GENES IN MAIZE MITOCHONDRIA

By

ADAM DAWSON

DOCTOR OF PHILOSOPHY

UNIVERSITY OF EDINBURGH

1983

I declare that this thesis was composed by myself, and that the work presented herein, unless otherwise stated, is my own.

CONTENTS

CHAPTER 2 MATERIALS AND METHODS

ACKNOWLEDGEMENTS

I am indebted to my family for constant moral and financial support prior to the completion of this thesis, and to my supervisor, Dr C J Leaver, for guidance and encouragement during the course of my research.

I should also like to thank my colleagues in the Departments of Botany and Molecular Biology at Edinburgh, in particular Sue Dunbar, Pete Isaac and Val Jones, for helpful discussions and practical advice. I am also grateful to Professor L Grivell and his staff at the University of Amsterdam for allowing me to work in his laboratory in July 1982; to Anna Nowosielska for typing this thesis; to Sandra Isaac for assistance in the completion of some of the figures; and to Val for cooking my tea as I wrote it.

This research was financed by a three year SERC postgraduate studentship and was supplemented by travel awards from the SERC and from the University of Edinburgh Rennie Bequest Fund.

ABSTRACT

Maize mitochondrial (mt) DNA fragments containing the gene encoding apocytochrome b (COB) were identified by 'heterologous' hybridisation with DNA probes containing this gene from fungi and mammals. These fragments were cloned and subsequent sequence analysis showed that the maize COB gene is 1164 BP long and probably does not contain introns. The predicted polypeptide encoded by COB shares 47 and 52% amino acid sequence homology with its homologues in S.cerevisiae and H.sapiens respectively. Hydrophobicity profiles of these three polypeptides were similar, and revealed the presence of nine, hydrophobic, membrane-spanning domains in each case. Comparison of these predicted amino acid sequences indicated that in maize TGG is the preferred tryptophan codon and that TGA codons are absent.

RNA hybridisation experiments showed that the COB gene is actively transcribed in maize mitochondria. The transcription pattern is complex and may result from sequential processing of a longer precursor transcript. Analysis of DNA sequences immediately 5' to the AUG initiation codon, revealed an octanucleotide $^{5'}$AGTTGTCA, which has 62.5% complementarity with a sequence $^{5'}$UGAAUCCU at the 3' end of the maize 18S rRNA. It is proposed that this sequence may act as a ribosome binding site in the mature mRNA. Similar octanucleotides precede the maize cytochrome oxidase, subunit I (COI) and Oenothera cytochrome oxidase, subunit II (COII) genes.

DNA was isolated from maize plants with fertile (N) and with cytoplasmic male sterile (cms-T) phenotypes. The presence of a major sequence rearrangement, probably involving a deletion, 5' to the presumed start of the COII gene in cms-T mitochondria was demonstrated by DNA hybridisation analyses. Nucleotide sequencing confirmed that

the rearrangement occurs 627 BP 5' to COII. In both N and T mt DNAs, this sequence contains an open reading frame (ORF) which precedes, and is continuous with, that encoding COII. The ORF extends into the sequence which is rearranged in cms-T mt DNA and it seems likely that its expression will be altered in these plants. Functional roles for this polypeptide *in vivo* are discussed.

The COB, COI and COII genes have been located on a preliminary maize mt DNA restriction map prepared by D. Lonsdale. COB and COI are separated by about 40 KB whereas COII is at least 100 KB from both genes. All three genes are seemingly transcribed from the same DNA strand.

ABBREVIATIONS

AMPS	Ammonium persulphate
A_x	Absorbance at x nm wavelength
Ap, Ap^r, Ap^s	Ampicillin, Ap-resistant, Ap-sensitive
ATP	Adenosine 5' triphosphate
BP	Basepair(s)
BSA	Bovine serum albumin
CIP	Calf intestinal phosphatase
Cm, Cm^r, Cm^s	Chloramphenicol, Cm-resistant, Cm-sensitive
cms	Cytoplasmic male sterile
COB	The mitochondrial gene encoding apocytochrome b
COI, COII, COIII	The mitochondrial genes encoding cytochrome oxidase subunits I, II, III
cpm	Counts per minute
ct	Chloroplast
d(A,G,C and T)TP	2'deoxy (Adenosine, Guanosine, Cytidine and Thymidine) 5' triphosphates
DAPI	4', 6-Diamidino 2-phenyl indole.2 HCl
DCCD	N,N' Dicyclohexylcarbodiimide
dd(A,G,C and T)TP	2',3'deoxy (Adenosine, Guanosine, Cytidine and Thymidine) 5' triphosphates
DNA	Deoxyribonucleic acid
dpm	Disintigrations per minute
ds	Double stranded
DTT	Dithiothreitol
EDTA	Ethylene diamine tetra acetic acid
EGTA	Ethyleneglycol bis (β amino ethyl ether) N,N'tetra acetic acid
EM	Electron microscope
g_{ave}	Average relative centrifugal force
g_{max}	Maximum relative centrifugal force

GMAG	Genetic Manipulation Advisory Group
hr	Hour(s)
IPTG	Isopropyl thio ß D-galactoside
KB	Kilobase pair (= 1000 BP)(s)
$\log_{10}$	Logarithm (base 10)
LTB	Low tris buffer (defined section 2.2.11.1)
M Dal	Mega Dalton (= 10^6 Daltons)
MOPS	Morpholinopropane sulphonic acid
Mr	Relative molecular weight
mt	Mitochondrial
MW	Molecular weight
NAD^+	Nicotinamide adenine dinucleotide
NaPi	A mixture of aqueous solutions of Na_2HPO_4 and NaH_2PO_4, each having equimolar $[PO_4^{2-3}]$, in a ratio determined to yield a given pH
ORF	Open reading frame
PD	Potential difference
PEG	Polyethylene glycol
pfu	Plaque forming unit(s)
RBS	Ribosome binding site
RF	Replicative form
RNA; (m,t and r)RNA	Ribonucleic acid; (messenger, transfer and ribosomal) RNA
rpm	Revolutions per minute
RUBP	Ribulose 1,5 bisphosphate
S	Svedberg unit
SD sequence	Shine-Dalgarno sequence
SDS	Sodium dodecyl sulphate
ss	Single stranded
SSC	Standard saline citrate (defined section 2.1.5)
TAE	Tris acetate EDTA buffer (defined section 2.2.3.1)

TBE	Tris borate EDTA buffer (defined section 2.2.3.1)
Tc, Tc^r Tc^s	Tetracycline, Tc-resistant, Tc-sensitive
TEMED	N,N,N',N' tetramethylethylene diamine
Tm	Melting temperature (ds nucleic acid)
Tricine	N-tris (hydroxymethyl) methyl glycine
Tris	Tris (hydroxymethyl) amino methane
Triton X-100	Octylphenoxypolyethoxy ethanol
URF	Unassigned reading frame
uv	Ultraviolet
v/v	Volume per volume (as percentage)
w/v	Weight per volume (as percentage)
w/w	Weight per weight (as percentage)
X-gal	5-bromo 4-chloro 3-indolyl β D-galactoside

CHAPTER 1 GENERAL INTRODUCTION

1.1 INTRODUCTORY REMARKS

The biogenesis of a functional mitochondrion requires co-ordinate expression of genes contained in both the nucleus and the mitochondrion. This nucleo-mitochondrial interaction has been intensively studied in recent years and an understanding of some aspects of their interaction is now being reached. However, in almost all aspects, investigations of plant mitochondrial genes and the factors which regulate their expression, have lagged behind similar investigations in mammals and fungi.

The purpose of this Chapter is to review briefly the current state of knowledge concerning mitochondrial genes, with emphasis on those in higher plants, particularly maize (Zea mays L.). Historical aspects of mitochondrial investigation are the subject of recent reviews and textbooks (for example Tzagoloff, 1982; Grivell, 1983a; Dujon, 1981) and are not described here. Mitochondrial genome organisation and expression in higher plants has also been reviewed recently by Leaver and Gray (1982) and by Levings (1983a). More detailed introductions to certain aspects of the plant mitochondrial genome precede Chapters 4 and 5, and to avoid repetition, this material is not included in Chapter 1.

1.2 MITOCHONDRIA: AN OVERVIEW

Mitochondria are small cytoplasmic organelles found in almost all eukaryotic cells. They are bounded by two membranes, the inner of which is extensively invaginated. These two membranes define four compartments: the outer membrane, intermembrane space, inner membrane and matrix. The principal activity of the mitochondrion is the generation of ATP by the vectorial transport of high energy electrons, released during the metabolism of Acetyl CoA via the citric acid cycle.

The citric acid cycle enzymes are mostly soluble and located in the matrix, whereas those involved in electron transport are components of the protein-rich inner membrane. The details of these reactions and the enzymes catalysing them are discussed in most recent biochemistry textbooks (e.g. Tzagoloff, 1982) and are not included in this Chapter.

The number and conformation of mitochondria in a cell depends on the organism and its metabolic activity. Mammalian cells typically contain about 1000 ellipsoidal mitochondria (Bogenhagen and Clayton, 1974; Piko and Matsumoto, 1976), although these may become fused into a single spiral organelle in spermatozoa, or into a mitochondrial reticulum in muscle cells (Bakeeva et al, 1978). Yeast cells contain up to 50 mitochondria when growing aerobically (Stevens, 1981) although it is possible that the separate mitochondria observed in thin sections of tissue, may in fact represent branches of a single, highly convoluted 'chondriome'.

Higher plant mitochondria are generally similar in appearance, although the inner membrane may be convoluted to form tubuli rather than cristae as in other organisms. The form of the mitochondrion can also vary depending on growth conditions; for example, cytokinin starvation of soybean cells grown in tissue culture leads to marked mitochondrial elongation (A. Dawson, unpublished observations).

The most interesting feature of mitochondria from the point of view of this work, is that they contain an autonomous genetic system, distinct from that in the nucleus (Luck, 1963). It is therefore the mitochondrial genome which forms the basis for discussion in the remainder of this thesis.

1.3 SIZE AND CONFORMATION OF MITOCHONDRIAL DNAs

1.3.1 General characteristics

Mitochondrial (mt) DNAs can be isolated in a great many different forms and sizes, depending on the species and tissue studied. In HeLa, Drosophila and Paramoecium cells (Albring et al, 1977; Schatz and Mason, 1974; Olszewska and Tait, 1980), these mt DNAs are probably attached to the inner mitochondrial membrane and are tightly associated with protein, possibly having a structural role equivalent to that of the histones, but perhaps also involved in DNA replication. The situation in higher plants is unknown, though proteins associated with mt DNA have been visualised by electron microscopy in some species (Kim et al, 1982). The size and organisation of mt DNA in animals, fungi and higher plants is discussed below. The more esoteric mt DNAs found in certain protozoa (e.g. Tetrahymena, Paramoecium and the trypanosomes) are not discussed. More detailed descriptions of these DNAs can be found in Suyama and Miura (1968); Cummings et al (1980) and in Borst and Hoeijmakers (1979) respectively.

1.3.2 Animal mt DNAs

Electron microscopic examination of mt DNA isolated from a variety of animal species reveals the presence of small, double stranded, supercoiled molecules, of contour length 4.45-5.85 μm (reviewed by Borst, 1972). These mt DNAs normally display ideal second-order reassociation kinetics, indicating a lack of sequence reiteration (Clayton et al, 1970). The sequence complexities (ca. 10 M Dal or 15 KB) calculated from these experiments correlate closely with the sizes predicted by contour length measurements in the electron microscope.

Analysis of mt DNAs by analytical CsCl density gradient centrifugation indicates buoyant densities in the range 1.686-1.711 g/ml (Quetier and Vedel, 1980), equivalent to G + C contents of 36-51%

(Schildkraut et al, 1962). Mt DNA from Drosophila displays a lower G + C content and this is coupled with an unusually large inter-species variation in mt DNA contour length. These observations have now been attributed to the presence of an A + T-rich region of variable length, located in the mt DNA (see Fauron and Wolstenholme, 1980). If this sequence is discounted, the size and G + C content of the remaining DNA is very similar to that of other animal mt DNAs.

The results of the physical analyses described above have recently been confirmed by the complete sequence analyses of human (Anderson et al, 1981), mouse (Bibb et al, 1981) and bovine (Anderson et al, 1982) mt DNA, and by the partial sequence analyses of many others (e.g. rat, Xenopus and Drosophila). In all cases (except Drosophila, see above), the total sequence length was between 16.2 and 16.6 KB, forming a single circular molecule. Conclusions based on these sequence analyses are discussed in later sections (see below).

1.3.3 Fungal mt DNAs

Fungal mt DNAs display wide variation in their size and conformation. Circular DNA molecules, of contour length 25 μm (equivalent to 50 M Dal or 75 KB) have been isolated from yeast (Saccharomyces cerevisiae) mitochondria (Hollenberg et al, 1970), and circular mt DNAs, varying in size from 18.9 KB in Schizosaccharomyces pombe (Anziano et al, 1983) and Torolopsis glabrata (Clark-Walker et al, 1981a) to 108 KB in Brettanomyces custerii (Clark-Walker et al, 1981a; McArthur and Clark-Walker, 1983) have also been characterised. A few examples of linear fungal mt DNAs are known, for example in Hansenula mrakii (Wesolowski and Fukuhara, 1981) and Candida strain SR23 (Kovak et al, 1983), although the majority are circular.

In general, size values estimated from electron micrographs correlate well with those calculated from summation of restriction

endonuclease-generated mt DNA fragment sizes. However, estimates of sequence complexity based on renaturation kinetics are more ambiguous, at least in S.cerevisiae (Hollenberg et al, 1970). This probably results from the extremely high A + T content of this mt DNA, which has been shown to exceed 80% (Bernardi et al, 1970). Prunell and Bernardi (1974) have shown that A + T-rich regions span large portions of the yeast mt genome, and that this 'spacer' DNA is interspersed with several G + C-rich clusters; a finding which has been largely confirmed by DNA sequence analysis. – contains coding information. 70% transcribed.

Sequence analysis of the 70-80 KB S.cerevisiae mt genome is nearing completion (Grivell, 1983), as is that of Schizosaccharomyces pombe (18.9 KB) (Lang et al, 1983) and Aspergillus nidulans (33 KB) (Brown et al, 1983 and 1983a; Scazzocchio et al, 1983). A + T-rich regions are not found in these organisms, in contrast to S.cerevisiae. Neither are they found in the well-characterised Neurospora crassa (65 KB, Burke and Raj Bhandary, 1982; Luck and Reich, 1964) and Podospora anserina (94 KB, Wright et al, 1982) mt genomes. The function of these A + T-rich regions in S.cerevisiae is not well understood.

1.3.4 Higher plant mt DNAs

1.3.4.1 The multiple chromosome model

Early investigations of plant mt DNA failed to reach any consistent conclusions about either its complexity or conformation. The only generally agreed features were its remarkably constant buoyant density (ρ = 1.706 - 1.707 in most species - equivalent to a G + C content of ca. 47%; Wells and Ingle, 1970 - though as high as 1.710, equivalent to G + C = 51%, in Oenothera; Brennicke, 1980) and that it was probably even larger than that of S.cerevisiae.

Initial electron microscopic studies by Kolodner and Tewari (1972)

indicated that the major class of mt DNA molecule isolated from light-grown pea leaves was circular, with a contour length of 30-35 µm (equivalent to about 70 M dal or 100 KB). These results correlated well with their estimates of genome size (74 M dal) based on reassociaion kinetics. More recent reassessments of the mt genome size in pea (Ward et al, 1981) based on restriction enzyme analysis and renaturation kinetics have, however, suggested that a complexity of 200-250 M Dal may be more accurate. Numerous other studies based on restriction enzyme analysis have shown that higher plant mt genomes are generally much larger than initially indicated by Kolodner and Tewari. These values are summarised in Table 1.1 (see also Wallace, 1982).

Table 1.1 Sizes of plant mt DNAs

Species	Complexity (MDal)	Size (KB)	Method	Reference
Oenothera berteriana	120	185	a	Brennicke, 1980
Glycine max	60	90	b	Synenki et al, 1978
Zea mays	200	300	a	Pring & Levings, 1978
Zea mays	350-450	567-700	c	Lonsdale et al, 1983; D. Lonsdale, pers.comm.
Zea mays	320	480	d	Ward et al, 1981
Triticum aestivum	400	590	a/c	Falconet et al, 1983
Triticum aestivum	140	210	a	Quetier & Vedel, 1977
Curcurbits	220-1600	320-2400	a/d	Ward et al, 1981
Nicotiana tabacum	40	60	d	Wong & Wildman, 1972
Parthenocissus tricuspidata	168	240	a	Quetier & Vedel, 1977
Solanum tuberosum	>90	>135	a	Quetier & Vedel, 1977
Brassica spp.	120	180	a	Lebacq & Vedel, 1981
Brassica campestris	145	218	c	Palmer & Shields, 1983
Euglena gracilis	40	60	d	Talen et al, 1974
Pisum sativum	74	110	b/d	Kolodner & Tewari, 1972
Pisum sativum	240	360	a/d	Ward et al, 1981
Cucumis sativus	120	180	a	Quetier & Vedel, 1981
Lactuca sativa	140	210	d	Wells & Birnstiel, 1969

a = Summation of restriction enzyme fragment sizes

b = Electron microscopy contour length measurments

c = Restriction mapping

d = Reassociation kinetics

The table shows that genome sizes estimated by restriction analysis are consistently much larger than the largest molecules observed by electron microscopy. This is particularly marked for maize, where E.M. revealed circular molecules falling into three main size classes (33, 45 and 66 M Dal; Levings et al, 1979), yet summation of restriction fragment sizes indicated complexities between 183 M Dal (Spruill et al, 1980) and 320 M Dal (Ward et al, 1981). Two models have been proposed to reconcile this disparity:

1) Partial methylation of the DNA could increase the number of restriction fragments generated with certain methylation-sensitive enzymes. However, the use of isoschizomers recognising the same nucleotide sequence but differing in their ability to cleave methylated DNA (Bonen et al, 1980), or the use of enzymes specific for methylated DNA (Ward et al, 1981) has shown that this is unlikely to be the case in most plant mt DNAs.

