuratus mtRNA (section 4.4.2) using a probe covering the non-coding region and some adjacent tRNA genes, detected 3 RNAs transcribed from the opposite strand to the RNA primer. The probe did not, however, map the RNAs precisely within this region of the genome, so it is not possible to say whether any of these transcripts are complementary to the replication primer. Antisense RNAs spanning the non-coding region could, in theory, act to affect primer formation, and so modulate the rate of D-loop synthesis, and DNA replication. However, given the possible regulation of mtDNA replication at the level of D-loop expansion, it is unclear as to what additional advantages an antisense RNA regulation mechanism would bring.

* ref listed on p219.

6.3 REPLICATION PAUSE SITES - A CONSEQUENCE OF THE CONTROL OF mtDNA REPLICATION OR TRANSCRIPTION?

The pause sites observed in mtDNA replication in S. purpuratus can be divided into 2 categories: those which lie within the non-coding region of the genome, and those which occur outside it. The pause sites within the non-coding region have a potential role in the control of productive mtDNA replication (section 6.3.1). Those sites elsewhere in the genome, which all (with one exception) mapped in the region of the 3′ ends of protein-coding genes (Figure 3.11), could either occur as the direct result of one aspect of the control of the mtDNA replication process, or as an indirect consequence of the transcriptional regulation of the mitochondrial genome, or a combination of both processes.

6.3.1 PAUSES WITHIN THE NON-CODING REGION.

The formation of D-loops in mtDNA, as detected in mitochondrial genomes from S. purpuratus eggs, requires DNA synthesis to be blocked at the 3′ end of the non-coding region. This termination of D-strand synthesis has to be regulated to differentiate between D-strand formation and productive DNA replication, and could therefore provide the mechanism by which the initiation of productive DNA replication is controlled both in the sea urchin and in other animal mtDNAs, as discussed above (section 6.2).

In S. purpuratus mtDNA, the gel electrophoresis and RNase protection data described in Chapters 3 and 4 (sections 3.4.2, 3.5 & 4.4.1) suggest that lagging-strand synthesis pauses within the non-coding region, at or close to the leading-strand replication origin. In vertebrate mtDNA replication, lagging-strand (L-strand) synthesis has also been observed to pause at the H-strand replication origin until H-strand synthesis is completed, and the daughter molecules separate (Bogenhagen et al, 1978). This halt to lagging-strand synthesis could possibly be due to the

inability of the lagging-strand replication complex to act on double-stranded DNA, or to the action of a termination factor at this site. In either case, lagging-strand pausing could provide a mechanism to ensure that the reinitiation of DNA synthesis or D-loop formation does not occur until replication is completed, by maintaining the origin in an inactive form. By controlling the timing of the final ligation of the gaps in the daughter strands at the H-strand origin, it would be theoretically possible to provide a cell-cycle control mechanism for mtDNA replication.

In kinetoplastid protozoa, minicircle replication could be controlled by such a mechanism. Each molecule appears to replicate only once per nuclear division, during the S phase of the cell cycle (Cosgrove & Skeen, 1970; Ryan et al, 1988), but the final ligation of the daughter strands occurs during the G2 phase of the cell cycle, when the network divides (Englund*, 1978).

If a mechanism does exist in sea urchins to link mtDNA replication to the cell cycle, then the prediction would be that this mechanism would be modified in oogenesis and embryogenesis, since the nuclear and mitochondrial DNA replication cycles are no longer linked.

6.3.2 INTERFERENCE BETWEEN REPLICATION AND TRANSCRIPTION?

The fact that the RNA primer for leading-strand DNA replication in animal mitochondria is synthesised by RNA polymerase means that the processes of replication and transcription are inescapably linked. In the yeast S. cerevisiae, there is also evidence to suggest that mtRNA polymerase is also involved in yeast mtDNA replication primer formation (section 6.1.2).

It is possible that the transcriptional organisation of the sea urchin mitochondrial genome is affecting mtDNA replication, with conflicts between the 2 polymerases causing

* ref listed on p219.

the stalling of replication fork movement, as has been observed in the nuclear rRNA gene cluster of the yeast S. cerevisiae (Brewer & Fangman, 1988). The replication pause sites detected in the S. purpuratus mtDNA (section 3.6) may therefore represent the sites of collisions between the RNA and DNA polymerases.

It might, however, be the case that the DNA polymerase is pausing at the sites of proteins bound to the DNA involved in the regulation of transcription. Transcription termination has been observed in vertebrate mtDNA at the 3′ end of the rRNA genes, allowing the production of higher levels of rRNA than mRNA from the HSP. The early EM observations of Bogenhagen et al (1973) on mouse mtDNA reported an accumulation of H-strand molecules of 13 kb, which would imply a replication pause near the transcriptional termination site at the 3′ end of the rRNA genes. Attenuation has also been observed in some of the transcription units of yeast mtDNA (Mueller & Getz, 1986), and could also involve DNA binding proteins. In the sea urchin P. lividus, mitochondrial transcription studies (Cantatore et al, 1990) showed that some of the transcripts had ragged 3′ ends taken to be indicative of transcription termination, rather than transcript processing.

Studies in S. purpuratus into the mitochondrial protein factors involved in DNA-binding around the A6/COIII junction region revealed several protein binding sites (S.A.Qureshi & H.T.Jacobs, unpubl. data), and 2 proteins (mtPBPI & II) have been identified with binding sites close to the gene boundary (section 4.6.3). These 2 proteins could be involved in the regulation of mitochondrial transcription, possibly terminating transcription at the 3′ end of the A6 gene, although the RNase protection experiments described above (section 4.5.1) point to RNA processing generating the 3′ end of the A6 message, and the in vitro transcription reactions (section 5.4) produced no evidence that mtPBPI binding affects polymerase movement.

The observations that mitochondrial protein levels are subject to post-transcriptional regulation by RNA stability in other organisms argues, in my opinion, against the regulation of transcription by the simultaneous binding of proteins at several sites in the genome: proteins which then interfere with replication fork movement. In HeLa cell mitochondria, RNA stability is sufficient to produce 20-150 fold higher levels of H-strand tRNAs compared to mRNAs (Attardi & Schatz, 1988), and probably accounts for the differences in mRNA levels between genes. In mouse (Piko & Taylor, 1987) and sea urchin (Cabrera *et al*, 1983) mtDNA, the differences in mitochondrial protein levels between different genes have also been proposed as being due to the differential stability of the mRNAs, since transcription of the genes occurs at the same rate. Transcriptional attenuation only appears to be used in vertebrate mtDNAs to generate the higher levels of rRNA compared to mRNA (though RNA stability also has an effect).

In *S. purpuratus* mtDNA certain adjacent genes overlap (Elliott & Jacobs, 1989), so the mature transcripts cannot be produced from the same RNA precusor molecule (*e.g.* 16S rRNA and COI). Such a mode of transcription appears wasteful, in that a certain percentage of the RNA produced is discarded (Elliott, 1990), having no apparent function, so any arguments for the close regulation of mitochondrial transcription on the basis of economy do not appear to be valid. Studies of transcription in bacteria have shown that transcription termination does occur independently from protein factors (Yager & von Hippel, 1987), as was also apparent from the *in vitro* transcription studies of the A6/COIII region of the *S. purpuratus* mtDNA using phage polymerase (described in Chapter 5). This implies that it is possible that no protein factors are involved in the termination of transcription to generate many, if not all, of the mature transcripts in *S. purpuratus* mtDNA, which in turn argues against replication pausing being due to transcriptional interference.

6.3.3 THE CONTROL OF LAGGING-STRAND REPLICATION.

The mapping of both a prominent pause site and a prominent lagging-strand origin to the region of the genome around the A6/COIII gene junction, suggests that replication pause sites could occur at the sites of initiation of lagging-strand DNA synthesis. A block to the movement of the leading-strand replication fork could allow the DNA to adopt the necessary conformation for lagging-strand synthesis to initiate, and could even lead to the dissociation of the DNA polymerase from the leading-strand and its switch to lagging-strand synthesis; a "U-turn" replication mechanism.

Assuming that replication pause sites do represent sites of lagging-strand initiation, the question then arises as to why pause sites are located near to the 3′ ends of protein-coding genes (section 3.6; Figure 3.11), and why so many are present in the sea urchin mitochondrial genome. Their location could be explained as being to ensure that the formation of a lagging-strand origin causes minimum disruption to the transcription of the genome. It is possible that the number of lagging-strand origins reflects an adaptation of sea urchin mtDNA to keep the regions of single-stranded DNA in the mitochondrial genome to a minimum during replication, and so reduce the potential effects of DNA mutation. Alternatively, S. purpuratus mtDNA could represent an intermediate stage between the semi-discontinuous DNA replication mechanisms of nuclear genomic DNA, and the continuous synthesis, "single origin per strand" mechanism of vertebrate mtDNA.

6.4 FUTURE WORK.

As stated in the earlier chapters, further experiments need to be performed to locate more precisely the lagging-strand origins, and the replication pause sites within the genome. The prediction from the model outlined above (section 6.3.3), which proposes that the replication pause sites have a role

in the initiation of lagging-strand synthesis, is that the locations of the pause sites and lagging-strand origins will coincide.

Another important question which needs to be addressed is how closely does the replication mechanism observed in AEEF mtDNA resemble that seen under physiological conditions *in vivo*. As mentioned above (section 6.1.1), the results obtained with AEEFs would have to be compared with those from naturally replicating mtDNA. Both oocyte and pluteus larval mtDNA could be used, which would enable mtDNA replication under 2 different forms of regulation to be studied. Such comparisons would indicate whether the RIs observed were peculiar to AEEF mtDNA, or to a replication control mechanism acting at a particular stage of development, or were common to all 3 replication systems studied.

The control of mtDNA replication during different stages of development, and the role of the D-loop in this process, could also be addressed. A possible mechanism for the control of mtDNA replication in *S. purpuratus* involves the regulation of the termination of D-loop synthesis (section 6.2). This model predicts that D-loops will be present in all cell types, even if no mtDNA replication occurs, and is supported by the detection of D-loops in egg mtDNA (Jacobs *et al*, 1989). However, an alternative mechanism could be used in early embryogenesis, when mtDNA replication is blocked, involving the control of replication at an earlier point in the events giving rise to the initiation of leading-strand DNA synthesis. The prediction from this alternative mechanism is that D-loop forms of mtDNA would be absent (section 6.2).

The ends of the D-strand molecules could be mapped in mtDNA samples from different developmental stages (*e.g.* oocytes, blastula) by RNase protection and/or primer extension (5′ ends only) experiments, using probes covering the non-coding of the genome. Comparisons of the results would reveal any developmental differences, and pretreatment of the mtDNA samples using RNase H could be utilised to detect any

ribonucleotides present in the D-strand molecules, and so determine the site(s) of RNA to DNA transition.

The control of mtDNA replication probably involves the action of protein factors at the non-coding region of the genome, which could control events such as D-loop termination. In P. lividus, footprinting has shown that a DNA-binding protein interacts at a site within the non-coding region (Roberti et al, 1991), and gel retardation experiments carried out in this laboratory (D.T. Segreto, pers. comm.) using mitochondrial protein extracts and the cloned fragments from S. purpuratus non-coding region, have revealed several different binding activities. Further experiments using these protein fractions, and techniques such as footprinting, and Southwestern blotting, should lead to identification of the binding sites, and the proteins involved. The use of protein fractions from different developmental stages may give an indication of possible roles of DNA binding proteins in the cell cycle, or developmental control of sea urchin mtDNA replication.

Any more accurate assessment of the role of any such proteins in mtDNA replication, and the sequence requirements for DNA synthesis, would need the development of a full in vitro replication system, or at least a system for the initiation of leading-strand DNA synthesis. Then mutagenesis, or the removal of stretches of sequence, such as the G-string, could be used to investigate the role of the various different sequence elements in DNA synthesis. The use of protein extracts from different developmental stages could provide information about the different factors which control replication during development. An in vitro initiation system for lagging-strand synthesis may be harder to establish if DNA-binding proteins, and possibly leading-strand DNA synthesis, are required for the regions of DNA involved to adopt the correct structures for lagging-strand initiation to occur. The use of artificial single-stranded templates from M13 or phagemid recombinants may overcome such difficulties.

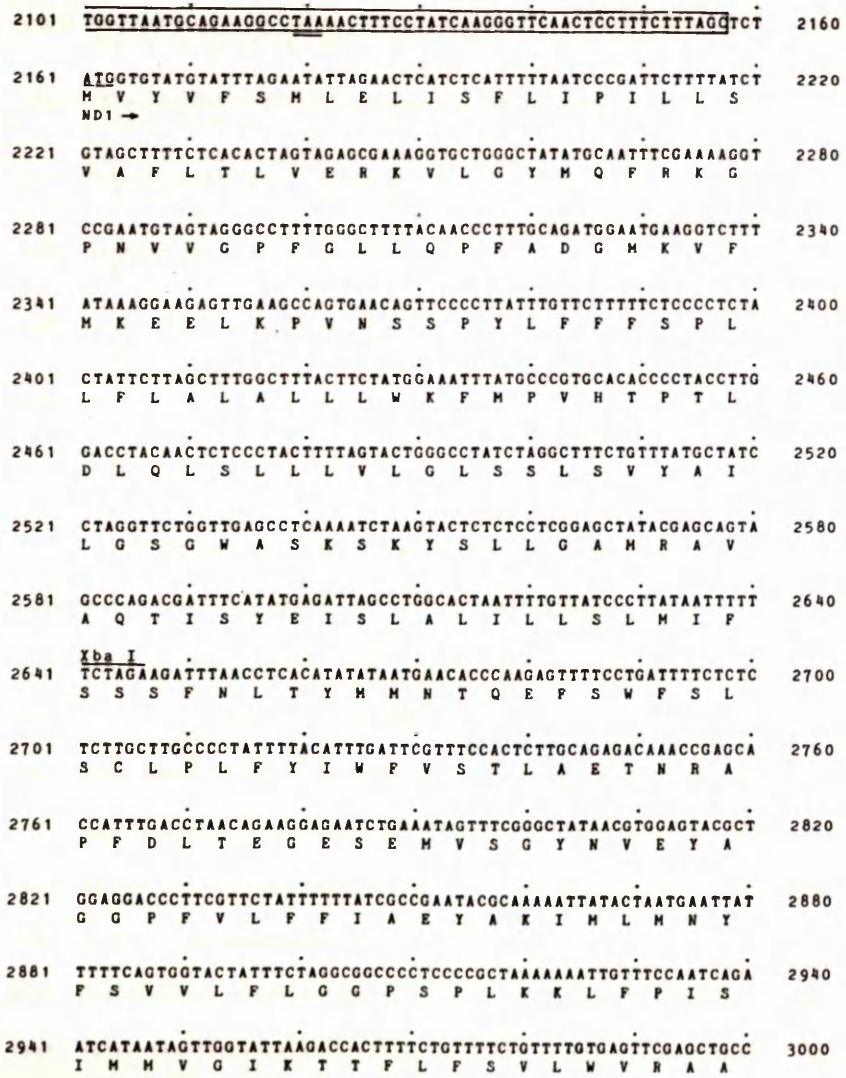
Further investigation of the question of a link between mitochondrial DNA replication and transcription requires the development of an *in vitro* transcription system using mitochondrial RNA polymerase. It could then be determined whether the termination sites observed in the A6/COIII region with phage polymerase can be recognised by the mitochondrial enzyme, and whether the proteins identified as binding to that region of the genome can affect transcription. Other DNA-binding proteins identified as having a role in mtDNA replication (*e.g.* in replication pausing; lagging-strand origin formation) could also be tested for possible effects on transcription.

So in conclusion, the experiments carried out in this study provide some basic information about the mechanism of mtDNA replication in *S. purpuratus*, opening the way for investigations into the mechanisms underlying the developmental regulation of the replication process.

APPENDIX

APPENDIX

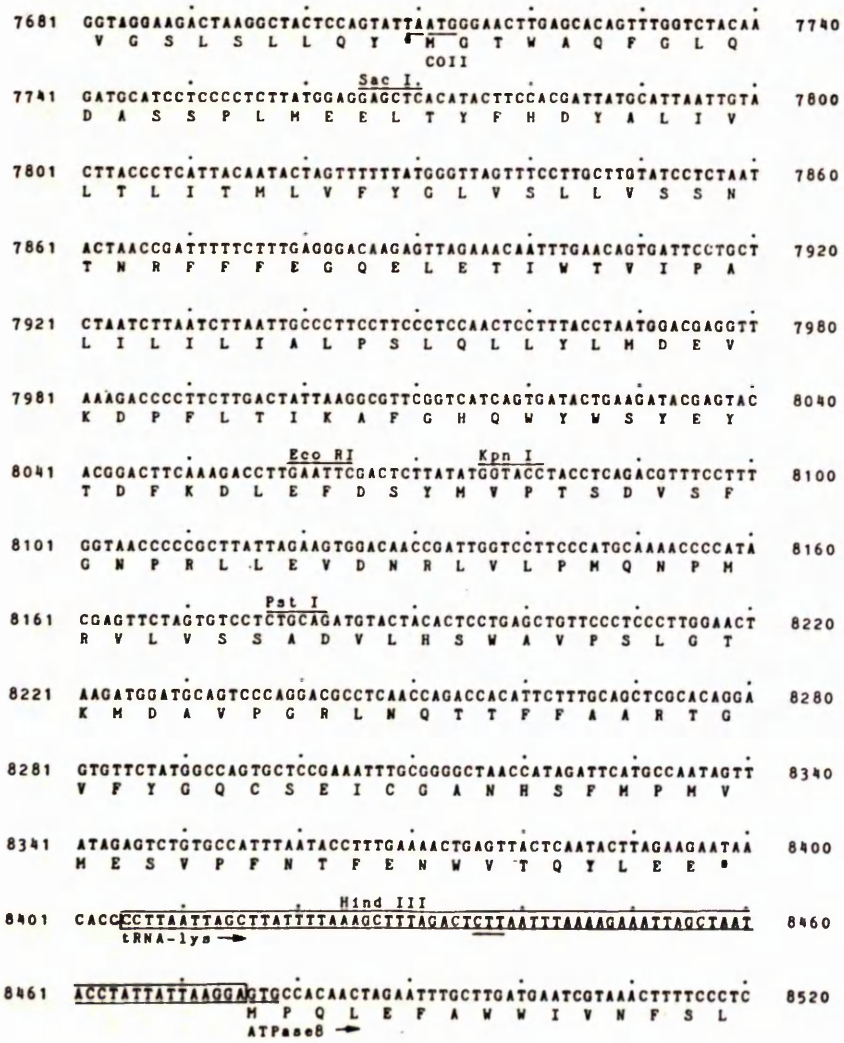
The complete nucleotide sequence of the mitochondrial genome of the sea urchin *Strongylocentrotus purpuratus*, from Jacobs *et al*, 1988, is given below.

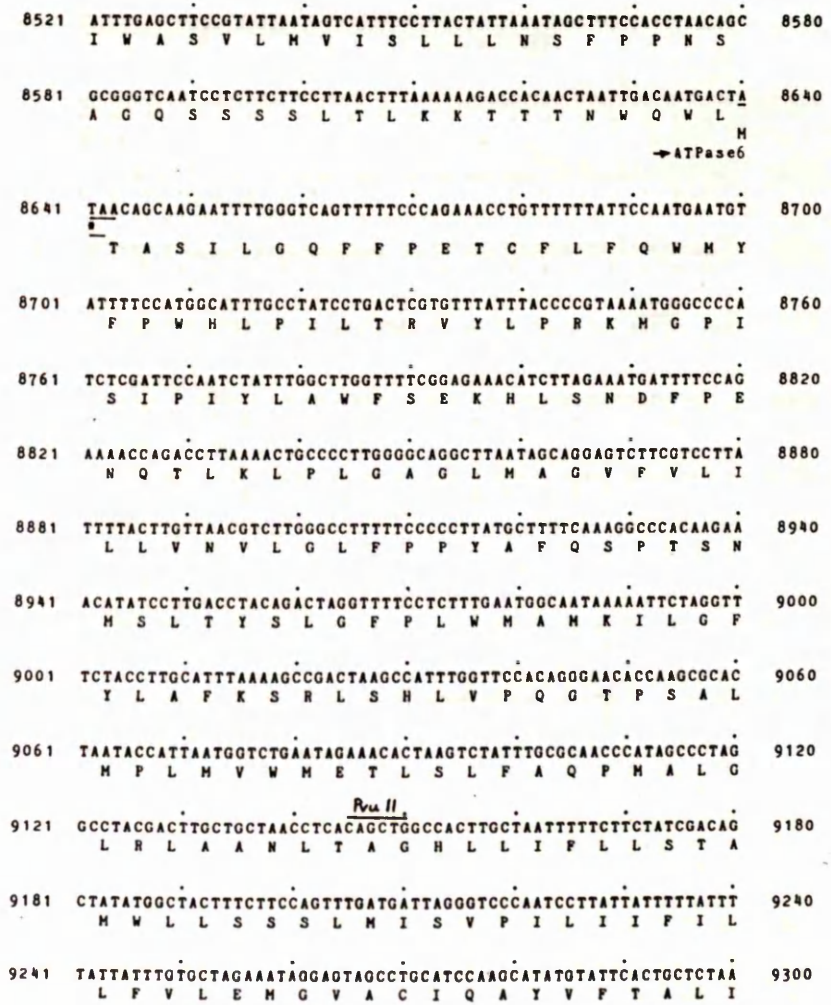

2101 TGGTTAATGCAGAAGGCCTAAAACTTTCCTATCAAGGGTTCAACTCCTTTCTTTAGCTCT 2160
2161 ATGGTGTATGTATTTAGAATATTAGAACTCATCTCATTTTTAATCCCGATTCTTTTATCT 2220
M V Y V F S M L E L I S F L I P I L L S
ND1 →
2221 GTAGCTTTTCTCACACTAGTAGAGCGAAAGGTGCTGGGCTATATGCAATTTCGAAAAGGT 2280
V A F L T L V E R K V L G Y M Q F R K G
2281 CCGAATGTAGTAGGGCCTTTTGGGCTTTTACAACCCTTTGCAGATGGAATGAAGGTCTTT 2340
P N V V G P F G L L Q P F A D G M K V F
2341 ATAAAGGAAGAGTTGAAGCCAGTGAACAGTTCCCCTTATTTGTTCTTTTTCTCCCCTCTA 2400
M K E E L K P V N S S P Y L F F F S P L
2401 CTATTCTTAGCTTTGGCTTTACTTCTATGGAAATTTATGCCCGTGCACACCCCTACCTTG 2460
L F L A L A L L L W K F M P V H T P T L
2461 GACCTACAACTCTCCCTACTTTTAGTACTGGGCCTATCTAGGCTTTCTGTTTATGCTATC 2520
D L Q L S L L L V L G L S S L S V Y A I
2521 CTAGGTTCTGGTTGAGCCTCAAAATCTAAGTACTCTCTCCTCGGAGCTATACGAGCAGTA 2580
L G S G W A S K S K Y S L L G A M R A V
2581 GCCCAGACGATTTCATATGAGATTAGCCTGGCACTAATTTTGTTATCCCTTATAATTTTT 2640
A Q T I S Y E I S L A L I L L S L M I F
Xba I
2641 TCTAGAAGATTTAACCTCACATATATAATGAACACCCAAGAGTTTTCCTGATTTTCTCTC 2700
S S S F N L T Y M M N T Q E F S W F S L
2701 TCTTGCTTGCCCCTATTTTACATTTGATTCGTTTCCACTCTTGCAGAGACAAACCGAGCA 2760
S C L P L F Y I W F V S T L A E T N R A
2761 CCATTTGACCTAACAGAAGGAGAATCTGAAATAGTTTCGGGCTATAACGTGGAGTACGCT 2820
P F D L T E G E S E M V S G Y N V E Y A
2821 GGAGGACCCTTCGTTCTATTTTTTATCGCCGAATACGCAAAAATTATACTAATGAATTAT 2880
G G P F V L F F I A E Y A K I M L M N Y
2881 TTTTCAGTGGTACTATTTCTAGGCGGCCCCTCCCCGCTAAAAAAATTGTTTCCAATCAGA 2940
F S V V L F L G G P S P L K K L F P I S
2941 ATCATAATAGTTGGTATTAAGACCACTTTTCTGTTTTCTGTTTTGTGAGTTCGAGCTGCC 3000
I M M V G I K T T F L F S V L W V R A A
3001 TACCCACGATTCCGGTATGATCAGTTAATGTTCTTAACATGAAAGAGGTACCTCCCCCTT 3060
Kpn I
Y P R F R Y D Q L M F L T W K S Y L P L
3061 TCAATAGGGGCTTTGTGTGCTATTTTGGCTTTAGTGGGCCTTACTGGGAATATCCCTCCC 3120
S M G A L C A I L A L V G L T G N I P P
3121 ACTATTTTAATGAGCTTGCCCCTAAAGTTAAGGTGTCTAGCTGATAATTAGATTTAAGGG 3180
T I * tRNA-Ile →
3181 TTAAATTCCCTTCAAGCTCTATGCGCCAAATAGTGTCTACCTTTTTATTTGTAACTGTAG 3240
M R Q M V S T F L F V T V V
ND2 →
3241 TCTCTGGGACTATAATCGTTGTTTCTTCCGAAAAATGATTCATTATCTGAGTTGGCTTAG 3300
S G T M I V V S S E K W F I I W V G L E
3301 AGCTTAGAACCTTAGCATTGGTTCCAATCCTTTGCTCAGGATTCTCCCCACGAAAAGTGG 3360
L S T L A L V P I L C S G F S P R K V E
Hind III
3361 AGGCGGACAATAAGTACTTTCTTGTTCAAGCTTCAAGGGCTGCACTTTTGCTAAAAGGCG 3420
A D N K Y F L V Q A S S A A L L L K G A
Hind III
3421 CCCTTGGACAAGCTTGGTTGACGGGATCATGATCAATCCTAGATCCTGTTAAAGAGGTAA 3480
L G Q A W L T G S W S I L D P V K E V T
3481 CCTCCATTTGCCTAAGAATAGCCCTTGCATTTAAGATAGGCCTTGCCCCTGTCCACTTCT 3540
S I C L S M A L A F K M G L A P V H F W
3541 GATTTCCAGATGTTTTACAAGGGTTGCCCTTTTTCCAAGGGCTAATAATAGCCACTTGGC 3600
F P D V L Q G L P F F Q G L M M A T W Q
3601 AAAAGATAGCCCCTCTAATACTAATGTTTTATTTTAGCCAGTTAGGTTTTTCTTACCTAC 3660
K M A P L M L M F Y F S Q L G F S Y L L
3661 TTATAACACCTAGGTTAATTTCTGTGCTGATAGGGGGCTGGGGAGGGCTAAATCAGACGC 3720
M T P S L I S V L M G G W G G L N Q T Q
3721 AAGTGCGTAAGATTTTAGCATTCTCTTCAATAGGAAAAATGGGCTGATTAGTCATAACAT 3780
V R K I L A F S S M G K M G W L V M T S
3781 CAGCTTACTCCTTTAACGCTGCGATCATTATGTTAGTTATTTACTTAATTATTAACACTT 3840
A Y S F N A A I I M L V I Y L I I N T S