2) The mitochondrial genome could be distributed over several small molecules (or 'chromosomes') represented by the small circular species seen in E.M. (Levings and Pring, 1976; Synenki et al, 1978; Quetier and Vedel, 1980; Spruill et al, 1980). This model goes some way towards interpreting the E.M. data, but does not explain why the majority of higher plant mt DNA, as isolated, is linear (Quetier and Vedel, 1980). It is also difficult to envisage how the 'chromosomes' become equally distributed during mitochondrial division, in the absence of any obvious 'spindle-like' mechanism for separating them. The possibility that different populations of mitochondria could contain different mt DNAs cannot be excluded, but is unlikely since DNA isolated from populations of mitochondria separated by velocity gradient centrifugation (Quetier and Vedel, 1980; Spruill et al, 1980) all display identical restriction digest profiles. By the same criteria, mitochondria isolated from different tissues of the same plant have been shown to contain the same mt DNA

(Ward et al, 1981; Quetier and Vedel, 1980). Spruill et al (1980) have shown that each 'chromosome' must contain different genetic information since hybridisation of cloned BamHI fragments of maize mt DNA to total mt DNA digested with BamHI indicated that the amount of unique DNA (121 M Dal) exceeded the coding capacity of even the largest circular molecule visualised by EM.

The multiple chromosome model explains most of the EM data but is unattractive on several counts. Moreover, intact chromosomes have never been recovered from normal plant cells. Supercoiled circular mt DNAs, separable by agarose gel electrophoresis, have only been isolated from plant cells grown in tissue culture (e.g. Oenothera: Brennicke and Blanz, 1982; Nicotiana: Sparks and Dale, 1980; Vicia faba and Zea mays: Dale, 1981; Dale et al, 1981). However, in contrast to the findings of Spruill et al (1980), hybridisation studies revealed that in bean, corn and tobacco, many of these molecules are related to one another in terms of the sequences they contain. The presence of similar molecules in non-cultured cells remains to be demonstrated and it has been shown, at least in maize, that mt DNA can undergo major rearrangement when cells are grown in tissue culture (Kemble et al, 1982). These supercoiled DNAs should not be confused with the plasmid-like DNAs found in higher plant mitochondria. These molecules are discussed in section 1.5.2.4.

1.3.4.2 Recombinational models.

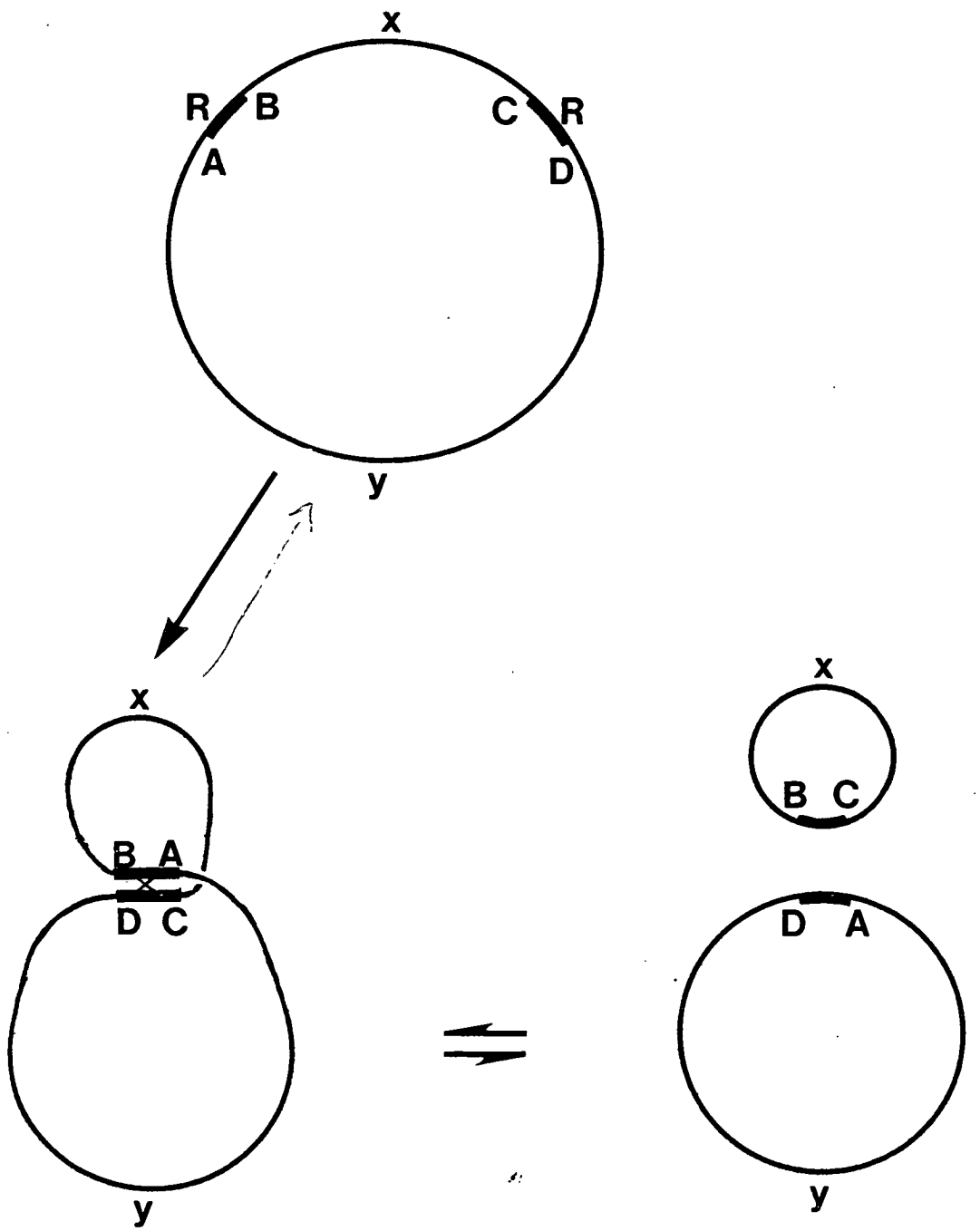
Recently, alternatives to the multiple chromosome model of higher plant mt genome organisation have been proposed. Palmer and Shields (1983) have, by restriction mapping clones containing portions of the Brassica campestris mt DNA, been able to construct a single circular restriction map of 218 KB. Contained within this sequence are two

direct repeats, each of 2 KB, separated by 135 and 83 KB. Palmer and Shields suggest that by 'illegitimate' recombination between these direct repeats, two smaller circles (of 135 and 83 KB) could be generated (Fig. 1.1). The mechanism of recombination they envisage is similar to that used by bacteriophage λ DNA when integrating into the host E.coli chromosome.

Mitochondrial genome organisation in maize is not yet so well defined, though it appears that a similar 'tripartite' structure could pertain (Lonsdale et al, 1983). By restriction mapping cosmid clones, this group has constructed a preliminary restriction map which can take the form of a single circular molecule of about 550-600 KB (see also section 4.5). This restriction map contains numerous direct and inverted repeats, and recombination seems to occur between many of these, leading to difficulties in the completion of the map. It is proposed that recombination occurs at high frequency between two direct repeated sequences of 3 KB each (although previously thought to be 26 KB: Lonsdale et al, 1981) separated by about 250 and 350 KB, in a manner exactly analogous to that proposed by Palmer and Shields for Brassica. More recent mapping data have shown that the larger circle formed by this recombination may also contain further direct repeats, leading to the formation of sub-circles of 250 and 67 KB (D. Lonsdale, pers. commun.). Lonsdale et al (1983) suggest that the three major circles (600, 350 and 250 KB) may be present in roughly equimolar proportions, since cloned DNA probes specific to the 3 KB repeat (R) hybridise to four XhoI fragments of mt DNA with equal intensity. The regions flanking the 3 KB repeat have been designated α, β, S1 and S2 (see Fig. 1.1 and also section 1.5.2.3 for description of S1 and S2 sequences). The four fragments correspond to the α-R-S1, α-R-S2, β-R-S1 and β-R-S2 combinations (see Fig. 1.1). The large molecule contains the α-R-S1 and β-R-S2 configuration.

Fig 1.1 The recombinational model for higher plant mitochondrial genome organisation.

In: B. campestris, R = 2KB; x = 83B and y = 135KB
Z. mays, R = 3KB; x = 250KB and y = 350KB
A ≡ ∝; B ≡ S1; C ≡ β and D ≡ S2
I. aestivum, R = 3.95KB (≡ 18S + 5S rRNA genes); x = ? and y = ?
A = 0.8KB ≡ u; B = 0.75KB (≡ w); C = 13.7KB (≡v) and
D = 1.45KB ≡ y (u - y are distances from termini of R
to the nearest Sal I site).



The mt genome of wheat (Triticum aestivum) probably assumes a similar configuration to that of Brassica and maize. However, hybridisation studies have revealed that the direct repeat unit in this species probably contains the 18S and 5S rRNA genes (Falconet et al, 1983) (see Fig. 1.1). This would contrast with maize, in which these genes are only present in single copies (Stern et al, 1982; Iams and Sinclair, 1982). The length of DNA sequence separating the repeat units in wheat is not known.

It seems possible that the single circle/intramolecular recombination model might apply to all higher plant mt DNAs. This model is certainly attractive and overcomes many of the objections levelled at the multiple chromosome model. If correct, the preponderance of linear molecules observed in electron micrographs of plant mt DNA is not unexpected since it is unlikely that circular molecules as large as 600 KB could be isolated intact. The smaller circular molecules visualised in these experiments may arise from rare recombination events between other repeated sequences in the mt genome. Stabilisation of these 'rare' events may give rise to the mt DNA concatemers isolated from plant cells grown in tissue culture.

Additional support for the model can be derived from studies of chloroplast (ct) DNA in Phaseolus vulgaris. Palmer (1983) has shown that this circular genome contains two inverted repeats of about 25 KB each, separated by 20 and 80 KB of unique sequence. Recombination between the repeats occurs at high frequency, leading to the formation of two equimolar ct DNA populations, each having the unique sequences in opposite orientations. These findings suggest that organellar genome reorganisation and instability could be widespread throughout the higher plants, therefore offering an explanation for the apparent complexity of their mt genomes. In addition, it has been shown that

different mt genomes of tobacco can recombine with each other when introduced into the same cell by cybrid fusion (Belliard, 1979). Higher plant mitochondria would therefore appear to contain the molecular machinery capable of supporting the recombinational model outlined above.

1.3.5 Mitochondrial DNA replication

Mt DNA replication in mammals has been well characterised by the examination of replicative intermediates separated by CsCl gradient centrifugation. It has been shown that replication of the two DNA strands (H and L) is initiated at two different points (designated O_H and O_L) in the genome. O_H is in the 'D-loop' region, a short portion of the genome formally equivalent to a DNA triplex. Replication of H-strand DNA proceeds by extension of the triplex structure until O_L is displaced, allowing initiation of L-strand replication. The complete process takes about 2 hr to complete and requires the activity of a nuclear gene-encoded γ polymerase (Bolden *et al*, 1977). Mt DNA replication appears to occur independently of nuclear genome replication. For a more detailed description of the mt DNA replication process see Clayton (1982) or Anderson *et al* (1982a).

Yeast mt DNA replication differs from that of the animals in that at least seven origins of replication exist (designated *ori* 1-7) and that both DNA strands are replicated from each (see review by Bernardi, 1982). Analysis of the DNA sequences retained in petite mt DNA mutants (see later) shows that the *ori* sequences are each about 400 BP long, and contain three G + C clusters separated by A + T rich sequences. These *ori* sequences appear to lie outside all known coding regions, and are capable of folding to form secondary structures similar to those of mammalian origins of replication (de Zamaroczy *et al*, 1981). In addition, a number of 'surrogate' *ori* sequences (*ori*^{S}) have been

identified which probably act as replication origins in petite mutants lacking canonical *ori* sequences. The precise mechanism by which replication is initiated at these sequences is not clearly understood (Blanc and Dujon, 1982), although the coincidence of RNA transcription promoters with *ori* sequences suggests that RNA priming could be involved (Bernardi, 1982).

In view of the much greater complexity of higher plant mt DNAs, it seems likely that different mechanisms will operate for its replication. At present no origins of replication have been identified, and it seems likely that more than one such origin will exist in each genome. Quetier and Vedel (1980) have observed putative replicating-intermediates by EM of a variety of plant mt DNAs. Likewise, Kim *et al* (1982) have described 'lasso-like' molecules, which they presumed to be replicative intermediaries, in the mt DNA of pearl millet (*Pennisetum typhoides*).

1.4 MITOCHONDRIAL GENOME ORGANISATION, EXPRESSION AND CODING CAPACITY

1.4.1 Genes in mitochondrial DNA

The experiments described in section 1.3 show that there is wide variation in the amount of unique sequence, and therefore potential coding capacity, in mt DNAs isolated from different organisms. The genes identified in these mt DNAs are discussed briefly below.

1.4.1.1 Animal mitochondrial genes

Sequence analyses of the human (Anderson *et al*, 1981), mouse Bibb *et al*, 1982) and bovine (Anderson *et al*, 1982) mt DNAs has revealed, in each, the presence of 13 open reading frames (ORFs) capable of encoding polypeptides containing more than 60 amino acids. These coding sequences are conserved in all animal species studied to date, as is the overall order of the genes. In addition, the mt DNAs contain

genes encoding Large (L, 16S) and Small (S, 12S) rRNAs and 22 tRNAs. These genes occupy almost all of the 16 KB genome, with little or no noncoding DNA separating them. All but one protein-coding ORF and 5 tRNA genes are transcribed from the same DNA strand. The only major noncoding region is the 'D-loop' sequence, which is approximately 1000-1500 BP long, depending on the species. This region may be involved in DNA replication (see section 1.3.5). Transfer RNA genes separate many of the protein-coding ORFs, and it has been suggested (Ojala et al, 1981) that these genes could form 'punctuation' signals recognised by enzymes involved in processing the primary RNA transcript of the mt DNA (see section 1.4.2.1). None of the mammalian mt genes have been shown to contain introns.

Comparison of the predicted amino acid sequences of the proteins encoded by the ORFs with published amino acid sequences of known mitochondrial proteins has allowed the genes encoding apocytochrome b and subunits I, II and III of cytochrome oxidase to be identified. Comparison with yeast mitochondrial gene sequences identified by analysis of mutants (see section 1.4.1.2) has also resulted in the identification of genes encoding ATPase subunits 6 (Macino and Tzagoloff, 1980) and 8 (Macreadie et al, 1983; Anderson et al, 1982a). These comparisons also showed that mammalian mitochondria use a novel genetic code (Barrell et al, 1979) in which UGA specifies tryptophan, and AUA specifies methionine (in the 'universal' code these triplets specify termination and isoleucine respectively). In addition, the absence of AGA and AGG from any of the ORFs suggests they may be used as termination codons. Codon usage in Drosophila melanogaster mitochondria is slightly different again, and is discussed by de Bruijn (1983).

The identity of the proteins encoded by the seven remaining open reading frames is not known. They are not merely confined to animal mt DNAs however since similar URF sequences have been identified in

Aspergillus nidulans (Scazzocchio, 1983), Neurospora crassa (Nelson et al, 1983) and Zea mays (Brown et al, 1983) mitochondria. All the identified reading frames encode hydrophobic polypeptides which are components of the electron transport and ATP generating systems. It is possible that the URFs code for further subunits of these multi-enzyme complexes, or possibly parts of the RNA-translating machinery. In any event, it seems that most mitochondrial proteins are encoded by nuclear genes, translated by cytoplasmic ribosomes and imported into the mitochondrion (see Schatz and Butow, 1983; also section 1.4.3). Mitochondria are not therefore 'autonomous' organelles in any meaningful sense of the word.

Attardi and co-workers (Attardi, 1981; Attardi et al, 1982) have studied protein synthesis in isolated HeLa cell mitochondria. They have shown that up to 26 translation products can be identified by one and two dimensional polyacrylamide gel electrophoresis. The origin of many of these polypeptides is obscure since the mt genome only has the capacity to encode at most 13 of them, and mRNA import into mitochondria has never been demonstrated. It is possible that proteolytic processing, or mRNA 'readthrough' may give rise to some of the additional apparent translation products (Attardi et al, 1982).

1.4.1.2 Fungal mitochondrial genes

The mt DNAs of most fungi are larger than those of the animals, and since their genomes largely contain unique DNA sequences, they therefore have a greater coding potential. It seems unlikely, however, that much of this excess coding capacity is actually realised, however, for the following reasons:

1) Isolated S.cerevisiae mitochondria only synthesise about 7 major polypeptides (Tzagoloff et al, 1979), which is actually fewer than the number synthesised by mammalian mitochondria.

2) The mt genomes of closely related fungi vary over five-fold in the amount of unique DNA which they contain (Clark-Walker et al, 1981). It seems unlikely that the different species could have acquired additional genes in the relatively short period since speciation occurred.

3) Much of the additional mt DNA in S.cerevisiae is very (> 80%) A + T rich (section 1.3.3) and unlikely to have coding capability (Prunell and Bernardi, 1974). This prediction has largely been verified by sequence analysis, with the notable exception of the Var-1 gene (Hudspeth et al, 1982) which escaped detection until recently because the DNA sequence which encodes it is almost entirely (89%) composed of A + T.

The identification of mitochondrially-encoded genes was pioneered in S.cerevisiae and identification of homologous genes in other species frequently relies solely on comparisons with published gene sequences from this yeast. Three properties of S.cerevisiae have been exploited in these analyses:

1) Mt DNA mutations which disrupt mitochondrial function are not necessarily lethal, since yeast can survive by anaerobic fermentation.