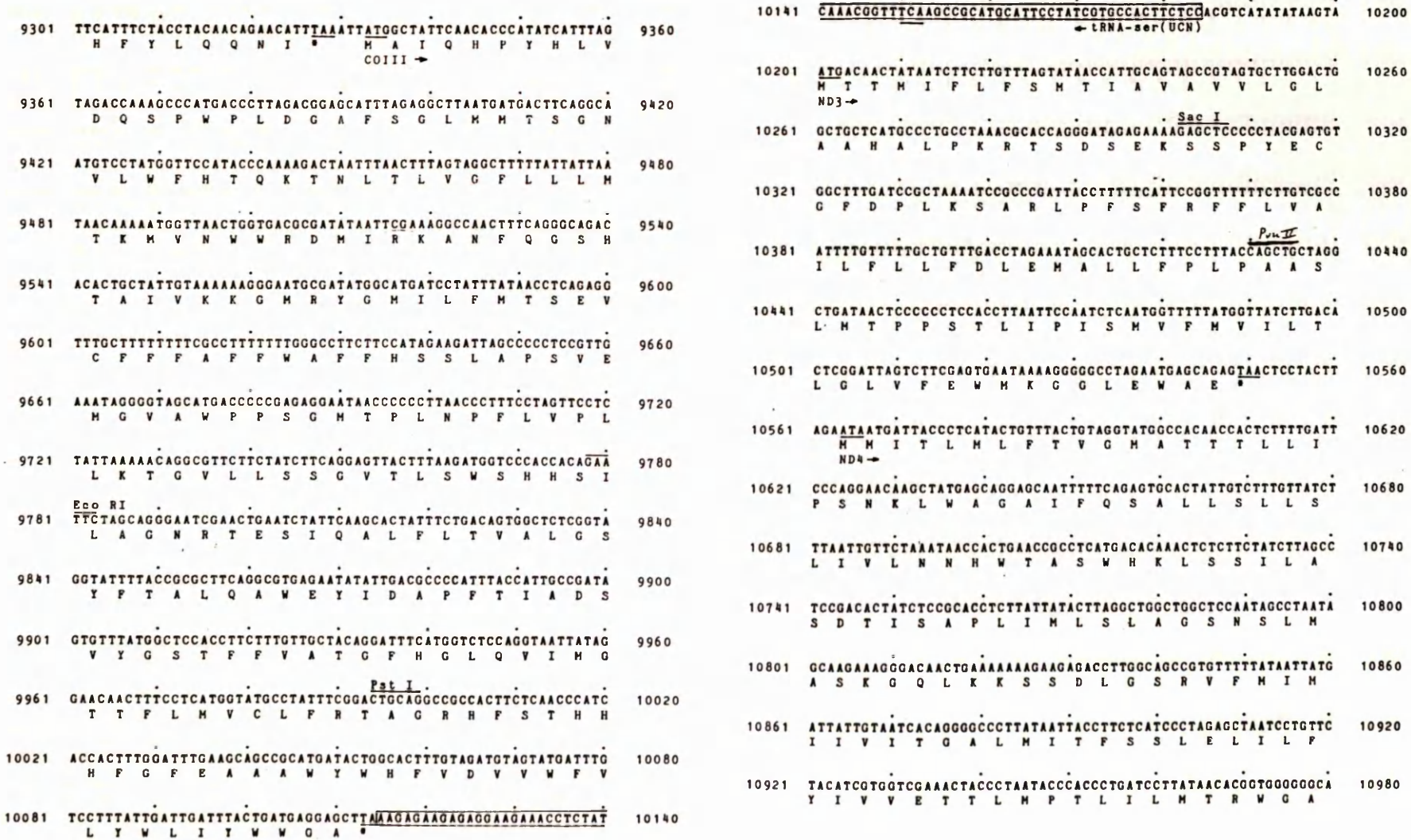
```
3841  CTTTGTTTTTATTGTTTGACCACTTAAAGGTCTCCACATTGGGACACTTAAAAACTATTT  3900
        L  F  L  L  F  D  H  L  K  V  S  T  L  G  H  L  K  T  I  S

3901  CTCAGCTTTCACCAATTAGAGTTGCTCTTGTTCTCCTAGTGATGCTCTCTCTAGGAGGCC  3960
        Q  L  S  P  I  S  V  A  L  V  L  L  V  M  L  S  L  G  G  L

3961  TTCCCCCATTAACCGGGTTTATCCTAAAGTTTACCTCCCTTTATTTCTTGGTTGCCAAAA  4020
        P  P  L  T  G  F  I  L  K  F  T  S  L  Y  F  L  V  A  K  N

4021  ATTTTATCATTTTATCTTCTATTATGATAATTGGAAATCTTCAAGATTATTTTTTTTATC  4080
        F  I  I  L  S  S  I  M  M  I  G  N  L  Q  D  Y  F  F  Y  L

4081  TCCGAATTTCGTTTAAAACTAGCTTATTTCTGTTTCCCCAACACATTATTAGATCCGCCT  4140
        R  I  S  F  K  T  S  L  F  L  F  P  Q  H  I  I  S  S  A  S

4141  CATGGCGAAATAGGACAATAATTTCACCTCTCGCCCCAAAGGCATGATTAAGTTCGGTCT  4200
        W  R  N  S  T  M  I  S  P  L  A  P  K  A  W  L  S  S  V  S

4201  CCACTGTGTTGAGTACTCTTGCAATACCCCTTACCCTCCCCTTATATATAATTACATAGA  4260
        T  V  L  S  T  L  A  M  P  L  T  L  P  L  Y  M  I  T  *

4261  AAAGTTATGACTCTAGGTGCATAGAAAATCCAAACAATAAATTTCTATAAAATAGAAAAC  4320
      16S rRNA

4321  ACTCTTACTCTCTAGTAAATTCATTTGAAATCTTATTTTAACCAAAAGAAGCAGTACCGC  4380

4381  AAGGGAAAGATGAAATACCCATAATTAAACAAACCTAAAAAAGGAAAGACTAAACCTGTG  4440

4441  ACCTGTGTATAATGGATTAACGAGAAATATAAGAAAACTAGTCTAATCCCGAAACTGGGC  4500

4501  GAGCTAATCTTCCCTCCTTTTTAGAAGGATACCCCCACTGTTGCAATAGTGGAAAAAAAG  4560

4561  GGAAGATTAGATGTGAAATCTAACCGCGCCCAGAGATAGCTGGTTTCCCAAAAAATTAGT  4620

4621  TTGAGCTAAGCCTCTATAAGAAAAATAAAACTCCTTTAATATTATAAAAAGAATCTTTTA  4680

                                  Hind III.
4681  ATTTTTAAGCAGAGGTTAGGCCCTTAAGGATAAGCTTAAAGCCCACAATGGAAAAATCCA  4740

4741  ACATTGATAGTTAAAGACAACAAGCCCCAAAATAGCTATTCTCGAAGGAATAGGCCTAGA  4800

4801  AGCAGCCACTTAACAAGAAAGCGTTAAAGCTCAATTGTTCTTGCTAGCTAAAAATTTTTG  4860

4861  GGCAATCATCTCTAACACTACAAATTATTGGGACATTCTGTAATCAGAAGAGACAATGTT  4920

4921  AATATAAGTAAGATAACTATACCTAGCGTTTTATACGCTTTTAACCATGGAAGAAAAACA  4980

4981  TTGAAACTAAAAATCCTCTTTTTAAAGTCGTCTTCCAACTCAGGAGAAGAAAATAAAAA  5040

5041  AAGTGGGAAAAGGAACTCGGCAAATAAGGTTTCGCCTGTTTACCAAAAACATCGCTCCCC  5100

5101  AAAATTTAAAGCCTGGGGAGTCCTGCCTGCCCAGTGACTAGAGGTTAAACGGCCGCTGTA  5160

5161  TCTTGACCGTCGGAGGTAGCATAATCATTGTCTTCCTAATTAGAGACTAGTATGAATGGC  5220

5221  AAGAGGGAAATAAATGACATTTTTTCTATAGCCCTTAATACTACCTCCCCGTGAAGAGGC  5280

5281  GGGGATAAAACGGTTAGACGAGAAGACCCTGTGGAGCTTTTAAGCGGAAGTTAAATTTTA  5340

5341  ACACACTTACCTTGTGACTAACTAATATACCTATCCAACAGTAAGTTTTAAACATTTTAG  5400

          Hind III
5401  CAAAAGCTTTGGTTGGGGCAACGCGGAGTAAGAAGACCCTCCGCTAATAAGATATTACTA  5460

5461  TAAAAAGAATTACGGTTCTACAATCAAAATGAAAGAATGATCCACTAAGGTGATCAAAGA  5520

5521  AACAAGTTACCGCAGGATAACAGCGTTATCTTTTCTGAGAGTTCACATTGACGAAAAGGT  5580

                                                  Hind III.
5581  TTGCGACCTCGATGTTGGATCGGGACATCCTAAGGGTGCAGAAGCTTTTAAGGGTTGGTC  5640

5641  TGTTCGACCATTAAAGTCCTACGTGATCTGAGTTCAGACCGGCGAGAGCCAGGTCAGTTT  5700

5701  CTATCTACGTTAAGGTCTCTCTTAGTACGAAAGGACCAGAGAGAGAGTGTCTATGCAAAG  5760

5761  AACGTAAGCACTTCAATTAAAAACATGCAACTAAGACGATGATTATTTTCTACTAACCAC  5820
                                  M  Q  L  S  R  W  L  F  S  T  N  H
                                  COI →

5821  AAGGACATCGGAACACTTTATTTAATTTTTGGGGCCTGAGCTGGCATGGTAGGCACAGCT  5880
        K  D  I  G  T  L  Y  L  I  F  G  A  W  A  G  M  V  G  T  A

5881  ATGAGTGTGATTATCCGTGCCGAGTTGGCACAACCTGGTTCGCTGCTAAAAGATGACCAG  5940
        M  S  V  I  I  R  A  E  L  A  Q  P  G  S  L  L  K  D  D  Q
```

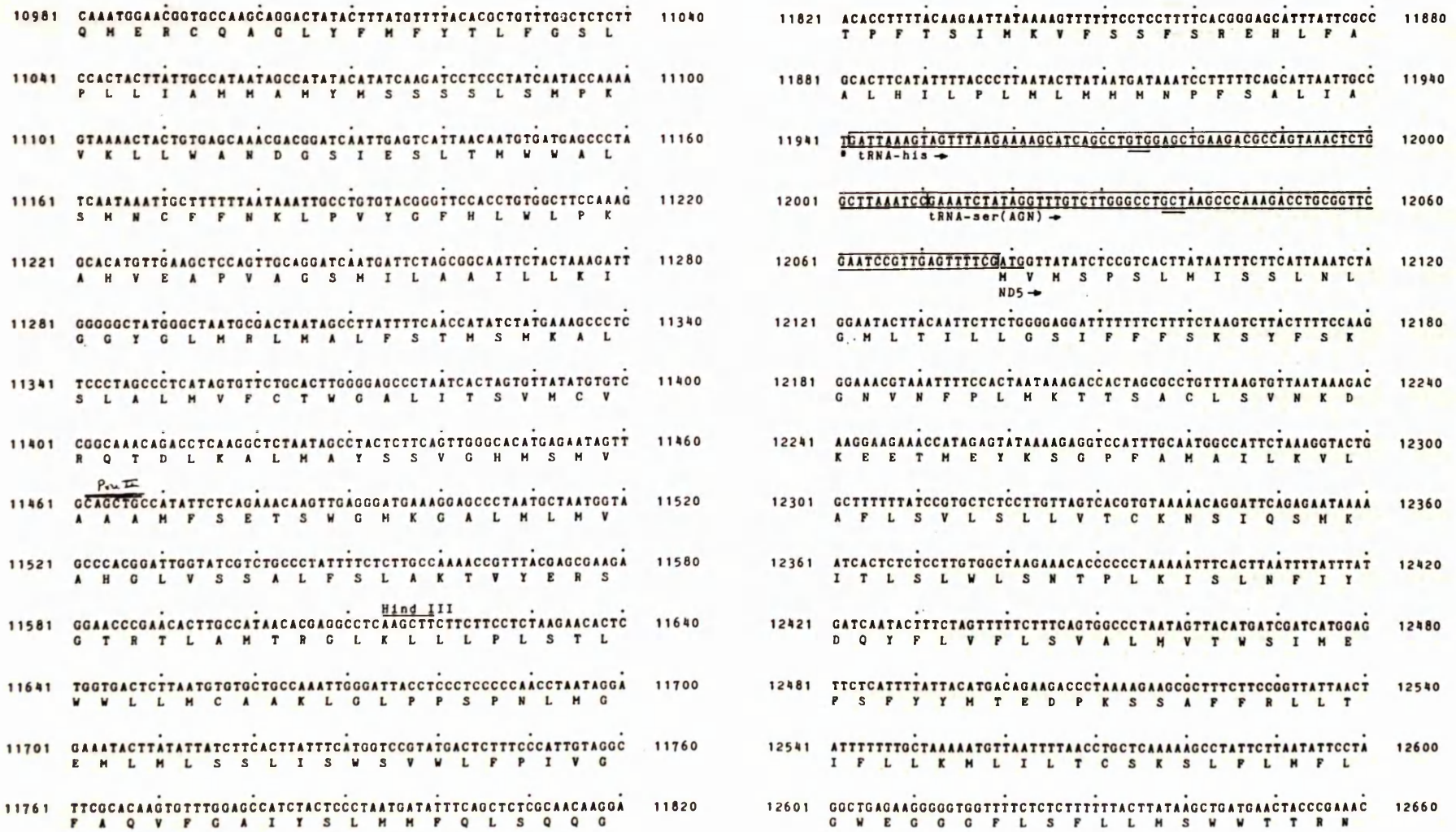
```
5941  ATATACAAAGTGGTCGTTACCGCACAATCGCTAGTCATGATTTTCTTCATGGTAATGCCA  6000
      M  Y  K  V  V  V  T  A  Q  S  L  V  M  I  F  F  M  V  M  P

6001  ATAATGATTGGTGGATTTGGGAATTGACTCATTCCACTAATGATCGGTGCGCCAGATATG  6060
      M  M  I  G  G  F  G  N  W  L  I  P  L  M  I  G  A  P  D  M

6061  GCCTTCCCCCCGCATGAAAAATATGAGTTTTTGACTTATTCCCCCTTCTTTTATATTACTT  6120
      A  F  P  R  M  K  N  M  S  F  W  L  I  P  P  S  F  M  L  L

6121  TTAGCGTCCGCAGGAGTAGAAAAAGGAGCAGGAACTGGCTGAACTATCTACCCTCCTCTC  6180
      L  A  S  A  G  V  E  K  G  A  G  T  G  W  T  I  Y  P  P  L

6181  TCTAGTAAAATAACACACGCCGGTAGGTCCGTTGATTTAGCAATCTTCTCCCTTCACCTG  6240
      S  S  K  M  T  H  A  G  S  S  V  D  L  A  I  F  S  L  H  L

6241  GCCGGTGCCTCTTCCATCTTGGGCCTCATTAAATTTATAACAACAATTATTAATATGCGG  6300
      A  G  A  S  S  I  L  G  L  I  K  F  M  T  T  I  I  N  M  R

6301  ACACCGGGGATGTCTTTGGATCGTCTTCCTTTATTCGTCTGATCCGTCTTTGTCACTGCC  6360
      T  P  G  M  S  L  D  R  L  P  L  F  V  W  S  V  F  V  T  A

6361  TTCTTGCTCCTCCTTTCTCTTCCAGTATTAGCAGGAGCAATTACAATGCTTCTCACAGAT  6420
      F  L  L  L  L  S  L  P  V  L  A  G  A  I  T  M  L  L  T  D

                  Pst I
6421  CGTAAAATAAACACAACTTTCTTTGACCCTGCAGGAGGGGGAGATCCAATTCTATTTCAA  6480
      R  K  M  N  T  T  F  F  D  P  A  G  G  G  D  P  I  L  F  Q

6481  CACCTATTCTGGCTTTTTGGACACCCCGAGGTGTATATTCTTATCTTACCGGGATTTGGT  6540
      H  L  F  W  L  F  G  H  P  E  V  Y  I  L  I  L  P  G  F  G

6541  ATGATCTCACACGTTATAGCTCACTACTCTGGTAAGCGAGAGCCTTTCGGATACCTGGGG  6600
      M  I  S  H  V  M  A  H  Y  S  G  K  R  E  P  F  G  Y  L  G

6601  TTGGTTTATGCCATGATTGCAATAGGAGTTTTAGGATTCCTTGTCTGGGCCCACCATATG  6660
      L  V  Y  A  M  I  A  M  G  V  L  G  F  L  V  W  A  H  H  M

6661  TTTACAGTAGGGATGGATGTTGATACACGAGCATACTTCACTGCCGCCACAATGATTATT  6720
      F  T  V  G  M  D  V  D  T  R  A  Y  F  T  A  A  T  M  I  I

6721  GCTGTCCCAACAGGATTAAAGGTTTTCAGATGAATGGCAAAGCTCCAAGGGTCTAATCTA  6780
      A  V  P  T  G  L  K  V  F  S  W  M  A  K  L  Q  G  S  N  L

6781  CAATGAAGACTCCCTTTATTATGAACCTTGGGGATTGTATTTTTATTCACATTAGGAGGA  6840
      Q  W  S  L  P  L  L  W  T  L  G  I  V  F  L  F  T  L  G  G
```