2) Cytoplasmic DNA is not inherited strictly maternally, unlike other fungi, animals and plants (Birky et al, 1982). This means that different mt DNAs introduced into the same cytoplasm by sexual crossing have the opportunity to recombine, thus allowing marker mutations (e.g. drug resistance; Dujon et al, 1977) to be mapped by classical genetic analyses.

3) Mt DNA will spontaneously mutate at a high rate to give 'petite' mutants which lack 80-100% their mt genomes. The mechanisms of 'petite' formation are now well understood and are reviewed by Bernardi (1982).

Combining these properties, mitochondrial loci identified in mt protein synthesis-deficient (syn^-) and respiratory deficient (mit^-) mutants could be mapped. Using a variety of biochemical and physical techniques, it has been possible to identify the variant polypeptides and RNAs produced by these mutants, and thereby to ascribe coding functions to the various genetic loci (see Borst and Grivell, 1978 for review). The repertoire of genes encoded is similar to that in animal mitochondria, with the addition of F_o ATPase subunit 9 (Hensgens et al, 1979; Macino and Tzagoloff, 1979), a ribosomal protein var-1 (Groot et al, 1979; Hudspeth et al, 1982) and a few short ORFs of unknown function (Coruzzi et al, 1981). No counterparts of the animal URFs have yet been discovered. Since the sequence analysis of the S.cerevisiae mt genome is almost complete, it seems unlikely that many more coding sequences will be identified.

S.cerevisiae uses a genetic code which differs again from both the 'universal' code and that used by mammalian mitochondria. Like the mammals, UGA encodes tryptophan (Fox, 1979b) and AUA specifies methionine (Hudspeth et al, 1982). However, CU(A,G,C and U) specify threonine rather than leucine (Li and Tzagoloff, 1979) and AGA and AGG are used as normal arginine codons (Hensgens et al, 1979). Similar codon usages have been deduced in other fungal genes, though the use of CUN as a threonine codon appears restricted to Saccharomyces.

In addition to the known protein coding genes, sequence analysis combined with genetic and physical approaches, has shown the capacity to encode two rRNAs (15S and 23S, see for example, Borst and Grivell, 1980b or Bos et al, 1980) and about 25 tRNAs. The order of these genes is different from that in the animals: notably, the tRNA genes are clustered and do not 'punctuate' the protein coding loci (Borst and Grivell, 1981b).

The same basic set of genes (encoding ATPase subunits 6 and 8; cytochrome oxidase subunits I, II and III; apocytochrome b, L and S rRNAs and about 25 tRNAs) are encoded in all other fungi examined by sequence analysis. The major differences are in the species-specific URFs and the presence in the mitochondrion of an active gene encoding ATPase subunit 9. In some species, this gene is apparently encoded in the nucleus (see Chapter 6). The order of the genes is not, however, preserved and varies even within the yeasts (Anziano et al, 1983; Clark-Walker and Sriprakash, 1981). Schizosaccharomyces pombe, which has one of the smallest fungal mt genomes (18.9 KB) appears to have adopted tRNA 'punctuation' like the animals (Lang et al, 1983). For more detailed discussions of genes and genome organisation in N.crassa, A.nidulans, Kloeckera africana and Podospora anserina mitochondria see Burger and Werner (1983), Scazzocchio et al (1983), Clark-Walker et al (1981b) and Wright et al (1982) or Wright and Cummings (1983) respectively. Many of these genes (particularly the L rRNA gene, COB and COI) contain introns, unlike their mammalaian counterparts. One such 'model' gene, cob-box or COB, (encoding apocytochrome b) from S.cerevisiae is described below.

1.4.1.3 The yeast COB gene

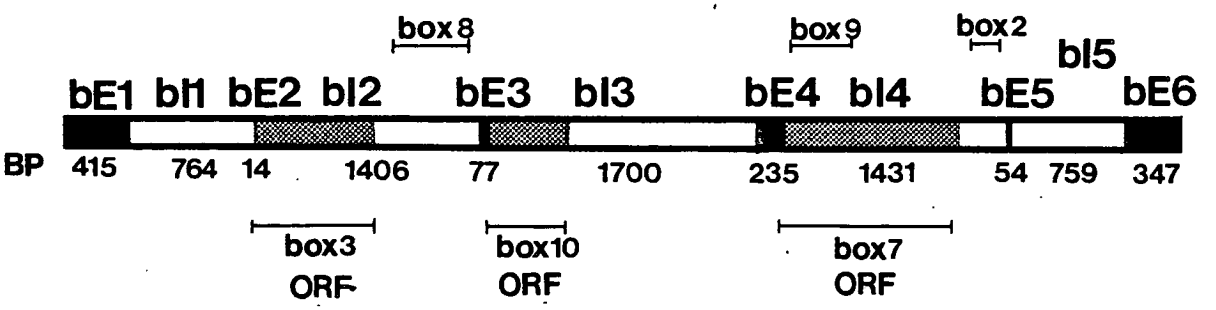
An unexpected observation from early mapping studies was that several complementation groups, e.g. oxi3 (the cytochrome oxidase subunit I gene, or COI) and COB mapped over inordinately large stretches of DNA: up to 12,882 BP for COI (Borst and Grivell, 1978; Hensgens et al, 1983a). The presence of introns in these genes predicted from the mapping studies has now been confirmed by sequence analysis (Nobrega and Tzagoloff, 1980; Lazowska et al, 1980; Bonitz et al, 1980; Hensgens et al, 1983a). These analyses also showed that a number of the introns are 'optional' and are only found in certain yeast

strains (Sanders et al, 1977; Borst, 1980). In the 'long' yeast strains KL14-4A and 777-3A, COB contains five introns (bI1-bI5), whereas the 'short' strain, D273-10B lacks the first three (Labouesse and Slominski, 1983).

It has been shown (Lazowska et al, 1980) that mutation in certain COB intron sequences (known as box loci 3, 7, 9 and 10) leads to the synthesis of defective and altered forms of cytochrome b (see, for example, Alexander et al, 1980). This led Lazowska et al (1980) to suggest that an intron-coded function, dubbed a 'maturase', was essential for the correct expression of COB. They postulated further that the maturase could be involved in the excision and splicing of the intron sequence which encodes it, from the mature mRNA. Both trans and cis acting intron box loci have been identified, and they correspond to the maturase encoding sequence, and the sequence recognised by the maturase, respectively (Borst and Grivell, 1981a). The 'maturase' is envisaged to be a readthrough protein derived from an mRNA combining both exon and intron sequences. It must be emphasised, however, that as yet an active maturase has never been isolated, so their existence is purely speculative. The most probable splicing pathway for COB is shown in Fig. 1.2 and is reviewed by Grivell et al (1982). The maturases, if genuine, probably form components of multi-enzyme splicing complexes, along with subunits encoded by nuclear genes. This nuclear involvement has been highlighted by the discovery of a nuclear mutation, NAM 2-1, which is able to suppress lesions in the box 7 maturase (Groudinsky et al, 1981). This is particularly interesting since this maturase is pleiotropic on the expression of COI, and is thought to splice intron aI4 from this gene, as well as bI4 from COB. Box 7 mutations can also be suppressed by the mim 2-1 mitochondrial mutation which maps in COI exon aI4. It is thought that this mutation alters a normally inactive maturase encoded in this

Fig 1.2 The COB gene in S. cerevisiae ('long' strains) From Jacq et al (1983); Nobrega and Tzagoloff (1980) and Lazowska et al (1980)

Splicing events:
1) bI1 → circular 10S RNA (nuclear maturase)
2) bE1 + bE2 + bI2 ORF → box 3 maturase
3) bI2 spliced by box 3 maturase, destroying its own mRNA
4) bE1 + bE2 + bE3 + bI3 ORF → box 10 maturase
5) bI3 spliced by box 10 maturase, destroying its own mRNA
6) bE1 + bE2 + bE3 + bE4 + bI4 ORF → box 7 maturase
7) BI4 and aI4 (COI) spliced by box 7 maturase, destroying its own mRNA
8) bI5 spliced by nuclear maturase.



intron, enabling it to splice bI4 (Dujardin et al, 1982). It has recently been shown (Dujardin et al, 1983) that NAM 2-1 is blocked by mutations in aI4, suggesting that this nuclear mutation also in some way activates the aI4 maturase.

The function of these introns in COB is not known, especially as all of them appear dispensible (Labouesse and Slominski, 1983). Mutants lacking all the COB introns apparently grow as well as strains containing five introns (provided aI4 is activated by NAM 2-1 or mim 2-1). The presence of ORFs in many of the introns and the fact that they are optional shows that they are not like 'normal' introns in nuclear DNA and probably do not confer the same advantages as proposed for these DNA sequences (Doolittle and Sapienza, 1980; Gilbert, 1978; Crick, 1979). Borst and Grivell (1981a) suggest that RNA splicing may provide an additional level at which the nucleus could exercise control over mitochondrial gene activity, and note that COI and COB are both specifically repressed during anaerobic growth on glucose.

The origin of these introns is not well understood. However, the finding that introns (and ORFs) displaying similar primary and secondary sequences (Davies et al, 1982) are widely dispersed throughout fungal mt genomes (e.g. S.cerevisiae L rRNA: Borst, 1979; N.crassa L rRNA: Burke and Raj Bhandary, 1982; A.nidulans L rRNA: Lazarus et al, 1980; P.anserina L rRNA: Wright and Cummings, 1983; N.crassa COB: Citterich et al, 1983; A.nidulans COB; Waring et al, 1983; S.pombe COB: Lang et al, 1983; S.cerevisiae COI: Hensgens et al, 1983; A.nidulans COI; Scazzocchio et al, 1983; N.crassa COI; Burger et al, 1982; S.pombe COI: Lang et al, 1983; T.glabrata and K.africana COI: Clark-Walker and Sriprakash, 1981; N.crassa ATPase 6 gene: Citterich et al, 1983a and N.crassa URF5: Nelson and Macino, 1983) suggests that all could have a common progenitor, perhaps analogous

to a bacterial transposon (see Grivell et al, 1982 for review). The finding that many of these introns are 'optional' (see for example Collins and Lambowitz, 1983) certainly appears to support this notion.

1.4.1.4 Higher plant mitochondrial genes.

The large complexity and low A + T content of most higher plant mt DNAs indicates a potential coding capacity greater than that of even the largest fungal mt genome (Ward et al, 1981). Probably because of this large size, the genes contained in plant mt DNAs have not been studied in depth until recently. The subject has been reviewed extensively by Leaver and Gray (1982), Leaver et al (1982) and Leaver et al (1983), and the salient features are described below.

1.4.1.4.1 Mitochondrial translation products. Isolated maize mitochondria synthesise at least 18-20 polypeptides which can be separated by one dimensional SDS-polyacrylmide gel electrophoresis (Leaver and Forde, 1980). Two dimensional separation reveals more (up to 50 in total; Hack and Leaver, 1980). Similar spectra of translation products have been identified in a variety of other higher plants (Leaver and Pope, 1976; Forde et al, 1979; Vedel et al, 1982). Because the synthesis of these polypeptides is insensitive to inhibitors of cytoplasmic (extramitochondrial) protein synthesis, and because mRNA import into mitochondria has not been demonstrated, these proteins are presumed to be encoded by genes in the mt DNA. These genes cannot be identified by classical genetic analyses for the reasons outlined in section 1.4.1.2. Moreover, the large size of the mt genome at present precludes complete sequence analysis to identify potential coding regions. A variety of techniques have been adopted to identify the following polypeptides synthesised by maize mitochondria:

1) Mitochondrially synthesised cytochrome oxidase subunits I and II cross reacted with antibodies raised against the homologous S.cerevisiae polypeptides, and were identified as translation products of M_r ca. 38,000 and 34,000 respectively (Forde and Leaver, 1979).

2) A major translation product of M_r 8000, which partitions into organic solvents and binds ^{14}C-DCCD (Hack and Leaver, in preparation). These properties suggest this protein is subunit 9 of the F_o ATPase.

3) A M_r 44,000 protein which co-purifies with mitoribosomes, suggesting a functional equivalence to the yeast Var-1 polypeptide (C.J. Leaver, pers. commun.).

4) The α subunit of the F_1 ATPase is also thought to be encoded in the mt DNA since a mitochondrial translation product of M_r 58,000 co-migrates with the α subunit on one- and two-dimensional SDS-polyacrylamide gels (Hack and Leaver, 1983). Peptide mapping and immunoprecipitation with antibodies raised against the yeast α subunit have confirmed this identification. The α subunit is also a mitochondrial translation product in sorghum, cucumber, pea and V.faba (Hack and Leaver, 1983; Boutry et al, 1983), in contrast to all other eukaryotic mitochondria which do not synthesise this protein.

The identity of the remaining polypeptides is unknown, although one, at least, must be cytochrome b (see Chapter 4). By analogy with mammals and fungi, it is likely that ATPase subunits 6 and 8, and cytochrome oxidase subunit III, will be among the remainder.

1.4.1.4.2 Hybridisation studies. Leaver and Harmey (1976) showed that higher plant mitochondria, unlike those from other eukaryotes, contain a 5S rRNA. In addition, higher plant mitoribosomes contain large (26S) and small (18S) rRNAs (Leaver and Harmey, 1973). Hybridisation experiments, using end labelled RNA, showed that the genes

encoding these rRNAs were located on specific restriction fragments of wheat mt DNA (Bonen and Gray, 1980). Mapping studies indicate that in wheat and maize (Stern et al, 1982; Iams and Sinclair, 1982) the 18S and 5S rRNA genes are closely linked, whereas the 26S gene is distant from this locus (see also section 1.3.4.2). Sequence analysis of the wheat 5S and 18S rRNA genes has recently been completed (Spencer et al, 1983), and a partial sequence of the maize 18S-5S unit has been reported by Chao et al (1983). These analyses show that higher plant 18S rRNAs are very much more eubacterial-like than those from either fungal or mammalian mitochondria. The significance of this finding and its implications for mitochondrial protein synthesis are discussed in section 4.4.4. Similar hybridisation experiments have shown that wheat and maize mt genomes encode several tRNAs, the genes for which, in maize, appear to be clustered (J.M. Grienenberger and D. Lonsdale, pers. commun.). The sequence of a wheat $tRNA_f^{met}$ gene has been reported by Gray and Spencer (1983).

Fox and Leaver (1981) have exploited the conservation of cytochrome oxidase subunit II sequences predicted from cross-reactivity with yeast antibodies, to locate the gene (COII, formerly mox-1) encoding this polypeptide in maize mt DNA. A single 2.4 KB EcoRI fragment of maize mt DNA was found to hybridise to 'heterologous' DNA probes containing the yeast COII gene. Sequence analysis confirmed that this fragment contained an ORF capable of specifying a polypeptide with 47% amino acid sequence homology to the equivalent yeast protein. Unexpectedly, the ORF was split by a single intron of 794 BP, unlike the COII sequence in other eukaryotes. The COII gene has recently been sequenced in Oenothera (Heisel and Brennicke, 1983) and does not contain an intron.

Sequence analysis of the maize COII gene and comparison of the predicted amino acid sequence with that from yeast and beef suggests

that the higher plant mitochondrial genetic code differs from the universal code and also from that used in animal and fungal mitochondria. Specifically, UGG is the major tryptophan codon, rather than UGA as in other mt DNAs. UGA codons were not found, suggesting that they may act as termination codons, as in the universal genetic code. In addition, CGG appears to be used as a tryptophan codon rather than specifying arginine as in all other genetic systems.

Extending the approach developed by Fox and Leaver (1981), it has been possible to identify and sequence the maize mt genes encoding COB (Chapter 4, this thesis) and COI (P. Isaac, pers. commun.). The results of these analyses are discussed in Chapters 4, 5 and 6. Additionally, an ORF sharing homology with the human mt URF-1 gene has recently been identified in maize mt DNA and partially sequenced (Brown et al, 1983; T.A. Brown, pers. commun.). The COB, URF-1 and COI genes are loosely grouped in a 40 KB region of mt DNA, whereas COII is over 100 KB away (section 4.5.3). All four genes appear to be transcribed from the same DNA strand and are not 'punctuated' by tRNA genes.

The approaches described above have allowed a few plant mitochondrial genes, or mt gene products, to be characterised. However, these genes could only occupy a very small proportion of the total potential coding capacity of the mt genome. Even allowing for the presence of a few genes normally found in the nucleus (e.g. the ATPase α subunit gene), and for the presence of 'promiscuous' DNA sequences of nonmitochondrial origin (Ellis, 1982; discussed in Chapter 6), it appears that much of the higher plant mt genome, despite its high G + C content and lack of reiteration, is noncoding. The excess coding capacity may comprise 'selfish DNA', analogous to that found in the nucleus

(Orgel and Crick, 1980), which simply fills out the intergene regions. This would explain why the mt genomes of closely related curcurbit species can vary so enormously in size (Ward et al, 1981) without presumably encoding significantly different numbers of genes.

1.4.2 Transcription of mitochondrial DNA

1.4.2.1 Animal mitochondrial RNAs.

The complete sequence analysis of many mammalian mt DNAs, combined with extensive transcript analyses, has allowed a fairly complete model for the transcription of these mt DNAs to be proposed (Montoya et al, 1981; Ojala et al, 1981). Both the H and L strands are completely transcribed (Murphy et al, 1975) from promoters in the D-loop region, near O_H. Since the L strand only encodes one protein and eight tRNAs (Anderson et al, 1982a), the primary transcript from this strand is apparently relatively long-lived. The H-strand is the main coding strand and processing of individual transcripts, by endonucleolytic cleavage on either side of the 'punctuating' tRNA sequences, probably occurs co-transcriptionally (Ojala et al, 1981). Following processing, rRNAs and mRNAs become polyadenylated (Hirsch and Penman, 1974) with approximately 55 nucleotides. In subsequent processing steps, these residues are removed from the rRNAs. It has been shown that polyadenylation is necessary to complete the termination codon in some mammalian mt genes which end with the incomplete triplets U or UA. The mature mRNAs lack 5' cap structures common to other eukaryotic mRNAs, they also normally lack 5' 'leader' and 3' trailing sequences. The mechanisms by which mitochondrial 55S ribosomes bind to these mRNAs are discussed in section 4.4.1.