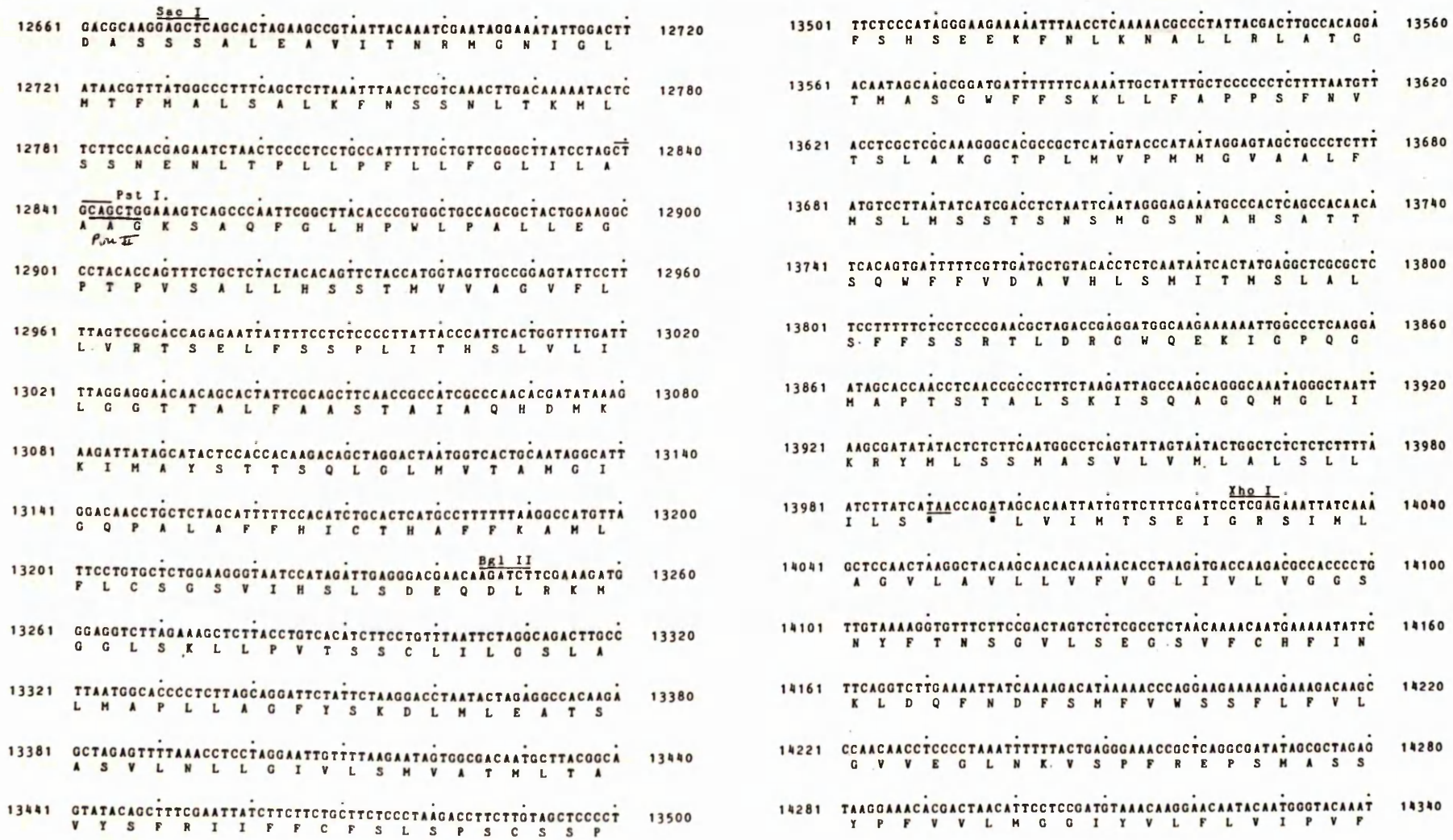
```
6841  CTCACAGGTATTGTTCTTGCCAATTCCTCCATTGACTTTGTTCTTCATGATACCTACTAC  6900
      L  T  G  I  V  L  A  N  S  S  I  D  F  V  L  H  D  T  Y  Y

6901  GTGGTAGCTCACTTTCACTACGTTCTTTCAATGGGGGCTGTATTTGCAATCTTCGCTGGT  6960
      V  V  A  H  F  H  Y  V  L  S  M  G  A  V  F  A  I  F  A  G

6961  TTCACTCACTGGTTTCCCCTCTTCTCTGGTTATAGCCTACACCCATTATGAGGAAAGGTT  7020
      F  T  H  W  F  P  L  F  S  G  Y  S  L  H  P  L  W  G  K  V

7021  CACTTCTTCATAATGTTTGTTGGAGTCAACTTAACCTTTTTCCCTCAACACTTCTTAGGT  7080
      H  F  F  M  M  F  V  G  V  N  L  T  F  F  P  Q  H  F  L  G

7081  CTAGCCGGAATGCCACGACGGTACTCAGACTATCCAGACGCCTATACACTTTGAAATACT  7140
      L  A  G  M  P  R  R  Y  S  D  Y  P  D  A  Y  T  L  W  N  T

7141  ATCTCCTCAATTGGATCAACCATCTCCGTAGTGGCTATGCTATTTTTCCTCTTCTTAATC  7200
      I  S  S  I  G  S  T  I  S  V  V  A  M  L  F  F  L  F  L  I

7201  TGGGAGGCCTTCGCTTCTCAACGGGAAGGAATCACCCCAGAGTTCTCACACGCCTCACTA  7260
      W  E  A  F  A  S  Q  R  E  G  I  T  P  E  F  S  H  A  S  L

7261  GAGTGACAATACACCTCCTTTCCCCCTTCTCACCACACCTTCGATGAAACACCCTCTACC  7320
      E  W  Q  Y  T  S  F  P  P  S  H  H  T  F  D  E  T  P  S  T

7321  ATAATTATTGTAAAGTAAGTGTCAGTTAAAGAGTTAGTTTAAGGAGAACCTCTGATTTCG  7380
      H  I  I  V  K  *                    tRNA-arg →

7381  GCTCAGATGGTTTTGGTTCAACCCCAAAACTCTTGAAATCGCCCTGCTTATAGTTATTTT  7440
                                           M  A  L  L  M  V  I  L
                                           ND4L →

7441  ATCAATGTTTTACCTAGGACTTATGGGAATCCTTTTAAAACGACTCCATTTCCTATCCAT  7500
      S  M  F  Y  L  G  L  M  G  I  L  L  K  R  L  H  F  L  S  I

7501  TCTATTATGCCTTGAGCTTACTCCTTATTTCCTTGTTTATCGGAATTGCGTTTGAAAAAA  7560
      L  L  C  L  E  L  T  P  Y  F  L  V  Y  R  N  C  V  W  K  K

7561  AAAAACAGGGGTCCCTCAAAAAACAACATTCAACTTATTCGTTTTGACTCTAGTCGCTTG  7620
      K  T  G  V  P  Q  K  T  T  F  N  L  F  V  L  T  L  V  A  C

7621  CGAAGCAAGCATAGGCCTCTCCCTAATGGTAGGGCTCTCCCGTACACACTCTTCAAATTT  7680
      E  A  S  M  G  L  S  L  M  V  G  L  S  R  T  H  S  S  N  L
```

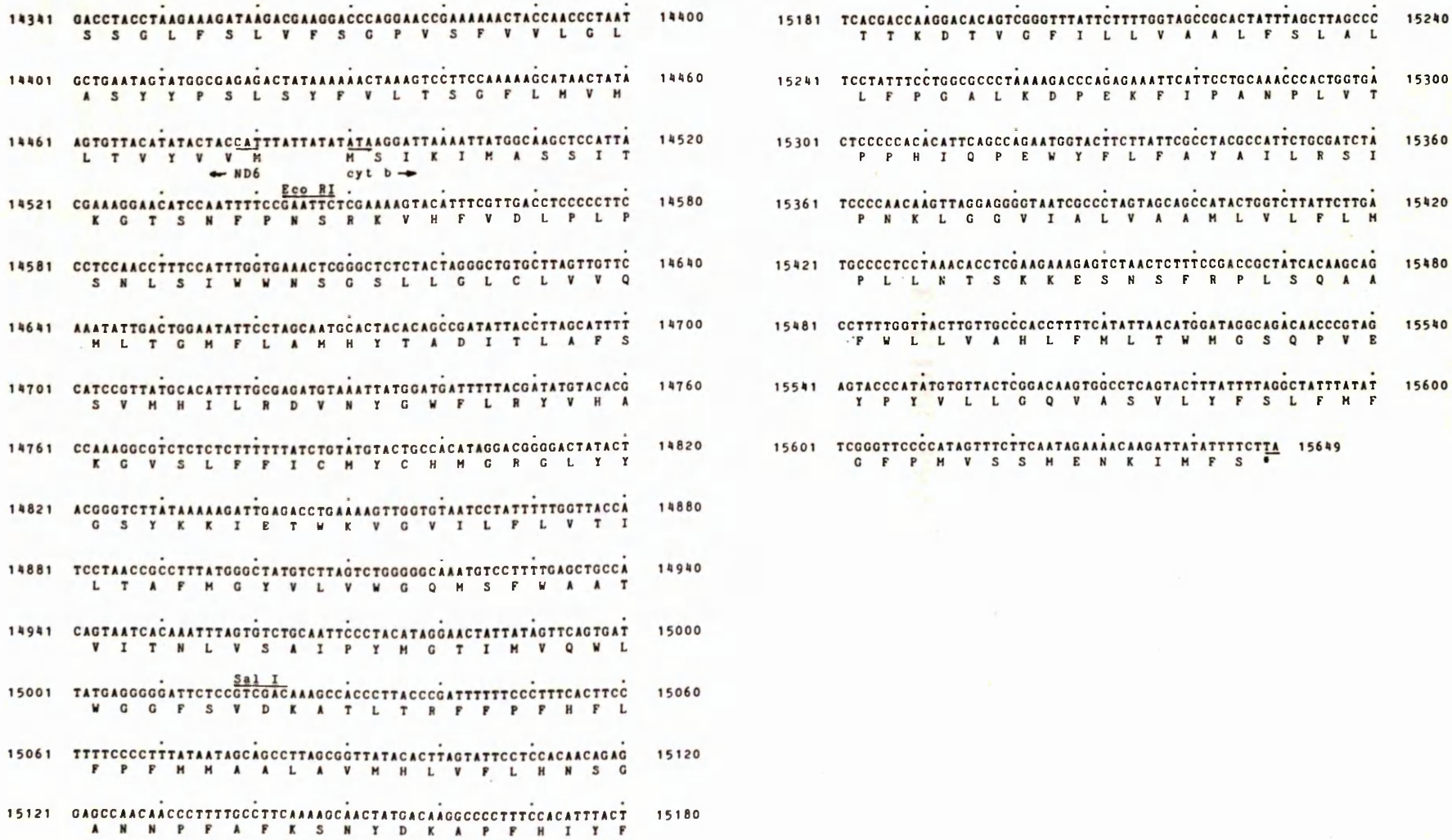

Sac I.
COII
Eco RI
Kpn I
Pst I
Hind III
tRNA-lys
ATPase8


ATPase6
Pvu II


9301 TTCATTTCTACCTACAACAGAACATTTAAATTATGGCTATTCAACACCCATATCATTTAG 9360
H F Y L Q Q N I * M A I Q H P Y H L V
COIII →
9361 TAGACCAAAGCCCATGACCCTTAGACGGAGCATTTAGAGGCTTAATGATGACTTCAGGCA 9420
D Q S P W P L D G A F S G L M M T S G N
9421 ATGTCCTATGGTTCCATACCCAAAAGACTAATTTAACTTTAGTAGGCTTTTTATTATTAA 9480
V L W F H T Q K T N L T L V G F L L L H
9481 TAACAAAAATGGTTAACTGGTGACGCGATATAATTCGAAAGGCCAACTTTCAGGGCAGAC 9540
T K M V N W W R D M I R K A N F Q G S H
9541 ACACTGCTATTGTAAAAAAGGGAATGCGATATGGCATGATCCTATTTATAACCTCAGAGG 9600
T A I V K K G M R Y G M I L F M T S E V
9601 TTTGCTTTTTTTTCGCCTTTTTTTGGGCCTTCTTCCATAGAAGATTAGCCCCCTCCGTTG 9660
C F F F A F F W A F F H S S L A P S V E
9661 AAATAGGGGTAGCATGACCCCCGAGAGGAATAACCCCCCTTAACCCTTTCCTAGTTCCTC 9720
M G V A W P P S G M T P L N P F L V P L
9721 TATTAAAAACAGGCGTTCTTCTATCTTCAGGAGTTACTTTAAGATGGTCCCACCACAGAA 9780
L K T G V L L S S G V T L S W S H H S I
Eco RI
9781 TTCTAGCAGGGAATCGAACTGAATCTATTCAAGCACTATTTCTGACAGTGGCTCTCGGTA 9840
L A G N R T E S I Q A L F L T V A L G S
9841 GGTATTTTACCGCGCTTCAGGCGTGAGAATATATTGACGCCCCATTTACCATTGCCGATA 9900
Y F T A L Q A W E Y I D A P F T I A D S
9901 GTGTTTATGGCTCCACCTTCTTTGTTGCTACAGGATTTCATGGTCTCCAGGTAATTATAG 9960
V Y G S T F F V A T G F H G L Q V I M G
Pst I
9961 GAACAACTTTCCTCATGGTATGCCTATTTCGGACTGCAGGCCGCCACTTCTCAACCCATC 10020
T T F L M V C L F R T A G R H F S T H H
10021 ACCACTTTGGATTTGAAGCAGCCGCATGATACTGGCACTTTGTAGATGTAGTATGATTTG 10080
H F G F E A A A W Y W H F V D V V W F V
10081 TCCTTTATTGATTGATTTACTGATGAGGAGCTTAAAGAGAAGAGAGGAAGAAACCTCTAT 10140
L Y W L I Y W W G A *
10141 CAAACGGTTTCAAGCCGCATGCATTCCTATCGTGCCACTTCTCCACGTCATATATAAGTA 10200
← tRNA-ser(UCN)
10201 ATGACAACTATAATCTTCTTGTTTAGTATAACCATTGCAGTAGCCGTAGTGCTTGGACTG 10260
M T T M I F L F S M T I A V A V V L G L
ND3 →
Sac I
10261 GCTGCTCATGCCCTGCCTAAACGCACCAGGGATAGAGAAAAGAGCTCCCCCTACGAGTGT 10320
A A H A L P K R T S D S E K S S P Y E C
10321 GGCTTTGATCCGCTAAAATCCGCCCGATTACCTTTTTCATTCCGGTTTTTTCTTGTCGCC 10380
G F D P L K S A R L P F S F R F F L V A
PvuII
10381 ATTTTGTTTTTGCTGTTTGACCTAGAAATAGCACTGCTCTTTCCTTTACCAGCTGCTAGG 10440
I L F L L F D L E M A L L F P L P A A S
10441 CTGATAACTCCCCCCTCCACCTTAATTCCAATCTCAATGGTTTTTATGGTTATCTTGACA 10500
L M T P P S T L I P I S M V F M V I L T
10501 CTCGGATTAGTCTTCGAGTGAATAAAAGGGGGCCTAGAATGAGCAGAGTAACTCCTACTT 10560
L G L V F E W M K G G L E W A E *
10561 AGAATAATGATTACCCTCATACTGTTTACTGTAGGTATGGCCACAACCACTCTTTTGATT 10620
M M I T L M L F T V G M A T T T L L I
ND4 →
10621 CCCAGGAACAAGCTATGAGCAGGAGCAATTTTTCAGAGTGCACTATTGTCTTTGTTATCT 10680
P S N K L W A G A I F Q S A L L S L L S
10681 TTAATTGTTCTAAATAACCACTGAACCGCCTCATGACACAAACTCTCTTCTATCTTAGCC 10740
L I V L N N H W T A S W H K L S S I L A
10741 TCCGACACTATCTCCGCACCTCTTATTATACTTAGGCTGGCTGGCTCCAATAGCCTAATA 10800
S D T I S A P L I M L S L A G S N S L M
10801 GCAAGAAAGGGACAACTGAAAAAAAGAAGAGACCTTGGCAGCCGTGTTTTTATAATTATG 10860
A S K G Q L K K S S D L G S R V F M I M
10861 ATTATTGTAATCACAGGGGCCCTTATAATTACCTTCTCATCCCTAGAGCTAATCCTGTTC 10920
I I V I T G A L M I T F S S L E L I L F
10921 TACATCGTGGTCGAAACTACCCTAATACCCACCCTGATCCTTATAACACGGTGGGGGGCA 10980
Y I V V E T T L M P T L I L M T R W G A


10981 CAAATGGAACGGTGCCAAGCAGGACTATACTTTATGTTTTACACGCTGTTTGGCTCTCTT 11040
Q M E R C Q A G L Y F M F Y T L F G S L
11041 CCACTACTTATTGCCATAATAGCCATATACATATCAAGATCCTCCCTATCAATACCAAAA 11100
P L L I A M M A M Y M S S S S L S M P K
11101 GTAAAACTACTGTGAGCAAACGACGGATCAATTGAGTCATTAACAATGTGATGAGCCCTA 11160
V K L L W A N D G S I E S L T M W W A L
11161 TCAATAAATTGCTTTTTTAATAAATTGCCTGTGTACGGGTTCCACCTGTGGCTTCCAAAG 11220
S M N C F F N K L P V Y G F H L W L P K
11221 GCACATGTTGAAGCTCCAGTTGCAGGATCAATGATTCTAGCGGCAATTCTACTAAAGATT 11280
A H V E A P V A G S M I L A A I L L K I
11281 GGGGGCTATGGGCTAATGCGACTAATAGCCTTATTTTCAACCATATCTATGAAAGCCCTC 11340
G G Y G L M R L M A L F S T M S M K A L
11341 TCCCTAGCCCTCATAGTGTTCTGCACTTGGGGAGCCCTAATCACTAGTGTTATATGTGTC 11400
S L A L M V F C T W G A L I T S V M C V
11401 CGGCAAACAGACCTCAAGGCTCTAATAGCCTACTCTTCAGTTGGGCACATGAGAATAGTT 11460
R Q T D L K A L M A Y S S V G H M S M V
PvuII
11461 GCAGCTGCCATATTCTCAGAAACAAGTTGAGGGATGAAAGGAGCCCTAATGCTAATGGTA 11520
A A A M F S E T S W G M K G A L M L M V
11521 GCCCACGGATTGGTATCGTCTGCCCTATTTTCTCTTGCCAAAACCGTTTACGAGCGAAGA 11580
A H G L V S S A L F S L A K T V Y E R S
Hind III
11581 GGAACCCGAACACTTGCCATAACACGAGGCCTCAAGCTTCTTCTTCCTCTAAGAACACTC 11640
G T R T L A M T R G L K L L L P L S T L
11641 TGGTGACTCTTAATGTGTGCTGCCAAATTGGGATTACCTCCCTCCCCCAACCTAATAGGA 11700
W W L L M C A A K L G L P P S P N L M G
11701 GAAATACTTATATTATCTTCACTTATTTCATGGTCCGTATGACTCTTTCCCATTGTAGGC 11760
E M L M L S S L I S W S V W L F P I V G
11761 TTCGCACAAGTGTTTGGAGCCATCTACTCCCTAATGATATTTCAGCTCTCGCAACAAGGA 11820
F A Q V F G A I Y S L M M F Q L S Q Q G
11821 ACACCTTTTACAAGAATTATAAAAGTTTTTTCCTCCTTTTCACGGGAGCATTTATTCGCC 11880
T P F T S I M K V F S S F S R E H L F A
11881 GCACTTCATATTTTACCCTTAATACTTATAATGATAAATCCTTTTTCAGCATTAATTGCC 11940
A L H I L P L M L M M M N P F S A L I A
11941 TGATTAAAGTAGTTTAAGAAAAGCATCAGCCTGTGGAGCTGAAGACGCCAGTAAACTCTG 12000
tRNA-his
12001 GCTTAAATCCGAAATCTATAGGTTTGTCTTGGGCCTGCTAAGCCCAAAGACCTGCGGTTC 12060
tRNA-ser(AGN)
12061 GAATCCGTTGAGTTTTCGATGGTTATATCTCCGTCACTTATAATTTCTTCATTAAATCTA 12120
M V M S P S L M I S S L N L
ND5
12121 GGAATACTTACAATTCTTCTGGGGAGGATTTTTTTCTTTTCTAAGTCTTACTTTTCCAAG 12180
G M L T I L L G S I F F F S K S Y F S K
12181 GGAAACGTAAATTTTCCACTAATAAAGACCACTAGCGCCTGTTTAAGTGTTAATAAAGAC 12240
G N V N F P L M K T T S A C L S V N K D
12241 AAGGAAGAAACCATAGAGTATAAAAGAGGTCCATTTGCAATGGCCATTCTAAAGGTACTG 12300
K E E T M E Y K S G P F A M A I L K V L
12301 GCTTTTTTATCCGTGCTCTCCTTGTTAGTCACGTGTAAAAACAGGATTCAGAGAATAAAA 12360
A F L S V L S L L V T C K N S I Q S M K
12361 ATCACTCTCTCCTTGTGGCTAAGAAACACCCCCCTAAAAATTTCACTTAATTTTATTTAT 12420
I T L S L W L S N T P L K I S L N F I Y
12421 GATCAATACTTTCTAGTTTTTCTTTCAGTGGCCCTAATAGTTACATGATCGATCATGGAG 12480
D Q Y F L V F L S V A L M V T W S I M E
12481 TTCTCATTTTATTACATGACAGAAGACCCTAAAAGAAGCGCTTTCTTCCGGTTATTAACT 12540
F S F Y Y M T E D P K S S A F F R L L T
12541 ATTTTTTTGCTAAAAATGTTAATTTTAACCTGCTCAAAAAGCCTATTCTTAATATTCCTA 12600
I F L L K M L I L T C S K S L F L M F L
12601 GGCTGAGAAGGGGGTGGTTTTCTCTCTTTTTTACTTATAAGCTGATGAACTACCCGAAAC 12660
G W E G G G F L S F L L M S W W T T R N


Sac I
Pst I.
Pvu II
Bgl II
Xho I


14341 GACCTACCTAAGAAAGATAAGACGAAGGACCCAGGAACCGAAAAAACTACCAACCCTAAT 14400
S S G L F S L V F S G P V S F V V L G L
14401 GCTGAATAGTATGGCGAGAGACTATAAAAAACTAAAGTCCTTCCAAAAAGCATAACTATA 14460
A S Y Y P S L S Y F V L T S G F L M V H
14461 AGTGTTACATATACTACCATTTATTATATATAAGGATTAAAATTATGGCAAGCTCCATTA 14520
L T V Y V V M M S I K I M A S S I T
ND6
cyt b
Eco RI
14521 CGAAAGGAACATCCAATTTTCCGAATTCTCGAAAAGTACATTTCGTTGACCTCCCCCTTC 14580
K G T S N F P N S R K V H F V D L P L P
14581 CCTCCAACCTTTCCATTTGGTGAAACTCGGGCTCTCTACTAGGGCTGTGCTTAGTTGTTC 14640
S N L S I W W N S G S L L G L C L V V Q
14641 AAATATTGACTGGAATATTCCTAGCAATGCACTACACAGCCGATATTACCTTAGCATTTT 14700
M L T G M F L A M H Y T A D I T L A F S
14701 CATCCGTTATGCACATTTTGCGAGATGTAAATTATGGATGATTTTTACGATATGTACACG 14760
S V M H H I L R D V N Y G W F L R Y V H A
14761 CCAAAGGCGTCTCTCTCTTTTTTATCTGTATGTACTGCCACATAGGACGGGGACTATACT 14820
K G V S L F F I C M Y C H M G R G L Y Y
14821 ACGGGTCTTATAAAAAGATTGAGACCTGAAAAGTTGGTGTAATCCTATTTTTGGTTACCA 14880
G S Y K K I E T W K V G V I L F L V T I
14881 TCCTAACCGCCTTTATGGGCTATGTCTTAGTCTGGGGGGCAAATGTCCTTTTGAGCTGCCA 14940
L T A F M G Y V L V W G Q M S F W A A T
14941 CAGTAATCACAAATTTAGTGTCTGCAATTCCCTACATAGGAACTATTATAGTTCAGTGAT 15000
V I T N L V S A I P Y M G T I M V Q W L
Sal I
15001 TATGAGGGGGATTCTCCGTCGACAAAGCCACCCTTACCCGATTTTTTCCCTTTCACTTCC 15060
W G G F S V D K A T L T R F F P F H F L
15061 TTTTCCCCTTTATAATAGCAGCCTTAGCGGTTATACACTTAGTATTCCTCCACAACAGAG 15120
F P F M M A A L A V M H L V F L H N S G
15121 GAGCCAACAACCCTTTTGCCTTCAAAAGCAACTATGACAAGGCCCCTTTCCACATTTACT 15180
A N N P F A F K S N Y D K A P F H I Y F
15181 TCACGACCAAGGACACAGTCGGGTTTATTCTTTTGGTAGCCGCACTATTTAGCTTAGCCC 15240
T T K D T V G F I L L V A A L F S L A L
15241 TCCTATTTCCTGGCGCCCTAAAAGACCCAGAGAAATTCATTCCTGCAAACCCACTGGTGA 15300
L F P G A L K D P E K F I P A N P L V T
15301 CTCCCCCACACATTCAGCCAGAATGGTACTTCTTATTCGCCTACGCCATTCTGCGATCTA 15360
P P H I Q P E W Y F L F A Y A I L R S I
15361 TCCCCAACAAGTTAGGAGGGGTAATCGCCCTAGTAGCAGCCATACTGGTCTTATTCTTGA 15420
P N K L G G V I A L V A A M L V L F L M
15421 TGCCCCTCCTAAACACCTCGAAGAAAGAGTCTAACTCTTTCCGACCGCTATCACAAGCAG 15480
P L L N T S K K E S N S F R P L S Q A A
15481 CCTTTTGGTTACTTGTTGCCCACCTTTTCATATTAACATGGATAGGCAGACAACCCGTAG 15540
F W L L V A H L F M L T W M G S Q P V E
15541 AGTACCCATATGTGTTACTCGGACAAGTGGCCTCAGTACTTTATTTTAGGCTATTTATAT 15600
Y P Y V L L G Q V A S V L Y F S L F M F
15601 TCGGGTTCCCCATAGTTTCTTCAATAGAAAACAAGATTATATTTTCTTA 15649
G F P M V S S M E N K I M F S *