An obvious limitation of this model is the inability to effect differential regulation of gene expression. Whilst this may be acceptable for the protein-coding genes, whose products are required

in fairly constant stoichiometry, it does not explain how the rRNAs can be overproduced as compared to the mRNAs. Attardi et al (1983) have shown that two separate promoters for H-strand transcription occur in the D-loop region, separated by the $tRNA^{Phe}$ gene. They conclude that transcription from the upstream promoter initiates frequently but terminates at the 3' end of the L rRNA gene, leading to the synthesis of the 12S and 16S rRNAs and $tRNA^{Phe}$ and $tRNA^{Val}$. Transcription from the downstream promoter occurs less frequently but does not terminate until all the rRNAs, tRNAs and mRNAs have been transcribed. How initiation by the same polymerase molecule at closely separated but different promoters can lead to transcription termination at two very different termini is not known.

It has been shown that transcription in Xenopus follows a similar pattern to that described in mammalian mitochondria (Rastl and Dawid, 1979). In view of the apparent constancy of mt genome organisation, it seems likely that a similar model may be applicable throughout the animal kingdom.

1.4.2.2 Fungal mitochondrial RNAs.

With the exception of S.pombe which probably employs tRNA 'punctuation' in the transcription of its compact mt genome (Lang et al, 1983; see also section 1.4.1.2), most fungal mt genomes are much larger than required to encode the numbers of genes they actually contain, and their transcription is more complex. S.cerevisiae mt DNA contains at least ten, and probably many more, promoter sequences (Christianson et al, 1983) dispersed throughout the genome. These promoters all contain a nonanucleotide concensus sequence $^{5'}$ATATAAGTA (Osinga and Tabak, 1982), and a similar sequence is thought to be located in Kluyveromyces lactis promoters (Osinga et al, 1982). Transcription is probably initiated at the last residue in the nona-

nucleotide. It has been found that promoter sequences do not always precede known protein-coding genes. In these cases the mature mRNA must form part of a long polycistronic precursor RNA. For example, no putative promoter sequence is found between the 5' end of the COB gene and a $tRNA^{glu}$ gene, 1050 BP upstream. Christianson et al (1983) show that the tRNA and COB genes are co-transcribed from a promoter 391 BP 5' to the tRNA gene. The resulting precursor RNA is extensively processed to yield a mature mRNA for COB having a 900 BP 5' leader sequence (intron processing is also required, as described in section 1.4.1.3). These experiments show that transcript termini defined by S1 nuclease protection will not necessarily identify the points at which transcription is initiated, since the primary transcripts may be present in too low an abundance to cause significant protection.

The signals for transcription termination in yeast mitochondria have not been investigated in detail. However, mature mRNAs have long 3' (and 5') nontranslated regions (Thalenfeld et al, 1983) and are apparently not polyadenylated (Borst and Grivell, 1978).

Mitochondrial transcription in other fungal species does not always conform to the models established in either the mammals or in yeast. For example, palindromic DNA sequences containing multiple PstI recognition sites seem to be involved in controlling transcription in Neurospora (Yin et al, 1980) although a putative promoter sequence with 70% homology to the yeast nonanucleotide (de Vries et al, 1983) precedes the COI gene. Despite the variations, a central feature of transcription in all mitochondria is a requirement for a nuclear-encoded RNA polymerase. This prerequisite forms another obvious level at which nuclear regulation of mitochondrial activity can occur.

1.4.2.3 Higher plant mitochondrial RNAs

Transcription and processing of higher plant mt RNA has not been

extensively investigated. It appears, however, that many features ascribed to yeast mt RNAs also apply in higher plants, notably the complex processing patterns and lack of polyadenylation (C.J. Leaver, pers. commun.). Transcription of the maize COB, COI and COII genes is discussed in section 4.3.3, and a model to account for the observed transcription patterns is presented.

1.4.3 Translation of mitochondrial mRNA and assembly of mitochondrial proteins

Mitochondria contain a translational machinery distinct from that in the extramitochondrial cytoplasm (Kuntzel and Noll, 1967). Most notably, the ribosomes and constituent rRNAs differ considerably in their sedimentation coefficients and primary structure from their cytoplasmic counterparts (Table 1.2).

Table 1.2. Properties of mitochondrial ribosomes (Modified from Tzagoloff, 1982)

Ribosome	Sedimentation coefficient (S)				
	Small subunit	S rRNA	Large subunit	L rRNA	Holo-ribosome
Bacteria	30	5 & 16	50	23	70
Eukaryotic cytoplasm	40	5,5.8 & 18	60	28	80
Animal mitochondria	28-35	12-13	39-45	16	55-60
S.cerevisiae mitochondria	37	15-19	50	23-25	73-74
Z.mays mitochondria	44	5 & 18	60	26	77-78
Trypanosome mitochondria	-	9	-	12	-
Tetrahymena mitochondria	55	-	55	-	80

These structural differences have often been taken to suggest a bacterial ancestry of mitochondria (Kuntzel and Köchel, 1981; Gray and Doolittle, 1982; Spencer et al, 1983; see also Chapter 6). However,

mitoribosomes are so diverse that it seems more appropriate to place them in a separate group from their prokaryotic or eukaryotic homologues. A unifying feature, however, of the group is a sensitivity to inhibitors of bacterial protein synthesis, such as chloramphenicol, and resistance to compounds such as cyclohexmide, which inhibit cytoplasmic ribosomes (McLean et al, 1958).

The biochemical requirements for protein synthesis by isolated mitochondria have been well characterised (reviewed by Schatz and Mason, 1974), and this system has been useful in the identification of mitochondrial translation products. However, the mechanics of protein synthesis have not been studied in depth. It is clear, though, that the process requires a considerable investment by the nucleus in terms of the necessary enzymes it must supply, Mammalian mitoribosomes alone contain at least 85 proteins, few, if any, of which are encoded in the mitochondrion (O'Brien and Matthews, 1976). The details of initiation of translation by mitoribosomes in animals, fungi and higher plants are discussed in section 4.4.2.

The assembly of a functional mitochondrion requires the co-ordination of gene expression in both the nucleus and the mitochondrion. Furthermore, it is important that the polypeptides encoded by these genes and synthesised in separate cellular compartments arrive at their correct destinations within the mitochondrion. It appears that 'addressing' systems are attached to many polypeptides destined for specific mitochondrial locations. These frequently take the form of amino-terminal extensions (or 'pre-pieces') which are subsequently proteolytically cleaved prior to assembly. Such signal sequences have been reported on both mt DNA-encoded proteins (e.g. S.cerevisiae and N.crassa cytochrome oxidase subunit II: Pratje et al, 1983; van den

Boogaart et al, 1982; and N.crassa cytochrome oxidase subunit I: Citterich et al, 1983) and those synthesised in the cytoplasm (Gasser et al, 1982). The probable pathways of import and assembly of these proteins are reviewed by Neupert and Schatz (1981) and by Schatz and Butow (1983).

1.4.4 Developmental regulation of mitochondrial gene expression

The changing energy requirements of a developing eukaryotic cell can be met in two ways: either by changes in the number of mitochondria in the cell, or by alteration of the activity of each mitochondrion. In practice, mammalian cells normally adopt the first alternative (Bogenhagen and Clayton, 1974; Piko and Matsumoto, 1976), probably because mammalian tissues provide relative homeostatic environments for the mitochondria, and prolonged periods of dormancy are not encountered. Similarly, the number of mitochondria in a yeast cell varies according to the growth conditions (see review by Stevens, 1981). In addition, ultrastructural changes within the mitochondria have been observed (Stevens, 1981) during the change from anaerobic ('glucose repressed') to aerobic growth (Plattner et al, 1970). In this case, release from glucose repression correlates with the differential de novo synthesis of mitochondrially-encoded components of the respiratory chain. This is probably accomplished by the increased activity of nucleus-encoded components of the mitochondrial transcription/translation machinery, and by the increased abundance of polypeptides forming subunits of the respiratory complexes (Frontali et al, 1982).

Mitochondrial biogenesis in higher plants has been studied in two model systems (Leaver and Gray, 1982): during seed germination and during artificially-induced ageing of storage tissue. These systems are discussed briefly below.

Mitochondria isolated from dry V.faba seeds show very little respiratory activity and no evidence of respiratory control (Dixon et al, 1980; Leaver and Forde, 1980). During the first 18 hr of imbibition (prior to cotyledon emergence), respiration increases considerably. However, the protein synthetic capacity of isolated mitochondria changes little during this period and remains very low. Hence, the initial recovery of mitochondrial activity is achieved independently of mitochondrial protein synthesis. Subsequent increases in respiratory activity (12-48 hr from imbibition) are accompanied by ultrastructural changes in the mitochondria and an increase in their protein synthetic capacity, including the specific synthesis of six polypeptides only made at a low level by mitochondria isolated from dry seeds. These results show that close co-operation between nuclear and mitochondrial genomes is required during seed germination, but the exact nature of this co-operation is not clear. Experiments are in progress in this laboratory (V. Jones, pers. commun.) to quantify the mRNA levels for specific mt genes (COB, COI and COII) present in V.faba tissues during the developmental process, in an effort to determine the level at which these controls are being exercised.

Ageing of excised Jerusalem artichoke tuber discs in distilled water induces mitochondrial biogenesis (Forde et al, 1979). It was found that mitochondria isolated from 26 hr-aged discs were able to incorporate exogenously supplied ^{35}S-methionine into protein approximately three times as fast as those from non-aged tissue. Some of this difference could be accounted for by a depletion in the mitochondrial methionine pool during ageing. It was noted, however, that the relative rates of synthesis of three polypeptides (M_r 17,000, 19,000 and 34,500) altered during ageing. This feature parallels the changes occurring during seed germination and also during de-repression of

yeast cells, illustrating that regulation of the activities of individual mt genes forms an important part of the developmental process.

1.5 MITOCHONDRIAL MUTATIONS AND CYTOPLASMIC MALE STERILITY

1.5.1 Types of mitochondrial DNA mutation

Genome mutation can take two basic forms: 'point' changes, i.e. additions, deletions and substitutions of single nucleotide pairs, and large scale changes, involving deletion, insertion or rearrangement of longer stretches of DNA. Because of their compact organisation, large scale mutation in mammalian mt genomes would almost certainly be lethal. However, point mutations do seem to accumulate at a remarkably high rate (e.g. Hauswirth and Laipis, 1982); a feature which has been useful in tracing mammalian evolutional lineages (Brown et al, 1979; Brown, 1980; Castora et al, 1980). The reason for this high rate of nucleotide substitution (ca. 0.02 substitutions/BP/10^6 years, about ten times as high as for nuclear DNA) is probably related to the absence of an efficient mt DNA repair system (Clayton 1982).

Point mutations leading to drug resistance have been recorded in both mammals (Wallace et al, 1982) and in S.cerevisiae (reviewed by Dujon, 1981). Additionally, point mutations leading to the disruption of mitochondrial function in yeast are well characterised and have been used in genetic analyses of the mt genome (section 1.4.1.2).

Perhaps because of the larger size of their mt DNAs, rearrangements and other forms of large-scale mutation are common in the mt genomes of the fungi. The best characterised are the yeast 'petite' mutants, which have lost large sections of their mt genomes and amplified greatly the small remaining portion (reviewed by Bernardi, 1982). The seeming lack of constraint placed on mt gene order in even closely related fungi (Clark-Walker and Sriprakash, 1981), and the fact that

portions of their mt genomes can become duplicated and inverted (Clark-Walker et al, 1981b; Brown et al, 1983a) shows that these rearrangements need not necessarily interrupt mitochondrial function.

Commensurate with their even greater size, the mt genomes of higher plants appear to undergo rearrangement and reordering at extremely high frequency (see also section 1.3.4.2). A detailed discussion of some of these rearrangements in maize can be found in Chapter 5. In the following section, the evidence that mt DNA rearrangement in higher plants is involved in cytoplasmically inherited male sterility (cms) is reviewed.

1.5.2. The mitochondrion and cms

1.5.2.1 The cms phenotype

Mitochondrial mutations in all eukaryotes except a few facultative aerobes, such as S.cerevisiae, are normally considered to be either lethal or without discernible phenotypic effect. However, a class of maternally-inherited mutations have been discovered in several higher plant species, which lead to male-sterility (e.g. Duvick, 1965; Edwardson, 1970). These mutations interfere with the normal processes of pollen formation or anthesis, and are commercially useful in the production of hybrid seed, since self pollination of the female parent can be prevented. Several lines of evidence, reviewed by Leaver and Gray (1982) and in section 1.5.2.2 below, suggest that cytoplasmically inherited male sterility is a result of mitochondrial dysfunction, and furthermore that the cms trait is carried in the mt DNA (Leaver, 1980; Levings, 1983a).

In a sense, cms mutations are conditionally lethal because they only interfere with mitochondrial activity under certain growth conditions, i.e. during anthesis. However, maize plants carrying the 'Texas' male-sterile cytoplasm (cms-T; Rogers and Edwardson, 1952)

are specifically susceptible to a toxin (T-toxin) produced by the phytopathogenic fungus Dreschlera [formerly Helminthosporium] maydis, race T. This toxin causes NAD^+ leakage from cms-T mitochondria, which may be the principal cause of mitochondrial degeneration and thus cell death, in cms-T plants (Matthews et al, 1979). It appears, therefore, that this toxin exploits a defect in cms-T mitochondria, allowing the effects of the cms mutation to be manifested throughout the whole plant, rather than being confined to the developing anthers. The outbreak of Southern Corn Leaf Blight, caused by this pathogen, in the United States in 1970 is well documented (Ullstrup, 1972). Considerable damage to the maize crop ensued, since almost all of the hybrid seed planted in that year had been derived from cms-T parents (see Ullstrup, 1972).

1.5.2.2 Evidence that cms results from a mutation in mt DNA

1) Cytoplasmic male sterile maize plants can be restored to fertility by dominant nuclear 'restorer' (Rf) genes. F_1 hybrid progeny from the cross cms - $\frac{rf}{rf}$ (♀) x male fertile - $\frac{Rf}{Rf}$ (♂) will have the genotype cms - $\frac{Rf}{rf}$, which is phenotypically male fertile. This is an important consideration when marketing hybrid plants for seed crops - for example a cms - $\frac{rf}{rf}$ maize plant would not shed pollen and could not therefore fertilise itself to produce a good cob.

Three groups of male sterile cytoplasm (C, S and T) have been recognised in maize, which differ in the Rf genes required to restore fertility (Beckett, 1971). Mt DNAs isolated from normal, male fertile (N) or cms-C, S and T plants can be distinguished both by their appearance in the electron microscope (Levings and Pring, 1976; Pring et al, 1978; Levings et al, 1979) and by their restriction enzyme digestion profiles (Pring and Levings, 1978). In contrast, chloroplast DNA isolated from N, C, S and T cytoplasms displays almost identical

restriction enzyme digestion patterns. The significance of the mt DNA rearrangements inferred from these studies is discussed in Chapter 5.

2) Mitochondria isolated from C, S and T cytoplasms synthesise a spectrum of polypeptides similar to those made by N mitochondria (see section 1.4.1.4.1). However, in each case, the cytoplasms could be characterised by the synthesis of one (or more) novel polypeptides. Specifically, cms-T and -C mitochondria synthesise proteins of M_r 13,000 and 17,500 respectively, which replace products of M_r 21,000 and 15,500 in N. Cms-S mitochondria are characterised by the additional synthesis of eight, low-abundance polypeptides of M_r 42-85,000 (Forde and Leaver, 1980; Forde et al, 1978). The role these variant polypeptides play in cms is not clear. Forde and Leaver (1979) have suggested that the M_r 13,000 polypeptide synthesised by cms-T mitochondria could act as a binding site both for a substance specifically produced during pollen formation, and for D.maydis T-toxin. Binding could in some way destabilise the inner mitochondrial membrane, leading to leakage of NAD^+. The finding that cms-T restorer genes suppress synthesis of the 13,000 M_r polypeptide (Forde and Leaver, 1980) is perhaps significant in this context. Walton (1983) has, however, shown that this does not lead to increased resistance to T-toxin, and suggests that the absence of the M_r 21,000 polypeptide is more closely correlated with toxin sensitivity.

3) Ultrastructural analyses of developing maize anthers in both N and cms-T plants show that mitochondrial degeneration is the first sign of abnormality in cms-T plants (Warmke and Lee, 1977). The degeneration follows a period of rapid mitochondrial division, suggesting the

'conditional' nature if the cms phenotype may be attributable to an inability of the mutant mitochondrion to function efficiently enough to meet the ATP requirements of these rapidly developing tissues.

4) Susceptibility to D.maydis T-toxin and the cms-T genotype are inseparable (see, for example, Brettell et al, 1980), indicating the two traits are probably manifestations of the same mutation. Therefore, the specific susceptibility of mitochondria from T cytoplasm to T-toxin, whilst chloroplasts from this same cytoplasm, and also mitochondria from N cytoplasm, are unaffected, suggests that the mutation resides in the mitochondrion.

1.5.2.3 Nucleo-mitochondrial interactions in cms

The fact that nuclear genes are able to suppress mitochondrial mutations leading to cms in maize shows that close co-operation between co-evolved nuclear and mitochondrial genomes is essential for the biogenesis of a normally functional mitochondrion. Male sterility arises when mitochondria are placed in a nuclear background unable to support their particular requirements. In this context, it is not entirely inconceivable that a nuclear background could be discovered in which 'N' mitochondria gave rise to male-sterile plants.

Nucleo-cytoplasmic interactions, of the type described above, are exploited in the generation of male-sterile plants in a variety of crop species. For example in sorghum, the combination of 9E cytoplasm with a 9E nuclear background yields a male fertile plant. However, when 9E cytoplasm is crossed into a Martin nuclear background, the resulting plant is male-sterile. Similarly inter-species and inter-generic crosses have been used to produce nucleo-mitochondrial combinations yielding male-sterile plants in tobacco (Belliard et al, 1979) and rape (Vedel et al, 1982). The sorghum Martin/9E combination

is particularly interesting because mitochondria isolated from these plants appear to synthesise an altered form of cytochrome oxidase, subunit I with lower electrophoretic mobility (M_r = 42,000 cf. 38,000 in fertile plants) (Dixon and Leaver, 1982). 'Northern' hybridisations show that the pattern of COI-specific RNA transcripts is identical in 9E/9E and Martin/Martin combinations, but differs in Martin/9E (D. Hanson and J. Bailey-Serres, pers. commun.). In this system, at least, it might be possible therefore, to establish a causal relationship between the synthesis of a variant mitochondrial polypeptide and the cms phenotype.

1.5.2.4 'Plasmid'-like DNAs in higher plant mitochondria

Episomal DNAs have been reported in a variety of fungal mitochondria (e.g. N.crassa strain FGSC 2225: Collins et al, 1981; Claviceps purpurea: Tudzynski et al, 1983 and Podospora anserina: (e.g.) Oziewacz and Esser, 1983) but never in mammalian mitochondria. Similarly, the presence of plasmid-like episomal DNA (and ds RNA) molecules in the mitochondria of many higher plant species is well documented. A brief description of some of these molecules is included in this section because an association between plasmids and cms has been suggested.

Maize mitochondrial 'plasmids' were first reported by Pring et al in 1977. Mitochondria isolated from cms-S plants were found to contain two linear DNA species of 6.2 and 5.2 KB, designated S1 and S2 respectively. These DNAs have the following properties:

1) S1 and S2 are normally present in equimolar amounts, and in a five-fold excess over main band or 'chromosomal' mt DNA (Thompson et al, 1980), comprising 10-15% of the DNA in a cms-S mitochondrion.

2) The termini of the episomes are covalently linked to proteins which may be involved in their replication (Kemble and Thompson, 1982).

3) S1 and S2 contain terminal inverted repeats of 150-200 BP (Kim et al, 1982a; Levings and Sederoff, 1983).

4) S1 and S2 share a common sequence of about 1100-1500 BP located at one end of the molecules.

5) Complete sequence analysis of S2 (shown to be 5,452 BP) revealed the presence of two long ORFs (of 3294 and 1017 BP) which might encode some of the high molecular weight polypeptides synthesised by cms-S mitochondria (see sections 1.5.2.2 and 4.4.4). This analysis also showed that S2 has some structural similarities to the DNAs of Bacillus phages and Adenovirus, suggesting that it might replicate in a similar fashion.

6) Cms-S plants revert to fertility at high frequency (Laughnan and Gabay, 1973). Reversion can be at either a cytoplasmic or nuclear level. The latter reversions are functionally equivalent to the generation of new Rf loci, although these loci map at several different points in the nuclear genome, and even on different chromosomes (Laughnan and Gabay, 1975). Cytoplasmic revertants are characterised by the loss of free S plasmids from the mitochondria, and by the appearance of novel 'mainband' mt DNA restriction fragments having homology to S1 and S2 (Levings et al, 1980).

7) Cms-S cells can revert to fertility when grown in tissue culture (Chourey and Kemble, 1982). In this case, movement of S1 and S2 sequences to mainband mt DNA also occurs, and this correlates with a change in callus morphology from 'organised' to 'friable'. It seems possible, therefore, that the S-plasmids might encode some cell surface factor, perhaps involved in cell:cell interaction phenomena.

8) Sequences homologous to S1 (but not S2) have been detected in nuclear DNA from N, SRf, nuclear S revertant and cms-S plants (Kemble et al, 1983). In each case, the pattern of hybridisation to

restriction fragments of n DNA was similar, showing that S1 is not acting as a mobile 'fertility element' (e.g. Laughnan and Gabay, 1975a) in nuclear reversion. The apparent correlation of S-plasmid integration into mt DNA with cytoplasmic reversion does indicate that the fertility element model would apply in this case, however. The presence of S1 sequences in the nucleus is unexpected and difficult to explain, although could perhaps be associated with the virus-like sequence proposed to be the progenitor of S1 and S2 (Thompson et al, 1980; Kemble et al, 1983).

9) Mainband mt DNA from N maize mitochondria contains integrated sequences homologous to S1 and S2 (Thompson et al, 1980). The integrated forms of S1 and S2 are rearranged and lack the 1100-1500 BP shared sequence (Thompson et al, 1980; McNay et al, 1983). They are also flanked on one side by repeated sequences, formerly thought to be 26 KB long (Lonsdale et al, 1981), but recently re-estimated at 3KB (Lonsdale et al, 1983, see also Fig. 1.1). S1 and S2 sequences also hybridise weakly to C, S and T mainband mt DNA. This could be due to the presence of: a) 'integrated' copies S1 and S2 on low abundance, high molecular weight mt DNA molecules; b) short sequences having complete homology with S1 and S2 in the most abundant mt DNA molecule(s) or c) long sequences having only partial homology to S1 and S2 in the abundant mt DNAs.

10) A number of male fertile Mexican Z.mays lines contain two linear episomes, designated R1 and R2 (Weissinger et al, 1981, 1983). R2 is probably identical to S2, R1 is similar to S1 but about 1 KB longer and contains 2.6 KB of unique DNA. Levings (1983a) has suggested that recombination between R1 and R2 could have yielded S1, thus explaining the origin of the 1.1-1.5 KB shared sequence. Plants containing R1 and R2 are not male-sterile, so the recombination event generating S1 may have led to the loss of sequences required for the maintenance of fertility.

Episomal mt DNAs are not restricted to cms-S maize lines. Additional linear and closed circular small DNAs have been detected in N and in cms-C, S and T mitochondria (Kemble and Bedbrook, 1980). All maize lines examined to date (Kemble and Bedbrook, 1980; Kemble et al, 1983) contain either a linear mt DNA species of 2.35 KB (designated n) or a closely related episome (designated t) of 2.0 KB. This molecule has homology to both S1 and S2 DNAs. Its universality, even in male-sterile lines, suggests it could play an essential role in mitochondrial biogenesis.

Plasmid DNAs have also been detected in certain Brassica species (Palmer et al, 1983), in sugarbeet (Powling, 1981) and in Sorghum (Dixon and Leaver, 1982; Pring et al, 1982) mitochondria. Some sugarbeet lines also contain a ds RNA species (Powling, 1981). In all cases, the presence of plasmid DNAs correlates to some extent with the cms phenotype, although the processes involved are not understood. The finding that linear Sorghum N1 and N2 mt plasmids share some sequence homology with S1 and S2 from maize (Pring et al, 1982) and are also associated with the synthesis of high molecular weight mitochondrial proteins (Dixon and Leaver, 1982) suggests that at least in these two examples, similar mechanisms may be operating.

1.6 SCOPE OF THIS THESIS

It is clear from an analysis of the literature reviewed in the preceding sections that while animal and fungal mitochondria and the genes they contain are well characterised, an equivalent understanding of plant mitochondrial gene organisation and expression has not yet been reached. In order to redress (partly) this balance, the work described in this thesis was undertaken. Two particular aspects of the maize mitochondrial genome were investigated:

1) The maize mitochondrial COB gene was isolated and sequenced. The results of this sequence analysis, presented in Chapter 4, confirmed and extended predictions made by Fox and Leaver (1981) based on analysis of the maize COII gene. In addition, expression of the gene was investigated both at a transcriptional level, by RNA hybridisation, and at a translation level by the identification of nucleotide sequences likely to form ribosome binding sites in the mature mRNA. The location of COB on the preliminary maize mt DNA restriction map prepared by D. Lonsdale, was also ascertained.

2) In order to characterise more fully the (presumed) mitochondrial mutation leading to cytoplasmic male-sterility, an investigation of mt DNA sequence rearrangement, characteristic of cms maize lines, was initiated. These studies, outlined in Chapter 5, focussed on sequences 5' to the COII gene and showed that a major sequence rearrangement occurs in cms-T mt DNA about 600 BP 5' to the most likely start point of the gene. This rearrangement could model the types of sequence alteration which lead to expression of the cms phenotype.

These two experimental sections cover different aspects of genome organisation and expression in maize mitochondria and are therefore introduced and discussed in separate chapters. Overall conclusions from this work, and some prospects for future research are discussed in Chapter 6.

CHAPTER 2 MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Maize seed

For isolation of mt DNA, plants were grown from seed of nuclear background B37 and mitochondrial genotype N (normal), C, S or T (male sterile), and for preparation of mt RNA, the hybrid N line 3451 was used. All seed was from Pioneer Hi-Bred International, Des Moines, Iowa (USA) and was of the most recent available crop.

2.1.2 Chemical reagents

All chemicals, except those noted below, were obtained from British Drug Houses (BDH) Ltd or from the Sigma Chemical Company and were of 'Analar' or 'Sigma' grade. Other reagents were:

Acrylamide and NN' Methylenebisacrylamide (for sequencing gels) from BDH and of 'Electran' grade.

Deoxynucleotide triphosphates and Tris from the Boehringer Corporation (London) Ltd.

Dideoxynucleotide triphosphates from P-L Biochemicals Inc.

Triton X-100 from Hopkin and Williams.

Agarose (Type I, Low EEO) from Miles Laboratories (Pty) Ltd.

Caesium Chloride from Fisons Scientific Apparatus, and of SLR grade.

Spectinomycin U-18409E (sulphate) from the Upjohn Company, Kalamazoo.

Ampicillin ('Penbritin' regd.) from Beecham Research Laboratories.

Herring sperm DNA (defatted) from Serva Feinbiochemica, Heidelberg.

Polyvinyl pyrrolidone from L. Light and Co Ltd.

Sephadex for column chromatography from Pharmacia Fine Chemicals.

M13 primer pentadecamer was from New England Biolabs, Inc.

Components for bacterial media are identified in section 2.1.5.

2.1.3 Enzymes

Restriction endonucleases were from Boehringer Mannheim GmbH, Bethesda Research Laboratories (UK) Ltd (BRL), Amersham International plc, or P & S Biochemicals Ltd.

T4-DNA ligase, pronase (Streptomyces griseus) and lysozyme were from Boehringer.

DNA polymerase I (E.coli) was from New England Biolabs.

DNA polymerase I (E.coli) (Klenow fragment) was from Boehringer or BRL.

DNAse I (RNAse-free) was from the Worthington Biochemical Corporation.

RNAse (DNAse-free) was from the Sigma Chemical Company Ltd.

Proteinase K (fungal) was from BRL.

Reverse transcriptase (avian myeloblastosis virus) was from Life Sciences Inc.

Calf intestinal phosphatase was from Boehringer and supplied as an ammonium sulphate suspension. Prior to use it was passed through a Sephadex G75 column pre-equilibrated with 20 mM Tris HCl pH 7.4 and 100 mM KCl. Active fractions were identified by their ability to prevent self ligation of restricted DNA (section 2.2.9.3.2) or by removal of $^{32}PO_4^{3-}$ moieties from γ-^{32}P dATP (Efstradiatis et al, 1977). CIP-containing fractions were pooled and stored at -20°C after the addition of 50% v/v glycerol.

All enzymes were stored at -20°C except DNAse I, RNAse and proteinase K, which were stored lyophilised at 0°C. Pronase was made up as a 10 mg/ml stock solution in H_2O and incubated at 37°C for 2 hr to remove contaminating DNAse and RNAse activities, prior to storage at -20°C.

2.1.4. Radioisotopes

α-^{32}P labelled dCTP (triethylammonium salt, in aqueous solution), at >400 Ci/mMol and 10 µCi/µl was from Amersham International plc and stored at -20°C. The 'activity date' for the isotope was the date and time at which the activity was 10 µCi/µl precisely, and was indicated on each batch of isotope. ^{32}P dCTP was renewed every 28 days.

2.1.5 Stock buffers and media

TE80 : 10 mM Tris, 1 mM EDTA, pH adjusted to 8.0 with 11.7 M HCl.

20 x SSC : 3 M NaCl, 0.3 M Na_3 citrate, pH adjusted to 7.0 with 1 M HCl.

100 x Denhardt's solution (Denhardt (1966)) : 2% w/v Ficoll (MW 400,000), 2% w/v polyvinylpyrrolidone (MW 40,000), 2% w/v BSA in 3 x SSC.

L-broth (Luria and Delbrück, 1943) : 1% w/v Difco Bacto Tryptone, 0.5% w/v Difco Bacto yeast extract, 0.5% w/v NaCl, pH adjusted to 7.2 with 5 M NaOH.

L-agar : 1% w/v Difco Bacto Tryptone, 0.5% w/v Difco Bacto yeast extract, 1% w/v NaCl, 1.5% w/v Difco agar, pH adjusted to 7.2 with 5 M NaOH.

Minimal agar : 0.2% w/v $(NH_4)_2SO_4$, 1.4% w/v K_2HPO_4, 0.6% w/v KH_2PO_4, 0.1% w/v Na_3 citrate, 0.02% w/v $MgSO_4$, 0.2% w/v glucose, 2.5×10^{-4}% w/v thiamine hydrochloride, 1.5% w/v Difco Bacto agar, pH not adjusted.

BBL top layer agar : 0.65% w/v Baltimore Biological Laboratories trypticase, 0.5% w/v NaCl, 1% w/v Difco agar, pH not adjusted.

Unless otherwise stated all buffers and media described in this thesis were made up in double distilled water and autoclaved at 121°C for 20 mins prior to use, or made up from individually sterilised components.

2.1.6 Miscellaneous

Reactions involving small volumes were carried out in 1.5 ml polypropylene tubes (Sarstedt) and where necessary centrifuged in an Eppendorf type 5412 (12 place) or 5413 (40 place) microfuge developing 12000 x g_{max}.

Larger volumes were manipulated in 15 ml or 30 ml Corex glass tubes, in 50 ml polypropylene tubes or in 250 ml polycarbonate bottles which could be centrifuged in a Sorvall RC-5B refrigerated superspeed centrifuge. Samples were ultracentrifuged when appropriate in a Sorvall OTD-50B ultracentrifuge fitted with a fixed angle TFT 65.13 rotor and an AH 627 swing out rotor.

Glassware and polypropylene tubes coming into contact with DNA were autoclaved (20 minutes, 121°C) prior to use; polyallomer and cellulose nitrate tubes were not sterilised.

Dialysis tubing (the Scientific Instrument Centre Ltd) was boiled in 10 mM $NaHCO_3$ for 5 minutes, rinsed in double distilled H_2O, re-boiled in 1 mM EDTA for 10 minutes then in double distilled H_2O for 10 minutes. Prepared tubing was stored at 4°C in 1 mM EDTA, 50% ethanol.

2.2 METHODS

2.2.1 Preparation of maize mt DNA

2.2.1.1. Buffers

Grinding buffer : 0.4 M Mannitol, 1 mM EGTA, 0.1% w/v BSA, 25 mM MOPS, 8 mM Cysteine HCl, 5 mM KCl, pH adjusted to 7.8 with 5 M KOH.

Wash buffer : 0.4 M Mannitol, 1 mM EGTA, 0.1% w/v BSA, 5 mM MOPS, pH adjusted to 7.5 with 0.5 M KOH.

Diluting medium : 0.2 M Mannitol, 1 mM EGTA, 10 mM Tricine, pH adjusted to 7.2 with 0.5 M KOH.

Resuspending medium : 0.4 M Mannitol, 1 mM EGTA, 10 mM Tricine, pH adjusted to 7.2 with 0.5 M KOH.

Stock buffered sucrose : 2.0 M Sucrose, 10 mM Tricine, 1 mM EGTA, pH adjusted to 7.2 with 0.5 M KOH. This was diluted as appropriate for sucrose gradients with 10 mM Tricine, 1 mM EGTA, pH 7.2.

Lysis buffer : 0.1 M Tris, 0.05 M EDTA, pH adjusted to 8.0 with 11.7 M HCl.

2.2.1.2. Growth of plant material

Seeds were surface sterilised in 7% w/v sodium hypochlorite (10 minutes) then imbibed in cold running water overnight and sown on autoclaved 12-ply cellulose wadding (Robinson and Sons Ltd) soaked in sterilised 0.1 mM $CaCl_2$. Seed was sown at a density of 100 g per 45 x 27 cm tray, and allowed to germinate in darkness for 96 hr at 28-30°C.

2.2.1.3. Isolation of mitochondria

Mitochondria were prepared as described below, which is a slight modification of the protocol devised by Forde *et al* (1978).

Etiolated coleoptiles were harvested by hand, weighed and rinsed three times in ice-cold sterile water. In order to prevent possible contamination of material, disposable gloves were worn at all stages of the preparation. Tissue homogenation and all subsequent steps were carried out at 4°C. After immersion in 2 ml grinding buffer per gram of starting material (normally 100-500 g), coleoptiles were partially disrupted in an MSE Atomix for 10 seconds on high speed then ground to homogeneity in a pestle and mortar. More vigorous methods of extraction were found to destroy mitochondrial integrity. The homogenate was filtered through four layers of muslin and 4 layers of milk filters (Blow, 'Maxa') then centrifuged at 1000 x $\underline{g}_{ave}$ for

5 minutes in a Sorvall GSA rotor. The supernatant was recentrifuged at 14,700 x $\bar{g}_{ave}$ in the same rotor for 15 minutes to yield a crude mitochondrial pellet which was resuspended using a teflon in glass homogeniser in 10-20 ml of wash medium per 100 g starting material and subjected to a second round of differential centrifugation in a Sorvall SS-34 rotor. The enriched mitochondrial pellet was resuspended in 2 ml of wash medium per 100 g starting tissue and applied to a linear sucrose gradient prepared by layering, from bottom to top, 3 ml 2.0 M, 8 ml each 1.45 M, 1.2 M and 0.9 M, and 4 ml 0.6 M buffered sucrose in a 36 ml polypropylene centrifuge tube. Prior to use, gradients were allowed to linearise overnight at 4°C. Mitochondria from 50 g of starting tissue were layered on each gradient, overlain with wash medium diluted 1:1 with H_2O, and centrifuged at 21,000 x $\bar{g}_{ave}$ in a Sorvall SS-90 vertical rotor for 50 minutes.