BIBLIOGRAPHY

B.Alberts, D.Bray, J.Lewis, M.Raff, K.Roberts & J.D.Watson (1989) Molecular biology of the cell, 2nd edition. Garland Publishing, Inc., New York.

S.Anderson, A.T.Bankier, B.G.Barrell, M.H.L.de Bruijn, A.R.Coulson, J.Drouin, I.C.Eperon, D.P.Nierlich, B.A.Roe, F.Sanger, P.H.Schreier, A.J.H.Smith, R.Staden & I.G.Young (1980) Sequence and organisation of the human mitochondrial genome. Nature 290, 457-465.

S.Anderson, M.H.L.de Bruijn, A.R.Coulson, I.C.Eperon, F.Sanger & I.G.Young (1982) Complete sequence of bovine mitochondrial DNA. J. Mol. Biol. 156, 683-717.

H.Araki, P.A.Ropp, A.L.Johnson, L.H.Johnston, A.Morrison & A.Sugion (1992) DNA polymerase II, the probable homolog of mammalian DNA polymerase ϵ,replicates chromosomal DNA in the yeast Saccharomyces cerevisiae. EMBO J. 11, 733-740.

G.Attardi & Schatz (1988) Biogenesis of mitochondria. Ann. Rev. Cell Biol. 4, 289-333.

T.A.Baker & A.Kornberg (1991) DNA replication, 2nd edition. W.H.Freeman & Co, San Francisco.

G.Baldacci & G.Bernardi (1982) Replication origins are associated with transcription initiation sequences in the mitochondrial genome of yeast. EMBO J. 1, 987-994.

G.Baldacci, B.Cherif-Zahar & G.Bernardi (1984) The initiation of DNA replication in the mitochondrial genome of yeast. EMBO J. 3, 2115-2120.

J.Battey, C.Moulding, R.Taub, W.Murphy, T.Stewart, H.Potter, G.Lenoir & P.Leder (1983) The human c-myc oncogene: structural consequences of translocation into the IgH locus in Burkitts lymphoma. Cell 34, 779-787.

L.Bell & B.Byers (1983) Separation of branched from linear DNA by two-dimensional gel electrophoresis. Anal. Biochem. 130, 527-535.

A.J.Bendich & S.B.Smith (1990) Curr.Genet. 17, 421-425.

J.L.Bennett & D.A.Clayton (1990) Efficient site-specific cleavage by RNase MRP requires interaction with two evolutionary conserved mitochondrial RNA sequences. Mol. Cell. Biol. 10, 2191-2201.

A.J.Berk & D.A.Clayton (1974) Mechanism of mitochondrial DNA replication in mouse L cells: asynchronous replication of strands, segregation of circular daughter molecules, aspects of topology and turnover of an initiation sequence. J. Mol. Biol. 86, 801-824.

M.J.Bibb, R.A.Van Etten, C.J.Wright, M.W.Walberg & D.A.Clayton (1981) Sequence and gene organisation of mouse mitochondrial DNA. Cell 26, 167-180.

J.Blow (1989) Eukaryotic chromosome replication requires both α and δ polymerases. Trends Genet. 5, 134-136.

P.H.Boer, L.Bonen, R.W.Lee & M.W.Gray (1985) Genes for respiratory chain proteins and ribosomal RNAs are present on a 16-kilobase-pair DNA species from Chlamydomonas reinhardtii mitochondria. Proc. Nat. Acad. Sci. (USA) 82, 3340-3344.

P.H.Boer & M.W.Gray (1988) Scrambled ribosomal RNA gene pieces in Chlamydomonas reinhardtii mitochondrial DNA. Cell 55, 399-411.

D.Bogenhagen & D.A.Clayton (1977) Mouse L-cell mitochondrial DNA molecules are selected randomly for replication throughout the cell cycle. Cell 11, 719-727.

D.Bogenhagen & D.A.Clayton (1978) Mechanism of mitochondrial DNA replication in mouse L cells: kinetics of synthesis and turnover of initiation sequence. J. Mol. Biol. 119, 49-68.

D.Bogenhagen, A.M.Gillum, P.A.Martens & D.A.Clayton (1978) Replication of mouse L-cell mitochondrial DNA. Cold Spring Harbor Symp. Quant. Biol. 43, 253-262.

R.Bordonne, G.Dirheime & R.P.Martin (1987) Transcription initiation and RNA processing of a yeast mitochondrial tRNA gene cluster. Nucl. Acids Res. 15, 7381-7394.

P.Borst (1991) Why kinetoplast DNA networks? Trends Genet. 7, 139-141.

A.Brennicke & D.A.Clayton (1981) Nucleotide assessment of alkali-sensitive sites in mouse mitochondrial DNA. J. Biol. Chem. 256, 10613-10617.

H.Bresch (1978) Mitochondrial profile densities and areas in different developmental stages of the sea urchin Sphaerechinus granularis. Exp. Cell Res. 111, 205-209.

B.J.Brewer (1988) When polymerases collide: replication and the transcriptional organisation of the E. coli chromosome. Cell 53, 679-686.

B.J.Brewer & W.L.Fangman (1987) The localization of replication origins on ARS plasmids in S. cerevisiae. Cell 51, 463-471.

B.J.Brewer & W.L.Fangman (1988) A replication fork barrier at the 3′ end of yeast ribosomal RNA genes. Cell 55, 637-643.

W.M.Brown, J.Shine & H.M.Goodman (1978) Human mitochondrial DNA: analysis of 7S DNA from the origin of replication. Proc. Nat. Acad. Sci. (USA) 75, 735-739.

C.V.Cabrera, H.T.Jacobs, J.W.Posakony, J.W.Grula, J.W.Roberts, R.J.Britten & E.H.Davidson (1983) Transcripts of three mitochondrial genes in the RNA of sea urchin eggs and embryos. Dev. Biol. 97, 500-505.

P.Cantatore, M.Roberti, P.Morisco, G.Rainaldi, M.N.Gadaleta & C.Saccone (1987) A novel gene order in the Paracentrotus lividus mitochondrial genome. Gene 53, 41-54.

P.Cantatore, M.Roberti, P.L.Polosa, A.Muskih & M.N.Gadaleta (1990) Mapping and characterisation of Paracentrotus lividus mitochondrial transcription units. Curr.Genet. 17, 235-245.

P.Cantatore, M.Roberti, G.Rainaldi, M.N.Gadaleta & C.Saccone (1989) The complete nucleotide sequence, gene organisation and genetic code of the mitochondrial genome of Paracentrotus lividus. J. Biol. Chem. 264, 10965-10975.

T.R.Cech (1990) Self-splicing of group I introns. Ann. Rev. Biochem. 59, 543-568.

D.D.Chang & D.A.Clayton (1984) Precise identification of individual promoters for transcription of each strand of the human mitochondrial DNA. Cell 36, 635-643.

D.D.Chang & D.A.Clayton (1985) Priming of human mitochondrial DNA replication occurs at the light-strand promoter. Proc. Nat. Acad. Sci. (USA) 82, 351-355.

D.D.Chang & D.A.Clayton (1987a) A novel endoribonuclease cleaves at a priming site of mouse mitochondrial DNA replication. EMBO J. 6, 409-417.

D.D.Chang & D.A.Clayton (1987b) A mammalian mitochondrial RNA processing activity contains nucleus-encoded RNA. Science 235, 1179-1184.

D.D.Chang & D.A.Clayton (1989) Mouse RNase MRP RNA is encoded by a nuclear gene and contains a decamer sequence complementary to a conserved region of mitochondrial RNA substrate. Cell 56, 131-139.

D.D.Chang, R.P.Fisher & D.A.Clayton (1987) Roles for promoter and RNA processing in the synthesis of mitochondrial displacement-loop strands. Biochim. Biophys. Acta 909, 85-91.

D.D.Chang, W.W.Hauswirth & D.A.Clayton (1985a) Replication priming and transcription initiate from precisely the same site in mouse mitochondrial DNA. EMBO J. 4, 1559-1567.

D.D.Chang, T.W.Wong, J.E.Hixson & D.A.Clayton (1985b) Regulatory sequences for mammalian mitochondrial transcription and replication. In: Achievements and Perspectives of Mitochondrial Research Volume II (Eds. E.Quagliariello et al) pp. 135-144.

J.W.Chase & I.B.Dawid (1972) Biogenesis of mitochondria during Xenopus laevis development. Dev. Biol. 27, 504-518.

L.-J.Chen & E.M.Orozoco Jr. (1988) Recognition of prokaryotic transcription terminators by spinach chloroplast RNA polymerase. Nucl. Acids Res. 16, 8411-8431.

J.Christiansen (1988) The 9S precursor of Escherichia coli 5S RNA has 3 structural domains: implications for processing. Nucl. Acids Res. 16, 7457-7476.

T.W.Christianson & D.A.Clayton (1988) A tridecamer DNA sequence supports human mitochondrial RNA 3′ end formation in vitro. Mol. Cell. Biol. 8, 4502-4509.

T.Christianson & M.Rabinowitz (1983) Identification of multiple transcriptional initiation sites on the yeast mitochondrial genome by in vitro capping with guanylyltransferase. J. Biol. Chem. 258, 14025-14033.

D.O.Clary & D.R.Wolstenholme (1985) The mitochondrial DNA molecule of Drosophila yakuba: nucleotide sequence, gene organisation and genetic code. J.Mol.Evol. 22, 252-271.

D.A.Clayton (1982) Replication of animal mitochondrial DNA. Cell 28, 693-705.

D.A.Clayton (1991) Replication and transcription in vertebrate mitochondrial DNA. Ann. Rev. Cell Biol. 7, 453-478.

M.Cooney, G.Czernuszewicz, E.H.Postel, S.J.Flint & M.E.Hogan (1988) Site-specific oligonucleotide binding represses transcription of the human c-myc gene in vitro. Science 241, 456-459.

W.B.Cosgrove & M.J.Skeen (1970) The cell cycle in Crithidia fasciculata. Temporal relationships between synthesis of deoxyribonucleic acid in nucleus and in kinetoplast. J. Protozool. 17, 172-177.

S.Crews, D.Ojala, J.Posakony, J.Nishiguchi & G.Attardi (1979) Nucleotide sequence of a region of human mitochondrial DNA containingthe precisely identified origin of replication. Nature 277, 192-198.

D.J.Cummings, I.A.MacNeil, J.Domenico & E.T.Matsuura (1985) Excision-amplification of mitochondrial DNA during senescence in Podospora anserina. J. Mol. Biol. 185, 659-680.

D.J.Cummings, K.L.McNally, J.M.Domenico & E.T.Matsuura (1990) Complete DNA sequence of the mitochondrial genome of Podospora anserina. Curr.Genet. 17, 357-402.

X-W.Deng, R.A.Wing & W.Gruissen (1989) The chloroplast genome exists in multimeric forms. Proc. Nat. Acad. Sci. (USA) 86, 4156-4160.

P.Desjardins & R.Morais (1990) Sequence and gene organisation of the chicken mitochondrial genome. A novel gene order in higher eukaryotes. J. Mol. Biol. 212, 599-634.

J.F.X.Diffley & B.Stillman (1991) A close relative of the nuclear chromosomal high-mobility protein HMG1 in yeast mitochondria. Proc. Nat. Acad. Sci. (USA) 88, 7864-7868.

J.N.Doda, C.T.Wright & D.A.Clayton (1981) Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. Proc. Nat. Acad. Sci. (USA) 78, 6116-6120.

D.T.Dubin, J.Montoya, K.D.Timko & G.Attardi (1982) Sequence analysis and precise mapping of the 3′ ends of HeLa mitochondrial ribosomal RNAs. J. Mol. Biol. 157, 1-19.

J.J.Dunn & F.W.Studier (1983) Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. J. Mol. Biol. 166, 477-535.