The buff coloured mitochondrial band at ca 1.2 M sucrose was recovered from the gradients, slowly mixed with 1.5 volumes of diluting medium and pelleted for 15 minutes at 13,000 x $\bar{g}_{ave}$ in the SS34 rotor. Mitochondria were resuspended in 1-2 ml of resuspending medium per 100 g starting tissue and collected by centrifugation in a microfuge at 12,000 x $\bar{g}_{max}$ for 4 minutes. Mt DNA was either isolated directly from this pellet, or it was stored at -80°C until required.

2.2.1.4. Extraction of mitochondrial DNA

A variety of methods for isolating mt DNA were tried initially but the protocol described below, a modification of that described by Fox (1979), was judged most satisfactory on a number of criteria (yield, RNA and cp DNA contamination, restriction enzyme analysis of the DNA, and preservation of low MW DNA species) and was used for all subsequent preparations.

The mitochondrial pellet from 50-500 g starting tissue was resuspended in 3.5 ml lysis buffer using a teflon in glass homogeniser. 0.3 ml of fresh 2 mg/ml proteinase K (in H_2O) and 0.5 ml of 10% w/v N-lauroyl sarcosine were added and the total mass brought to 5.65 g by the addition of lysis buffer. After 1 hr incubation at 60°C, 1 ml of 700 µg/ml ethidium bromide in lysis buffer and 6.4 g CsCl were added. The lysate was centrifuged at 85,000 x g_{ave} in a Kontron TFT 65.13 rotor for 24 hr at 4°C and after centrifugation, the mt DNA-containing band which fluoresced brightly under 366 nm (UV Products, Inc) UV illumination was removed into a 1ml syringe fitted with a 19G needle.

Ethidium bromide was removed from the DNA by repeated extraction with an equal volume of 1:1 iso propanol:n butanol saturated with CsCl-saturated TE80, until no further colour was visible in the aqueous layer (normally 5-6 extractions). The DNA was then dialysed to a 10^6 fold dilution for 24 hr against TE80 at 4°C and precipitated at -20°C with 0.1 volumes 3 M sodium acetate (pH 5.5) and 2.5 volumes ethanol. After dessication, the mt DNA pellet was dissolved in 50-100 µl of TE80 and yields were estimated from an A_{320} to A_{220} scan. Assuming that A_{260} = 1 corresponds to a DNA concentration of 50 µg/ml (Maniatis et al, 1982), yields of 0.1 to 0.2 µg mt DNA per gram of starting tissue were routinely obtained.

2.2.2. DNA digestion with restriction endonucleases

2.2.2.1 Buffers

10 x Low Salt buffer : 10 mM Tris, 10 mM $MgSO_4$, 1 mM DTT, pH adjusted to 7.4 with 1 M HCl.

10 x Medium Salt buffer : 50 mM NaCl, 10 mM Tris, 10 mM $MgSO_4$, 1 mM DTT, pH adjusted to 7.4 with 1 M HCl.

10 x High Salt buffer: 100 mM NaCl, 50 mM Tris, 10 mM $MgSO_4$, pH adjusted to 7.4 with 1 M HCl.

10 x KCl buffer: 20 mM KCl, 10 mM Tris, 10 mM $MgSO_4$, 1 mM DTT, pH adjusted to 8.0 with 1 M HCl.

Conversion buffer I (Low or KCl buffer → medium salt): 0.5 M NaCl.

Conversion buffer II (Low or KCl buffer → high salt): 1 M NaCl, 0.8 M Tris, pH adjusted to 7.4 with 1 M HCl.

Conversion buffer III (Medium → high salt) : 0.5 M NaCl, 0.8 M Tris, pH adjusted to 7.4 with 1 M HCl.

2.2.2.2 Reaction conditions

1-2 µg DNA in a final volume of 20 µl was normally used as a substrate for restriction enzyme reactions (see Szalay et al, 1979, for review). Larger amounts of DNA were digested in suitably scaled up volumes. Reactions contained 0.1 volumes of the appropriate 10 x buffer (Table 2.1) and were initiated by the addition of 1-20 units of restriction enzyme/µg DNA. Most mt DNA preparations were found to contain a non-dialysable contaminant which inhibited restriction enzyme digestion unless a 10 to 15-fold excess of enzyme over the manufacturer's recommendation was used. Plasmid DNAs were normally digested to completion with 1-5 units of enzyme per µg of DNA.

Digestions were for 30-120 minutes at 37°C except for PstI and SmaI, which were carried out at 30°C, and TaqI which was at 65°C. The latter was overlain with mineral oil to prevent volume loss through evaporation. Multiple enzyme digestions were carried out either simultaneously for enzymes with the same buffer requirements, or sequentially for those with incompatible buffers. In this case, the second enzyme was added 45 minutes after the first, together with 2 µl of conversion buffer I, II or III.

TABLE 2.1 Buffer requirements and sequence specificities of commonly used restriction endonucleases.

Low Salt	Medium Salt	High Salt	KCl buffer
BglII ($^{5'}$A$^{\blacktriangledown}$CATCT)	AccI (GT$^{\blacktriangledown}$[$^{AG}_{CT}$]AC)	EcoRI (G$^{\blacktriangledown}$AATTC)	SmaI (CCC$^{\blacktriangledown}$GGG)
HpaI (GTT$^{\blacktriangledown}$AAC)	AluI (AG$^{\blacktriangledown}$CT)	SalI (G$^{\blacktriangledown}$TCGAC)	
KpnI (GGTAC$^{\blacktriangledown}$C)	BamHI (G$^{\blacktriangledown}$GATCC)	XhoI (C$^{\blacktriangledown}$TCGAG)	
MspI (C$^{\blacktriangledown}$CGG)	ClaI (AT$^{\blacktriangledown}$CGAT)		
SacI (GAGCT$^{\blacktriangledown}$C)	HaeIII (GG$^{\blacktriangledown}$CC)		
TaqI (T$^{\blacktriangledown}$CGA)	HhaI (GCG$^{\blacktriangledown}$C)		
XmaI (C$^{\blacktriangledown}$CCGGG)	HindIII (A$^{\blacktriangledown}$AGCTT)		
	PstI (CTCGA$^{\blacktriangledown}$G)		
	Sau3A ($^{\blacktriangledown}$GATC)		

Reactions were terminated by heating to 65°C for 10 minutes followed by snap cooling on ice, except for TaqI which was inactivated by phenol extraction followed by ethanol precipitation of the DNA.

2.2.3 Electrophoresis of double stranded (ds) DNA

2.2.3.1 Buffers

20 x TAE : 0.8 M Tris, 0.4 M Na Acetate, 0.02 M EDTA, pH adjusted to 8.2 with glacial acetic acid.

10 x TBE : 0.89 M Tris, 0.89 M Boric acid, 0.02 M EDTA, pH = 8.3.

Loading buffer (vertical agarose gels) : 90% v/v glycerin, 0.1% w/v bromophenol blue.

Loading buffer (horizontal agarose gels) : 60% w/v sucrose, 0.3% w/v bromophenol blue, 80 mM EDTA.

Loading buffer (vertical polyacrylamide gels) : 25% v/v glycerol, 0.25% w/v bromophenol blue, 0.25% w/v Xylene cyanol FF, in 10 x TBE.

2.2.3.2 Agarose gel electrophoresis

1% w/v agarose slab gels were used to separate most DNA species between 1.0 and 10.0 KB (Fangman, 1978). 0.7 or 0.8% w/v gels were used for separation of larger DNA fragments (10-30 KB). Agarose was mixed with the appropriate volume of 1 x TAE and boiled for 5 minutes until completely dissolved, then cooled to 55°C and cast into horizontal or vertical moulds. Gels were run using 0.5-2.0 l of 1 x TAE in each tank, without recirculation, depending on the design of the apparatus.

2.2.3.2.1 Vertical gels . Glass plates were cleaned with methanol and assembled with greased spacers to form a gel mould 220 mm long x 140 mm wide and 3 mm thick. After casting, a 10, 12 or 15 tooth comb was inserted into the molten agarose to form loading slots 5-10 mm wide. The gel was allowed to set for 30 minutes at room temperature then the bottom spacer was removed and replaced with a thick strip of Whatman 3 MM paper. The comb was carefully removed and the gel clamped to a vertical gel stand, notched plate inwards.

DNA samples (0.1-2.0 µg) were loaded in 20-100 µl per track, after the addition of 0.3 volumes loading buffer and 0.3 volumes molten (37°C) 0.7% w/v low gelling temperature (LGT) agarose (Marine Colloids, EEO <0.15). The LGT agarose was found to reduce the 'tailing' of DNA bands which was otherwise a considerable problem with this gel system. Electrophoresis was at 1.5 V/cm for 16 hr by which time the bromophenol blue dye had migrated about $^{2}/3$ the length of the gel.

Vertical gels were found to give high resolution of DNA bands and were convenient when large volumes were to be loaded. However, because of the 'tailing' problem and the time taken to set up these gels, horizontal gels were used for most electrophoresis purposes later in this work.

2.2.3.2.2 Horizontal gels. Gels, 4 mm thick, were cast in a 220 x 150 mm or 250 x 200 mm flat bed mounted between agarose wicks to provide electrical continuity with the buffer tanks. Slots, 1 mm thick, 8 mm wide and 3.5 mm deep were formed with a 13 or 20 tooth comb. Sample volumes were adjusted to 20 µl, mixed with 5 µl loading buffer and applied to the gel. Electrophoresis was at 7.5 V/cm for 20 minutes to load the DNA samples, when sample wells were topped up and the gel surface covered with plastic film to prevent dehydration.

Electrophoresis was then continued at 1.4 V/cm for 16 hr, by which time the bromophenol blue dye had migrated about ¾ of the length of the gel.

2.2.3.2.3 Gel staining and photography. After electrophoresis, gels were immersed in 1 µg/ml ethidium bromide (Sharp et al, 1973) in H_2O for 20 minutes then destained in H_2O for 10 minutes. Gels were illuminated with incident short wave UV light (254 nm, Ultraviolet Products, Inc) and the fluorescent bands photographed through a Kodak Wratten no. 23A gelatin filter on Kodak 2415 technical pan film (10.2 x 12.7 cm).

2.2.3.3 Submerged agarose minigels

1% (w/v) agarose minigels, 100 mm wide x 69 mm long and 3 mm thick, were regularly used to assess DNA yields and to provide rapid analysis of restriction enzyme digests prior to loading the DNA on larger gels.

Gels were cast in 1 x TBE and run submerged under 30 ml of the same buffer. Sample volumes were adjusted to 7 µl and mixed with 2 µl loading buffer (for horizontal gels) prior to loading in 10 µl wells 5 mm wide. Electrophoresis was for 30-60 minutes at 60 mA and DNA was visualised with a mid range (302 nm) UV transilluminator (New England Biolabs, Inc) after staining for 10 minutes in 2.5 µg/ml ethidium bromide.

2.2.3.4 Polyacrylamide gels

This system was used for the analysis of small DNA fragments (10-1000 BP; Jovin, 1971), particularly those generated by tetra-nucleotide-recognising restriction enzymes.

6% gels were normally used and were cast in the same moulds as vertical agarose gels (section 2.2.3.2.1) with 1 mm spacers.

Gels contained:

5.8% w/v acrylamide

0.2% w/v NN' methylene bis acrylamide

0.07% w/v ammonium persulphate (AMPS)

0.05% v/v TEMED

5% v/v glycerol

in 1 x TBE.

Acrylamide and bis acrylamide were mixed and stored as a 5 x concentrated stock in H_2O at 4°C. AMPS was made up as a 10% w/v stock in H_2O and discarded immediately after use. All gel components except TEMED were mixed then TEMED was added and the gel poured. A 10 or 12 place comb was used to form loading wells 1 cm wide and the gel was allowed to polymerise for 30-60 minutes at room temperature. The comb and bottom spacer were removed when the gel had set and the assembly clamped to a vertical gel stand. Gels were run in 1 x TBE, and before loading were pre-run for 30 minutes at 5 V/cm to electrophorese contaminants from the gel. DNA samples (10-30 µl) were mixed with 0.1 volumes of loading buffer and applied to the gel using a drawn out pasteur pipette.

Electrophoresis was for 3 hr at 10 V/cm, by which time the bromophenol blue (found to co-migrate with fragments of about 50 BP) had moved approximately $^2/_3$ the length of the gel and the xylene cyanol $^1/_3$. DNA was stained and photographed as described in section 2.2.3.2.3, except that destain time was reduced to 5 minutes.

2.2.4 Recovery of DNA fragments from agarose gels

Several methods were used to isolate DNA fragments from agarose gels. These included "freeze-squeeze" (Thuring et al, 1975), electrophoresis onto DE81 paper (Dretzen et al, 1981), the use of

low gelling temperature agarose (Weislander, 1979) and electroelution into dialysis tubing (McDonnell et al, 1977). None proved entirely satisfactory though the latter, outlined below, was the easiest and gave the highest yield (10-50%) of DNA.

2.2.4.1 Buffer and phenol

20 x Electroelution buffer (EB) : 0.1 M Tris, 0.05 M acetic acid (pH = 7.8).

Phenol : Phenol was redistilled under N_2, equilibrated with TE80, and stored in small aliquots at -20°C.

2.2.4.2 Electroelution into dialysis tubing

DNA fragments were separated on horizontal or vertical agarose gels, stained with ethidium bromide and viewed under 302 nm UV light. The appropriate DNA band(s) were excised and placed in a section of sterile dialysis tubing with 2 ml of 1 x EB. The bag was sealed, immersed in 500 ml of 1 x EB and a potential gradient of 10 V/cm applied across the tank, with the dialysis sac at 90° to the current. Electrophoresis was continued until all of the ethidium bromide fluorescence was seen to have migrated out of the gel, when viewed under 302 nm UV light. The polarity of the current was then reversed for 10 minutes and the contents of the sac were transferred to a 5 ml disposable hypodermic syringe blocked with glass wool. The syringe was centrifuged briefly at 3000 x g_{ave} and the eluate collected in a 15 ml Corex centrifuge tube. The syringe was rinsed with a further 1 ml of electroelution buffer and re-centrifuged.

The pooled washings were extracted twice with an equal volume of phenol and the phenol was back extracted with 0.5 ml 1 x EB. Ethidium bromide and residual phenol were removed with two washings of n-butanol and DNA was precipitated at -80°C for 30 minutes by

the addition of $^{1}/_{9}$th volume of 3 M sodium acetate, pH 5.5 and 3 volumes ethanol.

DNA was pelleted in a Sorvall HB-4 rotor at 16,000 x $\bar{g}_{ave}$ for 15 minutes at -10°C, the supernatant drained, and the pellet washed by resuspension in 5 ml 70% ethanol then re-centrifuged as above. This second pellet was dried, dissolved in 200 µl TE80 and transferred to a 1.5 ml Sarstedt tube. DNA was re-precipitated with ethanol as above then centrifuged at 12,000 x $\bar{g}_{max}$ at room temperature. The final pellet was dried, dissolved in 5-10 µl TE80 and a 1 µl aliquot analysed by electrophoresis on a minigel to assess yield and integrity of the DNA.

The DNA prepared by this method was found to be a suitable substrate for nick-translation, restriction and ligation, though all these reactions occurred with lower efficiencies than with 'native' DNA prior to elution.

2.2.5 Transfer of DNA fragments from agarose gels to nitrocellulose

The technique devised by Southern (1975) was used with minor modifications to transfer and immobilise DNA on nitrocellulose filters and allow sequence homologies to be studied by hybridisation to ^{32}P-labelled DNA probes.

2.2.5.1 Buffers

Denaturation buffer (DB) : 0.5 M NaOH, 1.5 M NaCl.

Neutralisation buffer (NB) : 0.5 M Tris, 3.0 M NaCl, pH adjusted to 7.4 with 11.7 M HCl.

2.2.5.2 'Southern' transfer

Horizontal or vertical gels were run, stained and photographed as described in section 2.2.3, then soaked for 40 minutes, with one

change of buffer, in several volumes of DB with gentle agitation. After rinsing with distilled H_2O, the gel was further soaked in several volumes of NB for 40 minutes with one change, then placed on a wick of Whatman 3 MM paper soaked in 20 x SSC. The wick was suspended on a glass plate over a tank containing 200 ml of 20 x SSC.

A sheet of nitrocellulose (Schleicher and Schüll, BA85) the same size as the gel was first wetted with distilled H_2O then with 2 x SSC and carefully placed on the gel, using a soft pencil to mark the positions of the DNA tracks.

The filter was covered with 1 sheet of Whatman 3 MM paper soaked in 2 x SSC and the remainder of the apparatus was assembled as described by Southern (1975). Transfer was essentially complete after 16 hr, after which time the filter was removed, rinsed briefly with 100 ml 2 x SSC and baked in a vacuum oven at 80°C for 90 minutes. Filters were stored at 4°C between 2 sheets of Whatman 3 MM paper and sealed in polythene bags. Re-staining the compressed gel after transfer revealed that approximately 100% of DNA fragments less than 5 KB had left the gel, but that the efficiency of transfer declined considerably for larger fragments, probably to <10% for 20 KB molecules. To overcome this problem when studying larger fragments, gels were pre-soaked in 500 ml of 0.25 M HCl for 15 minutes prior to denaturation, to reduce fragment sizes by partial depurination (Wahl et al, 1979).