D.Dunon-Bluteau, A.Cordonnier & G.Brun (1987) DNA synthesis in a mitochondrial lysate of Xenopus laevis oocytes. H-strand replication in vitro. J. Mol. Biol. 197, 175-185.

N.J.Dyson, T.A.Brown, J.A.Ray, R.B.Waring, C.Scazzacchio & R.W.Davies (1989) Processing of mitochondrial RNA in Aspergillus nidulans. J. Mol. Biol. 208, 587-599.

P.Early, J.Rogers, M.Davis, K.Calame, M.Bond (1980) Two mRNAs can be produced from a single immunoglobulin u gene by alternative RNA processing pathways. Cell 20, 313-319.

D.J.Elliott (1990) Mitochondrial gene expression in early sea urchin development. PhD thesis, University of Glasgow.

D.J.Elliott & H.T.Jacobs (1989) Mutually exclusive synthetic pathways for sea urchin mitochondrial rRNA and mRNA. Mol. Cell. Biol. 9, 1069-1082.

D.Epel, R.Steinhardt, T.Humphreys & D.Mazia (1974) An analysis of a partial derepression of sea urchin eggs by ammonia; the existence of independent pathways. Dev. Biol. 40, 245-255.

I.C.Eperon, J.W.G.Janssen, J.M.J.Hoeijmakers & P.Borst (1983) The major transcripts of the kinetoplast DNA of Trypanosoma brucei are very small ribosomal RNAs. Nucl. Acids Res. 11, 105-125.

L.Ernster & G.Schatz (1981) Mitochondria: a historical review. J.Cell Biol. 91, 227s-252s.

J.E.Feagin, J.M.Abraham & K.Stuart (1988) Extensive editing of the cytochrome c oxidase III transcript in Trypanosoma brucei. Cell 53, 413-422.

R.P.Fisher, M.A.Parisi & D.A.Clayton (1989) Flexible recognition of rapidly evolving promoter sequences by mitochondrial transcription factor 1. Genes & Dev. 3, 2202-2217.

R.P.Fisher, J.N.Topper & D.A.Clayton (1987) Promoter selection in human mitochondria involves the binding of a transcription factor to orientation-independent upstream regulatory elements. Cell 50, 247-258.

P.J.Flory Jr. & J.Vinograd (1973) 5-bromodeoxyuridine labelling of monomeric and catenated circular mitochondrial DNA in HeLa cells. J. Mol. Biol. 74, 81-94.

R.Gelfand & G.Attardi (1981) Synthesis and turnover of mitochondrial ribonucleic acid in HeLa cells: the mature ribosomal and messenger ribonucleic acid species are metabolically unstable. Mol. Cell. Biol. 1, 497-511.

W.Gilbert & D.Dressler (1968) DNA replication: the rolling circle model. Cold Spring Harbor Symp. Quant. Biol. 33, 473-484.

A.M.Gillum & D.A.Clayton (1978) Displacenment-loop replication initiation sequence in animal mitochondrial DNA exists as a family of discrete lengths. Proc. Nat. Acad. Sci. (USA) 75, 677-681.

A.M.Gillum & D.A.Clayton (1979) Mechanism of mitochondrial DNA replication in mouse L-cells: priming during the initiation of heavy-strand synthesis. J. Mol. Biol. 135, 353-368.

J.M.Goddard & D.R.Wolstenholme (1978) Origin and direction of replication in mitochondrialDNA molecules from Drosophila melanogaster. Proc. Nat. Acad. Sci. (USA) 75, 3886-3890.

J.M.Goddard & D.R.Wolstenholme (1980) Origin and directionof replicationinmitochondrial DNA molecules from the genus Drosophila. Nucl. Acids Res. 8, 741-757.

H.A.Gold, J.N.Topper, D.A.Clayton & J.Craft (1989) The RNA processing enzyme RNase MRP is identical to the Th RNP and related to RNase P. Science 245, 1377-1380.

M.W.Gray (1989) The evolutionary origins of organelles. Trends Genet. 5, 294-299.

A.L.Greenleaf, J.L.Kelly & I.R.Lehman (1986) Yeast RPO41 gene product is required for transcription and maintenance of the mitochondrial genome. Proc. Nat. Acad. Sci. (USA) 83, 3391-3394.

L.A.Grivell (1989) Nucleo-mitochondrial interactions in yeast mitochondrial biogenesis. Eur. J. Biochem. 182, 477-493.

J.M.de Haas, J.Hille, F.Kors, B.van der Meer, A.J.Kool, O.Folkerts & H.J.J.Nijkamp (1991) Two potential Petunia hybrida mitochondrial replication origins show structural and in vitro functional homology with the animal mitochondrial heavy and light strand replication origins. Curr. Genet. 20, 503-513.

S.L.Hajduk, V.A.Klein & P.T.Englund (1984) Replication of kinetoplast DNA maxicircles. Cell 36, 483-492.

R.L.Hallberg (1974) Mitochondrial DNA in Xenopus laevis oocytes. I. Displacement loop occurrence. Dev. Biol. 38, 346-355.

K.Hancock & S.L.Hajduk (1990) The mitochondrial tRNAs of Trypanosoma brucei are nuclear encoded. J. Biol. Chem. 265, 19208-19215.

J.C.Hanvey, M.Shimizu & R.D.Wells (1988) Intramolecular DNA triplexes in supercoiled plasmids. Proc. Nat. Acad. Sci. (USA) 85, 6292-6296.

J.F.Hess, M.A.Parisi, J.L.Bennett & D.A.Clayton (1991) Impairment of mitochondrial transcription termination by a point mutation associated with the MELAS subgroup of mitochondrial encephalo-myopathies. Nature 351, 236-239.

J.H.Hoeijmakers & P.J.Weijers (1980) The segregation of kinetoplast DNA networks in Trypanosoma brucei. Plasmid 4, 97-116.

M.J.Hollingsworth & N.C.Martin (1986) RNase P activity in the mitochondrion of Saccharomyces cerevisiae depends on both mitochondrion and nuclear-encoded components. Mol. Cell. Biol. 6, 1058-1064.

J.E.Hixson, T.W.Wong & D.A.Clayton (1986) Both the conserved stem-loop and divergent 5′-flanking sequences are required for initiation at the human mitochondrial origin of light-strand DNA replication. J. Biol. Chem. 261, 2384-2390.

N.T.Hu & J.M.Messing (1982) The making of strand-specific M13 probes. Gene 17, 271-280.

U.Hubscher & P.Thommes (1992) DNA polymerase ϵ: in search of a function. Trends Biochem. 17, 55-58.

H.T.Jacobs (1991) Structural similarities between a mitochondrially encoded polypeptide and a family of prokaryotic respiratory toxins involved in plasmid maintenance suggest a novel mechanism for the evolutionary maintenance of mitochondrial DNA. J.Mol.Evol. 32, 333-339.

H.T.Jacobs, D.J.Elliott, V.B.Math & A.Farquharson (1988) Nucleotide sequence and gene organisation of sea urchin mitochondrial DNA. J. Mol. Biol. 202, 185-217.

H.T.Jacobs, E.R.Herbert & J.Rankine (1989) Sea urchin egg mitochondrial DNA contains a short displacement-loop (D-loop) in the replication origin region. Nucl. Acids Res. 17, 8949-8965.

A.Jacquier (1990) Self-splicing group II and nuclear pre-mRNA introns: how similar are they? Trends Biochem. 15, 351-354.

S.H.Jang & J.A.Jaehning (1991) The yeast mitochondrial RNA polymerase specificity factor, MTF1, is similar to bacterial sigma factors. J. Biol. Chem. 266, 22671-22677.

S.-T.Jeng, J.F.Gardner & R.I.Gumport (1990) Transcription termination by bacteriophage T7 RNA polymerase at rho-independent terminators. J. Biol. Chem. 265, 3823-3830.

H.Kasamatsu, L.I.Grossman, D.L.Robberson, R.Watson & J.Vinograd (1974) The replication and structure of mitochondrial DNA in animal cells. Cold Spring Harbor Symp. Quant. Biol. 38, 281-288.

H.Kasamatsu, D.L.Robberson & J.Vinograd (1971) A novel closed-circular mitochondrial DNA with properties of a replicating intermediate. Proc. Nat. Acad. Sci. (USA) 68, 2252-2257.

H.Kasamatsu & J.Vinograd (1973) Unidirectionality of replication in mouse mitochondrial DNA. Nature New Biol. 241, 103-105.

J.C.Kennel & A.M.Lambowitz (1989) Development of an *in vitro* system for *Neurospora crassa* mitochondrial DNA and identification of transcription initiation sites. Mol. Cell. Biol. 9, 3603-3613.

Y.Kohwi & T.Kohwi-Shigematsu (1988) Magnesium ion-dependent triple-helix structure formed by homopurine-homopyrimidine sequences in supercoiled plasmid DNA. Proc. Nat. Acad. Sci. (USA) 85, 3781-3785.

A.Kornberg (1980) DNA replication. W.H.Freeman & Company, San Francisco.

B.Kruse, N.Narasimhan & G.Attardi (1989) Termination of transcription in human mitochondria: identification and purification of a DNA binding protein factor that promotes termination. Cell 58, 391-397.

R.Landick & C.Yanofsky (1987) Transcription attenuation. In: *Escherichia coli* and *Salmonella typhimurium* cellular and molecular biology, vol 2 (Eds: F.C.Neidhardt *et al*). American Society for Microbiology, Washington DC. pp.1276-1301.

C.S.Levings III & G.G.Brown (1989) Molecular biology of plant mitochondria. Cell 56, 171-179.

D.M.Lonsdale, T.Brears, T.P.Hodge, S.A.Melville & W.H.Rottmann (1988) The plant mitochondrial genome: homologous recombination as a mechanism for generating heterogeneity. Phil.Trans.R.Soc. Lond.B 319, 149-163.

D.M.Lonsdale, T.P.Hodge & C.M-R.Fauron (1984) The physical map and organisation of the mitochondrial genome from fertile cytoplasm of maize. Nucl. Acids Res. 12, 9249-9261.

D.M.Lonsdale, T.P.Hodge, C.J.Howe & D.B.Stern (1983) Maize mitochondrial DNA contains a sequence homologous to the ribulose-1,5-bisphosphate carboxylase large subunit gene of chloroplast DNA. Cell 34, 1007-1014.

R.Maleszka, P.J.Skelly & G.D.Clark-Walker (1991) Rolling circle replication of DNA in yeast mitochondria. EMBO J. 10, 3923-3929.

L.Marechal-Drouard, J.-H.Weil & P.Guillemaut (1988) Import of several tRNAs from the cytoplasm into the mitochondria in bean Phaseolus vulgaris. Nucl. Acids Res. 16, 4777-4788.

P.A.Martens & D.A.Clayton (1979) Mechanism of DNA replication in mouse L-cells: localisation and sequence of the light-strand origin of replication. J. Mol. Biol. 135, 327-351.

L.Martin-Parras, P.Hernandez, M.L.Martinez-Robles & J.B.Schvartzman (1991) Unidirectional replication as visualised by two-dimensional agarose gel electrophoresis. J. Mol. Biol. 220, 843-853.

B.S.Masters, L.L.Stohl & D.A.Clayton (1987) Yeast mitochondrial RNA polymerase is homologous to those encoded by bacteriophages T3 and T7. Cell 51, 89-99.

H.Masukata & J.-I.Tomizawa (1990) A mechanism of formation of a persistent hybrid between elongating RNA and template DNA. Cell 62, 331-338.

L.Matsumoto, H.Kasamatsu, L.Piko & J.Vinograd (1974) Mitochondrial DNA replication in sea urchin oocytes. J.Cell Biol. 63, 146-159.

L.Matsumoto & L.Piko (1971) In vivo radioactive labelling of mitochondrial DNA in Arabacia punctulata. Biol. Bull. 141, 397.

A.G.Mayhook, A.-M.Rinaldi & H.T.Jacobs (1992) Replication origins and pause sites in sea urchin mitochondrial DNA. Proc. R. Soc. Lond. B 248, 85-94.

D.A.Mead, E.S.Skorup & B.Kemper (1985) Single-stranded DNA SP6 promoter plasmids for engineering mutant RNAs and proteins: synthesis of a "stretched" preparathyroid hormone. Nucl. Acids Res. 13, 1103-1118.

D.A.Mead, E.S.Skorup & B.Kemper (1986) Single-stranded DNA "blue" T7 promoter plasmids: a versatile tandem promoter system for cloning and protein engineering. Protein Eng. 1, 67-74.

D.A.Melton, P.A.Krieg, M.R.Rebagliati, T.Maniatis, K.Zinn & M.R.Green (1984) Efficient in vitro synthesis of biologically active RNA and RNA hybridisation probes from plasmids containing a bacteriophage SP6 promoter. Nucl. Acids Res. 12, 7035-7070.

F.Michel & B.Dujon (1983) Conservation of secondary structure in 2 intron families including mitochondrial-, chloroplast-, and nuclear-encoded members. EMBO J. 2, 33-38.

J.Montoya, G.L.Gaines & G.Attardi (1983) The pattern of transcription of the human mitochondrial rRNA genes reveals two overlapping transcription units. Cell 34, 151-159.

G.B.Morin & T.R.Cech (1986) The telomeres of the linear mitochondrial DNA of Tetrahymena thermophila consist of 53 bp tandem repeats. Cell 46, 873-883.

D.M.Mueller & G.S.Getz (1986) Transcriptional regulation of the mitochondrial genome of yeast Saccharomyces cerevisiae. J. Biol. Chem. 261, 11756-11764.

C.Nishida, P.Reinhard & S.Linn (1988) J. Biol. Chem. 263, 501-510. DNA repair synthesis in human fibroblasts requires DNA polymerase δ.

J.M.Ntambi, T.A.Shapiro, K.A.Ryan & P.T.Englund (1986) Ribonucleotides associated with a gap in newly replicated kinetoplast DNA minicircles from Trypanosoma brucei. J. Biol. Chem. 261, 11890-11895.