2.2.6 ^{32}P-labelling of DNA

2.2.6.1 Buffers

1 x nick translation buffer : 52.5 mM Tris, 5.25 mM $MgCl_2$, 20 µM each d(A,G,C and T)TP, 35 mM β-mercaptoethanol, pH adjusted to 7.5 with 1 M HCl. Stored in small aliquots at -20°C.

DNAse I dilution buffer : 45 mM $(NH_4)_2SO_4$, 22.5 mM Tris, 0.045% w/v BSA, 4.5 mM β-mercaptoethanol, 50% v/v glycerol, pH adjusted to 7.5 with 1 M HCl.

2.3 x reverse transcriptase buffer : 80 mM Tris, 13.9 mM $MgCl_2$, 83 µM each d(A,G,C and T)TP, 111.1 mM NaCl, 11 mM DTT, pH adjusted to 8.1 with 1 M HCl.

C-buffer : 5.2 mM Tris, 0.1 mM EDTA, 74 µM each d(A,G and T)TP, pH adjusted to 8 with 1 M HCl.

2.2.6.2 Nick translation

Double stranded DNA species were labelled by nick-translation (Maniatis et al, 1975; Rigby et al, 1977) under the following conditions:

0.5-1 µg of DNA in 1-2 µl of TE80 was mixed, in order, with:

(i) 1 x nick translation buffer to a total final volume of 33 µl.

(ii) 0.5 units DNA polymerase I.

(iii) 1 µl of 2 x 10^{-5} mg/ml DNAse I, recently diluted in DNAse I diluting buffer.

(iv) 10-20 µCi ^{32}P-dCTP.

Components were mixed and centrifuged briefly to start the nick translation reaction, which was at 15°C for 90 minutes. The reaction was terminated by the addition of 50 µl TE80 and by placing on ice.

Single stranded M13 clones were also found to label by nick translation under these conditions. Presumably, single stranded nicks in internally base paired DNA provide suitable 5' priming ends for the template-dependent 5'→3' polymerase and exonuclease activities of E.coli DNA Polymerase I (Kelley and Stamp, 1979).

2.2.6.3 Second strand synthesis

Single stranded DNA clones in phage M13 were labelled using reverse transcriptase by a modification of the 'random primer' technique of Taylor et al, 1976.

8 µl of single-stranded template DNA solution, prepared as in section 2.2.11.7, was annealed with 1 µl New England Biolabs M13 primer pentadecamer as described in section 2.2.12.2.1. After 45 min at 60°C, tubes were cooled over 15 minutes to 42°C then 10 µl of 2.3 x reverse transcriptase buffer, 10 µCi ^{32}P-dCTP and 10 units reverse transcriptase were added. Tubes were incubated for a further 2 hrs at 42°C then the reactions were terminated with 60 µl TE80 and by placing on ice.

When probes of particularly high specific activity were required, a modification of the sequencing reaction (section 2.2.12 and Sanger et al, 1977) was used to label single stranded clones with the Klenow fragment of E.coli DNA polymerase I. 8 µl of DNA was annealed with primer in the normal way and cooled to room temperature. 9 µl of C-buffer and 10 µCi ^{32}P-dCTP were added and the reaction initiated by the addition of 2.5 units of Klenow fragment diluted to 10 µl with 10 mM Tris HCl pH 8. After 1 hr at room temperature, the reaction was terminated by the addition of 50 µl TE80 and chilling on ice.

2.2.6.4 Separation of labelled DNA from unincorporated nucleotides.

^{32}P-DNA labelled by nick-translation or second strand synthesis was passed through a sterile Sephadex G-50 (fine) column, equilibrated with TE80, to remove unincorporated nucleotides. Columns were set up in 1 ml disposable hypodermic syringes plugged with glass wool. The syringes were filled with Sephadex slurry then allowed to settle and drain. Columns were dried by centrifugation at 500 x g_{ave} in

a Baird and Tatlock bench centrifuge for 1½ minutes, by which time they had shrunk to about 0.8 ml. The labelled DNA mixture, usually in 80-90 µl, was applied directly to the top of the column and the eluate, containing the purified probe, was collected in a 1.5 ml Sarstedt tube as the column was re-centrifuged at 750 x g_{ave} for 2 minutes. ^{32}P-labelled DNA probes were stored at -20°C until required.

The probe was largely free of unincorporated nucleotides since if ^{32}P-dCTP was applied to a column and treated as described above, <<1% of the label appeared in the eluate. Furthermore, applying a mixture of Orange G and Blue Dextran dyes in TE80 demonstrated that 80-90% of the labelled DNA, which co-migrates with the Blue Dextran, was recovered.

2.2.6.5 Estimation of DNA labelling efficiency.

Cerenkov light emitted by ^{32}P in aqueous solution (see Marshall, 1952, for review) was detected on the ^{3}H channel of an Intertechnique SL-3000 liquid scintillation counter. Observed counts were converted to dpm using the data in figure 2.1. and an estimate of incorporation of label into the probe could be made knowing the age of the isotope used, and assuming that 1 µCi = 2.2×10^6 dpm (by definition).

(Correction factor for decay of radioisotope = $2^{-t/t_{½}}$,

where t = time after activity date in days

and $t_{½}$ = half-life of isotope in days

= 14.31 days for ^{32}P)

This method was quicker than acid precipitation of DNA, required less handling of radiosiotope, and involved no loss of sample. Using this method, incorporation of ^{32}P into DNA by nick translation, second strand synthesis using reverse transcriptase and second strand synthesis using Klenow was estimated at 20-40, 20-60 and 80-90% respectively.

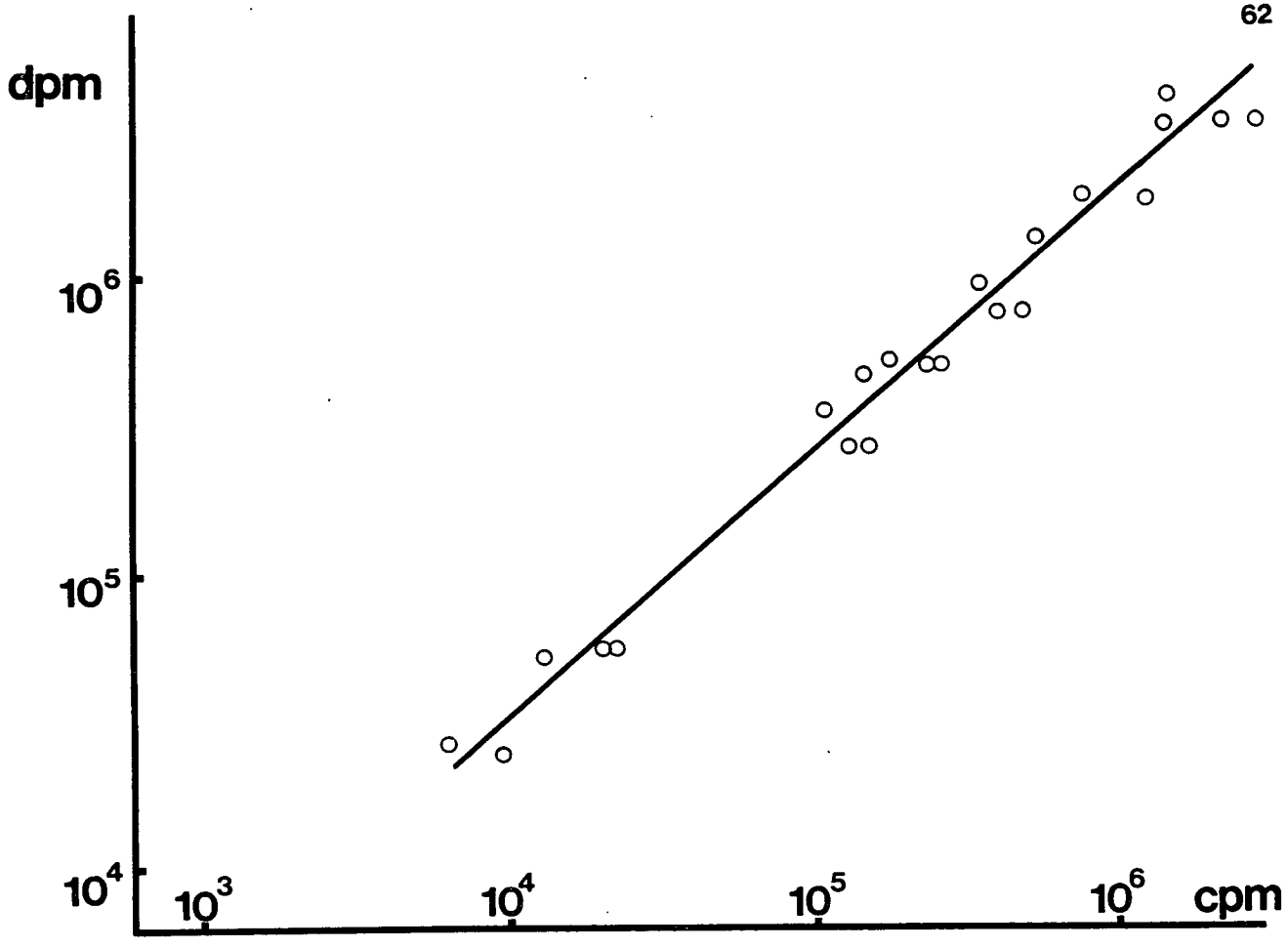


Fig. 2.1 Conversion of observed Cerenkov cpm on the ^{3}H channel of an Intertechnique SL-3000 liquid scintillation counter to dpm in the sample.

1 µl of α^{32}P-dCTP, 10 µCi/µl 6 days after this experiment was carried out, was sequentially diluted in TE80 and 80 µl aliquots counted in polypropylene Sarstedt tubes suspended in a glass scintillation vial. All measurements of radioactivity in DNA samples were carried out under similar conditions. Counting efficiencies in a separate experiment (not shown) were demonstrated to be largely volume-independent in the range 3.5 µl-1 ml.

2.2.7 Probing DNA immobilised on nitrocellulose

'Southern' blots of DNA prepared as described in section 2.2.5 were hybridised with appropriate ^{32}P-labelled DNA probes (section 2.2.6) under conditions determined by the expected degree of sequence homology between the probe and the blotted DNA. The conditions described below were experimentally determined to optimise signal:noise ratios when using both exactly and inexactly homologous probes.

2.2.7.1 Homologous probes

All stages except the final washes were carried out in heat-sealed polythene bags placed on a rocking table in an oven (Laboratory Thermal Equipment) of appropriate temperature.

2.2.7.1.1 Pre-hybridisation. Nitrocellulose filters were pre-hybridised (Denhardt, 1966) for 1-3 hr in 4 x SSC, 1 x Denhardt's solution, 50% v/v formamide at 37°C. 2 ml buffer/cm^2 of filter was used.

2.2.7.1.2 Hybridisation. Pre-hybridisation buffer was discarded and replaced with 0.1 ml/cm^2 of the same buffer to which 0.5-2 µg thermally denatured (5 minutes at 100°C and snap cooled on ice) probe (10^5-10^7 dpm) and 25 µg/ml sonicated herring sperm (average size 200 BP) DNAs were added. Hybridisations were carried out at 37°C, which for most higher plants is 26°C below the melting temperature (Tm) of the mt DNA, as calculated from the formula:

$$T_m\ (°C) = 81.5 + 0.41\ (GC) - 0.72\ F + 16.6\ \log_{10} M$$

(Marmur and Doty, 1962; McConaughy et al, 1969; Howley et al, 1979)

where: (GC) is the % G + C content of the DNA (about 47% for the mt DNA of most higher plants; calculated from Fox and Leaver, 1981; also from Wells and Ingle, 1970 and Schildkraut et al, 1962);

F is the % formamide in the hybridisation buffer;

and M is the molar concentration of univalent cations in the hybridisation buffer (= 0.78 for 4 x SSC).

16-20 hr hybridisations were found to be adequate to yield a strong signal under these conditions.

2.2.7.1.3 Washing. Following hybridisation, the solution containing the probe was removed from the bag and filter was washed 3 times for 30 minutes each in 1 ml/cm^2 of the pre-hybridisation buffer, at 37°C.

The probe was stored at -20°C and could be re-used up to 3 times if re-heated to 100°C for 5 minutes before use. The filter was then transferred to a large tray and washed 4 times for 5 minutes each in 500 ml of 2 x SSC, irrespective of filter size, at room temperature (20°C) with gentle agitation. Finally, the filter was blotted dry to remove most moisture, and autoradiographed as described in section 2.2.7.3. Under these conditions, the final wash is at T_m - 38°C.

2.2.7.2 "Heterologous" probes

When a DNA probe hybridises to a complementary DNA strand with which it does not have exact homology, a heteroduplex of reduced stability is established. Bonner et al (1973) have shown that a 1% increase in basepair mismatching lowers the T_m of this duplex molecule by 1°C. Hence when 'heterologous' probes, for example from mammals or fungi, were hybridised to maize mt DNA, the hybridisation and washing stringencies were lowered to allow stable heteroduplexes to form. The following procedure was found to work well for most heterologous probes:

2.2.7.2.1 Pre-hybridisation. Nitrocellulose filters were pre-hybridised in heat sealed polythene bags, at 52°C, for 2 hr in 1 ml/cm^2 of 10 x Denhardt's solution, 0.1% w/v SDS, 4 x SSC and 100 µg/ml boiled and sonicated herring sperm DNA.

2.2.7.2.2 Hybridisation. Pre-hybridised filters were rolled and transferred to glass scintillation vials containing 0.1 ml/cm^2 of pre-hybridisation buffer to which 2 x 10^6 - 2 x 10^7 dpm of thermally denatured probe had been added. Hybridisation was at 52°C for 48-72 hr, with gentle agitation. These conditions (47°C below the melting temperature of a maize mt DNA homoduplex) were empirically determined to maximise signal strength and minimise spurious

hybridisation of the probe to non-homologous DNA fragments. When probes of low G + C content, e.g. from S.cerevisiae mt DNA (Borst and Grivell, 1978) were used, the temperature of hybridisation and initial washes was reduced to 47°C, to accommodate the reduced stability of heteroduplexes formed with these probes (Marmur and Doty, 1962). In some hybridisations, 5% w/v dextran sulphate was incorporated in the buffer to reduce reaction volumes (Wahl et al, 1979), but this also led to higher 'background' hybridisation so its use was discontinued.

2.2.7.2.3 Washing. Non-hybridised probe was removed by washing, at 52 or 47°C, 3 times for 15 minutes each in 2 ml/cm^2 of 4 x SSC, 0.1% w/v SDS. Probe-containing buffers were stored at -20°C for re-use. Filters were washed in a tightly sealed plastic box, then blotted dry and autoradiographed as described in section 2.2.7.3.

2.2.7.3 Autoradiography

^{32}P labelled probes which had hybridised to blotted DNA were detected and visualised using X-ray film as described below.

Blots were mounted on cardboard, wrapped in plastic film (Dow 'Saran-Wrap' or Gelpack 'Cling-film') and placed in contact with Dupont 'Cronex' X-ray film in a metal cassette. When weak signals were to be detected, the film was pre-flashed (Laskey and Mills, 1975) and fluorographed at -80°C (Randerath, 1970) using Dupont 'Cronex' Lightning-Plus CF intensifying screens. Stronger signals were detected by autoradiography at room temperature. Exposure times were 3 hr to 30 days, depending on the intensity of the hybridisation. Films were developed as recommended by the manufacturer.

2.2.8 Transcript analysis by 'Northern' blotting.

This work was carried out in collaboration with V.P. Jones.

2.2.8.1 mt RNA preparation, electrophoresis and transfer to nitrocellulose.

Mitochondrial RNA was isolated from 4-day old etiolated maize coleoptiles by the method of Koller et al (1982). DNA was removed with RNAse-free DNAse I according to Smith and Ellis (1981) and mt RNA was electrophoresed through denaturing 1.3% formaldehyde-agarose gels essentially as described in Lehrach et al (1977). RNA was transferred immediately to nitrocellulose without staining, by the method of Thomas (1980). Molecular weight markers were provided by Tobacco Mosaic Virus RNA (6395 bases; Goelet et al, 1982), total E.coli RNA (23 S rRNA = 2904 bases, Brosius et al, 1980; 16 S rRNA = 1541 bases, Brosius et al, 1978; 5 S rRNA = 120 bases, Brownlee et al, 1967) and Cowpea Chlorotic Mottle Virus RNA (approximately 3200, 2900, 2300 and 900 bases, Davies and Verduin, 1979) which were electrophoresed in tracks parallel to the mt RNA but were not blotted. These RNAs were visualised under short wave (254 nm) uv light after staining for 30 minutes in 33 µg/ml acridine orange (McMaster and Carmichael, 1977) in 10 mM Na Pi buffer, pH 6.8.

2.2.8.2 DNA:RNA hybridisations.

Nitrocellulose RNA blots were pre-hybridised for 4 hr at 42°C in 0.25 ml/cm^2 of a buffer containing: 50% v/v formamide, 5 x Denhardt's solution, 0.1% w/v SDS, 5 x SSC and 200 µg/ml sonicated and denatured (see section 2.2.7.1.2) herring sperm DNA. After pre-hybridisation, 10^6 to 10^7 dpm of ^{32}P-labelled and denatured (section 2.2.7.1.2) probe DNA was added and hybridation continued for 16 hr at 42°C in the same buffer. Both hybridisation and pre-hybridisation were

performed in glass scintillation vials which were gently rocked in an oven of appropriate temperature.

Filters were washed twice for 10 minutes each in 250 ml of 2 x SSC 0.1% w/v SDS at room temperature then twice for 10 minutes each in 250 ml of 0.1 x SSC, 0.1% SDS at room temperature. After blotting the filters dry, hybridisation was detected by fluorography as described in section 2.2.7.3.

2.2.9 Cloning DNA restriction enzyme fragments into plasmid vectors

Note. All manipulations in sections 2.2.9, 2.2.10, 2.2.11, 2.2.13 and 2.2.14 involving in vitro recombined DNA were carried out in accordance with current GMAG guidelines.