D.Ojala, J.Montoya & G.Attardi (1981) tRNA punctuation model of RNA processing in human mitochondria. Nature 290, 470-474.

K.A.Osinga, E.De Vries, G.Van der Horst & H.F.Tabak (1984) Processing of yeast mitochondrial messenger RNAs at a conserved dodecamer sequence. EMBO J. 3, 829-834.

J.D.Palmer (1988) Intraspecific variation and multicircularity in Brassica mitochondrial DNAs. Genetics 118, 341-352.

J.D.Palmer & L.A.Herbon (1987) Unicircular structure of the Brassica hirta mitochondrial genome. Curr.Genet. 11, 565-570.
M.A.Parisi & D.A.Clayton (1991) Similarity of human mitochondrial transcription factor 1 to high mobility group proteins. Science 252, 965-969.

L.Pica-Mattoccia & G.Attardi (1972) Expression of the mitochondrial genome in HeLa cells. IX. Replication of mitochondrial DNA in relationship to the cell cycle in HeLa cells. J. Mol. Biol. 64, 465-484.

L.Piko (1969) Absence of synthesis of mitochondrial DNA during early development in sea urchins. Am. Zool. 9, 1118

L.Piko & K.D.Taylor (1987) Amount of mitochondrial DNA and abundance of some mitochondrial gene transcripts in early mouse embryos. Dev. Biol. 123, 364-374.

L.Piko, A.Tyler & J.Vinograd (1967) Amount, location, priming capacity, circularity and other properties of cytoplasmic DNA in sea urchin eggs. Biol. Bull. 132, 68-90.

L.Piko & L.Matsumoto (1976) Number of mitochondria and some properties of mitochondrial DNA in the mouse egg. Dev. Biol. 49, 1-10.

B.Polisky, X-Y.Zhang & T.Fitzwater (1990) Mutations affecting primer RNA interaction with the replication repressor RNA I in plasmid ColE1: potential RNA folding pathway mutants. EMBO J. 9, 295-304.

J.W.Posakony, J.M.England & G.Attardi (1977) Mitochondrial growth and division during the cell cycle in HeLa cells. J. Cell Biol. 74, 468-491.

A.-M.Rinaldi, A.Storace, A.Arzone & V.Mutolo (1977) Cell nucleus negatively controls mitochondrial RNA synthesis during early sea urchin development. Cell Biol.Int.Rep. 1, 249.

A.-M.Rinaldi, G.De Leo, A.Arzone, I.Salcher, A.Storace & V.Mutolo (1979a) Biochemical and electron microscopic evidence that cell nucleus negatively controls mitochondrial genomic activity in early sea urchin development. Proc. Nat. Acad. Sci. (USA) 76, 1916-1920.

A.-M.Rinaldi, I.Salcher-Cillari & V.Mutolo (1979b) Mitochondrial division in non-nucleated sea urchin eggs. Cell Biol.Int.Rep. 3, 179.

A.-M.Rinaldi, I.Salcher-Cillari & A.M.Valenti (1981) Mitochondrial DNA synthesis in Xenopus laevis enucleated eggs. Cell Biol.Int. Rep. 5, 987.

A.-M.Rinaldi & G.Giudice (1985) Nuclear-cytoplasmic interactions in early development. In: Biology of Fertilization, Vol 3 (Eds: C.B.Metz & A.Monroy). Academic Press Inc., Orlando, Florida. pp.367-377.

D.L.Robberson, H.Kasamatsu & J.Vinograd (1972) Replication of mitochondrial DNA. Circular replicative intermediates in mouse L-cells. Proc. Nat. Acad. Sci. (USA) 69, 737-741.

M.Roberti, A.Mustich, M.N.Gadaleta & P.Cantatore (1991) Identification of two homologous mitochondrial DNA sequences, which bind strongly and specifically to a mitochondrial protein of Paracentrotus lividus. Nucl. Acids Res. 19, 6249-6254.

D.R.Robinson & K.Gull (1991) Basal body movements as a mechanism for mitochondrial genome segregation in the trypanosome cell cycle. Nature 352, 731-733.

B.A.Roe, D.-P.Ma, R.K.Wilson & J.F.-H.Wong (1985) The complete nucleotide sequence of the Xenopus laevis mitochondrial genome. J. Biol. Chem. 260, 9759-9774.

J.Rogers, P.Early, M.Davis, K.Calame, M.Bond (1980) Two mRNAs with different 3′ ends encode membrane-bound and secreted forms of immunoglobulin u chains. Cell 20, 303-312.

K.A.Ryan, T.A.Shapiro, C.A.Rauch & P.T.Englund (1988) Replication of kinetoplast DNA in trypanosomes. Ann.Rev.Microbiol. 42, 339-358.

A.H.Schinkel, M.J.A.Groot Koerkamp & H.F.Tabak (1988) Mitochondrial RNA polymerase of Saccharomyces cerevisiae: composition and mechanism of promoter recognition. EMBO J. 7, 3255-3262.

A.H.Schinkel & H.F.Tabak (1989) Mitochondrial RNA polymerase: dual role in transcription and replication. Trends Genet. 5, 149-154.

W.Schuster & A.Brennicke (1987) Plastid, nuclear and reverse transcriptase sequences in the mitochondrial genome of Oenothera: is genetic information transferred between organelles via RNA? EMBO J. 6, 2857-2863.

T.Sekiya, M.Kobayashi, T.Seki & K.Koike (1980) Nucleotide sequence of a cloned fragment of rat mitochondrial DNA containing the replication origin. Gene 11, 53-62.

C.Sheline, T.Melendy & D.S.Ray (1989) Replication of DNA minicircles inkinetoplasts isolated from Crithidia fasciculata: structure of nascent minicircles. Mol. Cell. Biol. 9, 169-176.

A.Simpson, Y.Suyama, H.Dewes, D.A.Campbell & L.Simpson (1989) Kinetoplastid mitochondria contain functional tRNAs which are encoded in nuclear DNA and also contain small minicircle and maxicircle transcripts of unknown function. Nucl. Acids Res. 17, 5427-5445.

L.Simpson (1987) The mitochondrial genome of kinetoplastid protozoa: genomic organisation, transcription, replication, and evolution. Ann.Rev.Microbiol. 41, 363-382.

M.Singer & P.Berg (1991) Genes and genomes. University Science Books, CA / Blackwell Scientific Publs, Oxford.

P.J.Skelly & R.Maleszka (1991) Distribution of intron sequences among 21 yeast species. Curr.Genet. 19, 89-94.

I.Small, L.Marechal-Drouard, J.Masson, G.Pelletier, A.Cosset, J.-H.Weil & A.Dietrich (1992) In vivo import of a normal or mutagenised heterologous transfer RNA into the mitochondria of transgenic plants: towards novel ways of influencing mitochondrial gene expression? EMBO J. 11, 1291-1296.

A.E.Souza, P.J.Myler & K.Stuart (1992) Maxicircle CR1 transcripts of Trypanosoma brucei are edited and developmentally regulated and encode a putative iron-sulfur protein homologous to a NADH dehydrogenase subunit. Mol. Cell. Biol. 12, 2100-2107.

M.Steinert & G.Steinert (1962) La synthese de l'acide desoxyribonucleique au cours du cycle de division de Trypanosoma mega. J. Protozool. 9, 203-211

L.Stryer (1988) Biochemistry, 3rd edition. W.H.Freeman & Co, San Fransisco.

K.Stuart (1991) RNA editing in mitochondrial RNA of trypanosomatids. Trends Biochem. 16, 68-72.

D.P.Tapper & D.A.Clayton (1981) Mechanism of replication of human mitochondrial DNA. J. Biol. Chem. 256, 5109-5115.

D.P.Tapper & D.A.Clayton (1982) Precise nucleotide location of the 5' ends of RNA primed nascent light strands of mouse mitochondrial DNA. J. Mol. Biol. 162, 1-16.

J.N.Topper & D.A.Clayton (1990a) Secondary structure of the RNA component of a nuclear/mitochondrial ribonucleoprotein. J. Biol. Chem. 265, 13245-13262.

J.N.Topper & D.A.Clayton (1990b) Characterization of human MRP/Th RNA and its nuclear gene: full length MRP/Th is an active endoribonuclease when assembled as an RNP. Nucl. Acids Res. 18, 793-799.

P.E.Thorsness & T.D.Fox (1990) Escape of DNA from mitochondria to the nucleus in Saccharomyces cerevisiae. Nature 346, 376-379.

A.Tzagoloff & A.M.Myers (1986) Genetics of mitochondrial biogenesis. Ann. Rev. Biochem. 55, 249-285.

J.Vieira & J.Messing (1982) The pUC plasmids, an M13mp7 derived system for insertion mutagenesis with synthetic universal primers. Gene 19, 259-268.

J.Vieira & J.Messing (1987) Production of single-stranded plasmid DNA. Methods in enzymology 153, 3-11.

M.W.Walberg & D.A.Clayton (1981) Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA. Nucl. Acids Res. 9, 5411-5421.

B.L.Ward, R.S.Anderson & A.J.Bendich (1981) The mitochondrial genome is large and variable in a family of plants (Cucurbitaceae). Cell 25, 793-803.

A.C.Webb & C.J.Camp (1979) Mitochondrial DNA synthesis in oocytes of X. laevis. Exp.Cell Res. 119, 414-418.

A.C.Webb & L.D.Smith (1977) Accumulation of mitochondrial DNA during oogenesis in X. laevis. Dev.Biol. 56, 219.

C.Welter, S.Dooley, K.D.Zang & N.Blin (1989) DNA curvature in front of the human mitochondrial L-strand replication origin with specific protein binding. Nucl. Acids Res. 17, 6077-6086.

F.H.Wilt (1973) Polyadenylation of maternal RNA of sea urchin eggs after fertilization. Proc. Nat. Acad. Sci. (USA) 70, 2345-2349.

D.R.Wolstenholme, K.Koike & P.Cochran-Fouts (1973a) Single-strand-containing replicating molecules of circular mitochondrial DNA. J.Cell Biol. 56, 230-245.

D.R.Wolstenholme, K.Koike & P.Cochran-Fouts (1973b) Replication of mitochondrial DNA: replicative forms of molecules from rat tissues and evidence for discontinuous replication. Cold Spring Harbor Symp. Quant. Biol. 38, 267-280.

D.R.Wolstenholme, J.L.MacFarlane, R.Okimoto, D.O.Clary & J.A.Wahleithner (1987) Bizarre tRNAS inferred from the DNA sequences of mitochondrial genomes of nematode worms. Proc. Nat. Acad. Sci. (USA) 84, 1324-1328.

T.W.Wong & D.A.Clayton (1985) In vitro replication of human mitochondrial DNA: accurate initiation at the origin of light-strand synthesis. Cell 42, 951-958.

T.W.Wong & D.A.Clayton (1986) DNA primase of human mitochondria is associated with structural RNA that is essential for enzymatic activity. Cell 45, 817-825.

T.D.Yager & P.H.von Hippel (1987) Transcription elongation and termination. In: Escherichia coli and Salmonella typhimurium cellular and molecular biology, vol 2 (Eds: F.C.Neidhardt et al). American Society for Microbiology, Washington DC. pp.1242-1275.

D.Yang, Y.Oyaizu, H.Oyaizu, G.J.Olsen & C.R.Woese (1985) Mitochondrial origins. Proc. Nat. Acad. Sci. (USA) 82, 4443-4447.

C.Yanisch-Perron, J.Vieira & J.Messing (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequencesof the M13mp18 and pUC19 vectors. Gene 33, 103-119.

M.de Zamaroczy, G.Faugeron-Fonty, G.Baldacci, R.Goursot & G.Bernardi (1984) The ori sequences of the mitochondrial genome of a wild-type yeast strain: number, location, orientation, and structure. Gene 32, 439-457.

H.Zhu, I.G.Macreadie & R.A.Butow (1987) RNA processing and expression of an intron-encoded protein in yeast mitochondria: role of a conserved dodecamer sequence. Mol. Cell. Biol. 7, 2530-2537.

M.M.Bland, C.S.Levings III & D.F.Matzinger (1986) The tobacco mitochondrial ATPase subunit 9 gene is closely linked to an open reading frame for a ribosomal protein. Mol. Gen. Genet. 204, 8-16.

P.T.Englund (1978) The replication of kinetoplast DNA networks in Crithidia fasciculata. Cell 14, 157-168.

P.T.Englund (1979) Free minicircles of kinetoplast DNA in Crithidia fasciculata. J. Biol. Chem. 254, 4895-4900.

M.W.Gray & P.H.Boer (1988) Organisation and expression of algal (Chlamydomoonas reinhardtii) mitochondrial DNA. Phil. Trans. R. Soc. Lond. B 319, 135-147.

J.A.Wahleithner & D.R.Wolstenholme (1988) Ribosomal protein S14 genes in broad bean mitochondrial DNA. Nucl. Acids Res. 16, 6897-6913.

F.Berthier, M.Renaud, S.Alziari & R.Durand (1986) RNA mapping on Drosophila mitochondrial DNA: precursors and template strands. Nucl. Acids Res. 14, 4519-4533.

C.A.Breitenberger, K.S.Browning, B.Alzner-DeWeerd & U.L.RajBhandary (1985) RNA processing in Neurospora crassa mitochondria: use of transfer RNA species as signals. EMBO J. 4, 185-195.

C.S.Gibbs (1988) The structure and expression of mammalian glutamine synthetase genes. PhD thesis, University of Glasgow.

E.Sbisa, M.Nardelli & C.Saccone (1988) Symmetric transcription of the replication origin of rat mitochondrial DNA. Gene 72, 309-310.

E.Sbisa, M.Nardelli, F.Tanzariello, A.Tullo & C.Saccone (1990) The complete and symmetric transcription of the main non coding region of rat mitochondrial genome: in vivo mapping of heavy and light transcripts. Curr. Genet. 17, 247-253.

J.Tomizawa (1986) Control of ColE1 plasmid replication: binding of RNA I to RNA II and inhibition of primer formation. Cell 47, 89-97.