2.2.9.1 Buffers and antibiotics

10 x CIP buffer : 0.1 M Tris, pH adjusted to 8.0 with 1 M HCl.

10 x ligase buffer : 0.66 M Tris, 0.1 M $MgCl_2$, 0.1 M DTT, 1 mM ATP, 10 mM EDTA, pH adjusted to 7.2 with 1 M HCl. Stored at -20°C and discarded after 20-30 days.

Antibiotics : Ampicillin (Ap): 100 mg/ml in H_2O, stored at -20°C.

Tetracycline (Tc): 5 mg/ml in 95% ethanol, stored at -20°C.

Chloramphenicol (Cm): Stored dessicated at 4°C and used directly.

2.2.9.2 Vectors

The pMB1-derived plasmid, pBR328 (Soberon et al, 1980), was chosen for this work because it was the most generally versatile cloning vector available, containing EcoRI, BamHI and HindIII sites in phenotypically recognisable drug resistance genes (Soberon et al, 1980; Bolivar et al, 1977; Kleckner et al, 1977).

Occasionally when using this vector, clones arose having recombinant phenotype due to intramolecular recombination rather than to insertion of foreign DNA. These may be a consequence of the large size (4.9 KB) of the vector or the 482 BP inverted duplication (Prentki et al, 1981) which it contains. The smaller vector, pACYC 184 (Chang and Cohen, 1978) was sometimes used to overcome this problem.

pBR328 DNA was a gift of F. Bolivar and was used to transform E.coli HB101 (section 2.2.9.7). A single Ap^r, Tc^r, Cm^r colony was used to initiate a bulk preparation of plasmid DNA (section 2.2.10.2). Fidelity of the vector was checked by single and double digestion with a variety of restriction endonucleases and comparison of fragment sizes with those predicted from the restriction map (Soberon et al, 1980).

2.2.9.3 Preparation of vector DNA for ligation.

2.2.9.3.1 Restriction enzyme digestion. 0.5-2.0 µg of vector DNA was digested to completion with a 2 to 5-fold excess of the appropriate restriction enzyme for 1-3 hr. An assessment of the extent of digestion was made by electrophoresis of small aliquots of treated and untreated DNA in parallel tracks of a minigel. Disappearance of all bands comigrating with the closed circular and supercoiled DNA bands was taken to indicate complete digestion. The enzyme was then inactivated by heat treatment (section 2.2.2.2) and terminal 5' phosphates removed from the vector DNA with CIP to prevent self-ligation (Chaconas and Van de Sande, 1980).

2.2.9.3.2 Phosphatase treatment 0.1 volumes of 10 x CIP buffer and 3 units of CIP/µg DNA was added to the restricted vector and dephosphorylation was allowed to proceed for 30 minutes at 37°C. The reaction was terminated by heating to 65°C for 10 minutes and residual CIP activity was removed by two rounds of phenol extraction. DNA was

precipitated with ethanol in the normal manner and the pellet washed with 0.5 ml of 70% ethanol. The pellet was resuspended in H_2O, re-precipitated with ethanol, dessicated and dissolved in TE80 to a final concentration of 50-100 ng/µl. CIP treated vector was stored at -20°C for up to 28 days before use with little loss of ligation efficiency.

2.2.9.4 Preparation of foreign DNA for ligation.

DNA was digested to completion with the appropriate restriction enzyme(s) and heat inactivated as described in section 2.2.2.2. A small aliquot was electrophoresed on a minigel to assess DNA concentration by comparison with known masses of bacteriophage λ DNA (section 2.2.14).

2.2.9.5 Ligation conditions.

100 ng of vector DNA was ligated with an amount of prepared foreign DNA calculated to contain an equal number of compatible restriction fragments, in a final volume of 10-20 µl (Dugiaczyk *et al*, 1975). For mtDNA treated with a hexanucleotide-recognising enzyme which cuts on average every 4096 BP, cloned into pBR328 (4900 BP), (4096/4900) x 100 = 83 ng was used:

Ligation mixtures contained:

1-4 µl DNA (as appropriate)
2 µl 10 x ligase buffer
1 unit T4 DNA ligage
H_2O to 20 µl;

and the reaction was allowed to proceed at 4-8°C for 16 hrs.

2.2.9.6 Bacterial host and competent cells.

Escherichia coli strain HB101 (F^-, *hsdS20* (r_B^-, m_B^-), *recA13*, *ara-14*, *proA2*, *lacY1*, *galK2*, *rpsL20* (Sm^r), *xyl-5*, *mtl-1*, *supE44*, λ^-) (Bolivar

and Backman, 1979), henceforth referred to as HB101, was used for transformation by, and propagation of, all plasmid clones. Cells competent for transformation by plasmid DNA were produced by a modification (Dagert and Ehrlich, 1979) of the standard $CaCl_2$ procedure (Cohen et al, 1973) as described below.

HB101 was streaked onto an L-agar plate and a well isolated colony used to inoculate a 5 ml L-broth culture which was grown to stationary phase with vigorous shaking at 37°C, for 10-16 hr (an 'overnight'). 0.2 ml of this was used to inoculate a 50 ml L-broth culture which was grown at 37°C with shaking until A_{600} = 0.2. At the correct density, cells were chilled on ice for 10 minutes and divided into two sterile 24 ml McCartney bottles. Cells were pelleted at 1300 xg_{ave} at 4°C, in an MSE Mistral 4L centrifuge, fitted with a 62303 head, for 5 minutes, then resuspended in 0.4 ml/ml of starting culture of 0.1 M $CaCl_2$ at 0°C. After 20 minutes, cells were centrifuged as before and resuspended in 10 µl/ml of starting culture of 0.1 M $CaCl_2$ at 0°C. Cells were assumed to reach maximum transformability 24 hr after the second resuspension in $CaCl_2$ (Dagert and Ehrlich, 1979).

2.2.9.7 Transformation of competent HB101.

The procedure used for transformation was essentially as described in Dagert and Ehrlich (1979); see also Humphreys et al (1979) for review.

A 5 µl aliquot of the ligation mixture was added to 100 µl of competent cells in a sterile 6 ml metal capped glass tube then placed on ice for 15 minutes. The remainder of the ligation mix was stored at -20°C and it was found that this could be used as a substrate for transformation with little reduction in efficiency up to 14 days after ligation. Cells were heat-shocked (37°C, 5 minutes) to effect transformation then incubated in 2 ml of L-broth at 37°C for 1 hr to allow

recovery and expression of antibiotic resistance genes. 1, 10 and 200 µl aliquots of this mixture were plated onto 85 mm diameter L-agar plates containing the appropriate antibiotics to select for transformants. For pBR328 clones, these were:

Cloning site	Transformant phenotype	Selection antibiotic	Plate concentration
EcoR1	Ap^r Cm^s Tc^r	Tetracycline	10 µg/ml
HindIII	Ap^r Cm^r Tc^s	Ampicillin	100 µg/ml
BamHI	Ap^r Cm^r Tc^s	Ampicillin	100 µg/ml
EcoR1-HindIII	Ap^r Cm^s Tc^s	Ampicillin	100 µg/ml

Transformation efficiencies Ap^rTc^r plasmids were found to be enhanced approximately 10-fold (from 9×10^5 transformants/µg to 8×10^6 when using pBR328, and from 2×10^4/µg to 1×10^5 when using ligated DNA) when initially selected on Tc rather than Ap plates. All Tc^r clones were later shown to be Ap^r by replica plating onto L-agar plates containing 100 µg/ml Ap. The reasons for this discrepancy are unclear but may be related to a lag phase in the expression of Ap^r, due to a secreted enzyme (Ambler and Scott, 1978) compared with Tc^r which is conferred by changes in the bacterial membrane preventing transport of the drug (Chopra and Howe, 1978).

Transformant colonies were transferred by replica plating with a sterile velvet pad, or by picking with toothpicks, onto appropriate media to select for recombinant clones. When the vector DNA had been treated with CIP, approximately 90-95% of all transformants were recombinant. When CIP was not used, the fraction was 10-15%.

2.2.10 Isolation and purification of plasmid DNA

2.2.10.1 Plasmid minipreps

Plasmid DNA was prepared simultaneously from up to 40 clones by the rapid alkaline lysis method devised by Birnboim and Doly (1979) and outlined below.

2.2.10.1.1 Buffers

Lysis buffer : 25 mM Tris, 10 mM EDTA, 50 mM glucose, 0.2. w/v lysozyme, pH adjusted to 8.0 with 1 M HCl.

Alkaline SDS : 0.2 M NaOH, 1% w/v SDS

High salt solution : 3 M Na Acetate, 30 µg/ml yeast tRNA, pH adjusted to 4.8 with glacial acetic acid.

2.2.10.1.2 Method. Six clones of interest were streaked onto sectors of an 85 mm L-agar plate containing 100 µg/ml Ap and grown, inverted, for 16 hr at 37°C. Single colonies from these streaks were used to initiate 5 ml 'overnight' cultures which were supplemented with 20 µg/ml Ap. 0.5 ml aliquots of these cultures were transferred to 1.5 ml polypropylene tubes (Sarstedt) and centrifuged at 12000 x g_{max} for 15 seconds to pellet cells which were then resuspended in 100 µl of freshly prepared lysis buffer. Cells were incubated on ice for 30 minutes to digest partially the bacterial cell walls, then were lysed by the addition of 200 µl alkaline SDS. After a further 5 minutes at 0°C, 150 µl of high salt solution was added and the mixture incubated at 0°C for 1 hr. Under these conditions, chromosomal DNA was selectively denatured and precipitated, leaving supercoiled plasmid DNA in solution.

The lysate was cleared by centrifugation at 12000 x g_{max} (room temperature) for 5 minutes then 400 µl of supernatant was removed and nucleic acids were precipitated by the addition of 1 ml ethanol and chilling at -20°C for 30 minutes. The suspension was centrifuged at 12000 x g_{max} and the supernatant removed with a drawn out pasteur pipette. The pellet was dissolved in 100 µl of 0.1 M sodium acetate (pH 6) and nucleic acids precipitated a second time with 200 µl ethanol after chilling to -20°C for 10 minutes. The pellet was collected by centrifugation as above for 2 minutes, dessicated, then dissolved in 10 µl of TE80.

Minigel electrophoresis revealed yields of plasmid DNA to be 1-2 µg from 0.5 ml of 'overnight' culture. The DNA was a suitable substrate for all restriction enzymes but if it was to be used for nick-translation or ligation, it was further purified by RNAse treatment as follows:

DNA (1-2 µg) was diluted to 100 µl with TE80 containing 10 mM NaCl and heat treated (100°C, 15 minutes). RNAse was added to a final concentration of 50 µg/ml. Digestion was for 4 hr at 37°C and was terminated by the addition of NaCl to 100 mM. The solution was subjected to two rounds of phenol extraction and ethanol precipitation then the DNA pellet was washed by resuspension in 70% ethanol and finally dissolved in 10 µl of TE80.

2.2.10.2 Large scale plasmid preparation.

Plasmid DNA was prepared on a large scale by a modification (Hu and Messing, 1982) of the detergent lysis method (Godson and Vapnek, 1973) as outlined below. This method was found to be consistently more reliable and higher yielding than the lysis by boiling method of Holmes and Quigley (1981).

2.2.10.2.1 Buffers

Sucrose buffer : 50 mM Tris, 40 mM EDTA, 25% w/v sucrose, pH adjusted to 8.1 with 1 M HCl.

Triton mix : 50 mM Tris, 62.5 mM EDTA, 0.1% v/v Triton X-100, pH adjusted to 8.1 with 1 M HCl.

2.2.10.2.2 Method. Single colonies of the desired clone were used to inoculate 5 ml 'overnight' cultures containing 20 µg/ml Ap. 2.5 ml of this culture was diluted into 250 ml of L-broth supplemented with 0.2% w/v glucose and 20 µg/ml Ap. A 10-fold excess void space was included in the culture vessel to allow adequate aeration of the

culture, which was grown at 37°C in a New Brunswick orbital shaker (200-300 rpm) until A_{590} = 1. Cm (solid) was then added to a final concentration of 150 µg/ml (Clewell and Helinski, 1972). When Cm^r clones were to be amplified, Spectinomycin was added at 300 µg/ml at this stage. Cultures were incubated for a further 16 hr at 37°C to allow plasmid amplification, then cells were harvested by centrifugation at 4000 x $\underline{g}_{ave}$ for 15 minutes in a Sorvall GSA rotor (4°C).

Cells were resuspended in 6 ml (total) of sucrose buffer then 1 ml of 20 mg/ml lysozyme freshly dissolved in the same buffer and 1 ml of 0.5 M EDTA (pH 8.1) were added. After 15 minutes digestion at 0°C, 13 ml of triton mix was added and the preparation allowed to stand a further 15 minutes at 0°C. Cell lysis had normally occurred by this stage and the mixture was extremely viscous. Occasionally, lysis did not occur immediately so the cells were warmed to room temperature and a few drops of 10% v/v Triton X-100 were added.

The lysate was cleared by centrifugation at 48,000 x $\underline{g}_{ave}$ for 45 minutes at 4°C. The supernatant was decanted into a measuring cylinder and 0.95 g CsCl and 0.1 ml 5 mg/ml ethidium bromide added /ml of cleared lysate. The lysate was transferred to two 13.5 ml cellulose nitrate or polyallomer tubes and centrifuged to equilibrium (Radloff et al, 1967) for 65 hr at 90,000 x g_{ave} in a Kontron TFT 65.13 rotor (18°C). Following centrifugation, the gradients were examined under 366 nm uv illumination (UV Products Inc) and the lower fluorescent band, containing covalently closed circular plasmid DNA, was removed, and ethidium bromide extracted, as described in section 2.2.1.4. Following ethanol precipitation, plasmid DNA was dissolved in 100-200 µl of TE80 and yields and purity estimated from A_{320} to A_{200} scans. Yields normally varied from 250 µg for small clones (inserts <3 KB) to 10-20 µg for large clones (20-30 KB inserts) from a 250 ml starting culture.

2.2.10.3 Preparation of cosmid DNA

Cosmid DNA was prepared from 100 ml stationary-phase cultures exactly as described in Lonsdale et al (1981).

2.2.11 Cloning DNA restriction fragments into bacteriophage M13

2.2.11.1 Buffers and indicator solutions

LTB : 20 mM Tris, 20 mM NaCl, 1 mM EDTA, pH adjusted to 7.9 with 11.7 M HCl.

PEG/NaCl : 2.5 M NaCl, 20% w/v Polyethylene glycol (MW = 6000).

T(0.1)E80 : 10 mM Tris, 0.1 mM EDTA, pH adjusted to 8.0 with 11.7 M HCl.

XGal : 20 mg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactoside in dimethyl formamide. Stored at -20°C.

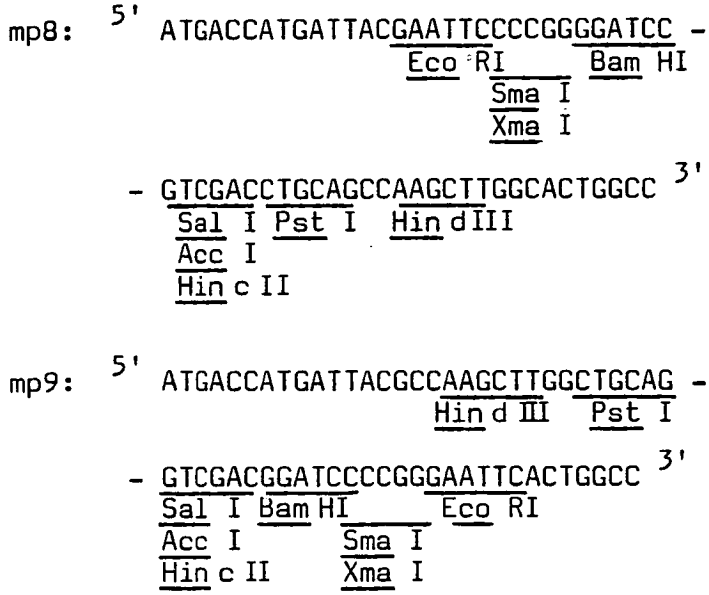
IPTG : 24 mg/ml isopropyl-thio-β-D-galactoside in dimethyl formamide. Stored at -20°C.

2.2.11.2 Vectors

DNA fragments were cloned into the replicative forms (RF) of the M13 vectors mp8 and mp9 (Messing and Vieira, 1982) in order to generate single stranded template DNA (Messing et al, 1977) for sequence analysis by the chain termination reaction (Sanger et al, 1977). An additional use was for the construction of clones having no homology to pBR328, which when labelled by second strand synthesis (section 2.2.6.3), were useful for probing libraries of pBR-derived clones.

mp9 RF DNA was a gift of K. Jones, Genetics Department, University of Edinburgh and mp8 phage was a gift of D. Finnegan, Department of Molecular Biology, University of Edinburgh. mp8 RF DNA was prepared from the phage as described in section 2.2.13 and the fidelity of the cloning vector checked by sequence analysis of the polyclonal linker (fig. 2.3) and also by comparison of the fragments generated by various restriction enzymes with the published restriction map (van Wezenbeck et al, 1980; Messing et al, 1977; Messing and Vieira, 1982).

Fig 2.2 DNA sequences containing the polyclonal linkers in M13 mp8 and mp9
From Messing and Vieira (1982)



2.2.11.3 Host *E.coli* strain

The F^+ *E.coli* strain JM101 (Δ(*lac pro*), *thi*, *supE*, F'*traD36*, *proAB*, *lac* i^q, *z*Δ*M15*; Messing, 1979), referred to as JM101, was used to propagate and maintain M13 clones. In order to maintain the F-plasmid required for M13 infection (Pratt *et al*, 1966), bacteria were plated on minimal medium plates lacking proline (section 2.1.5). Since some of the genes involved in proline biosynthesis have been transferred to the F-plasmid from the chromosome of JM101, cells lacking the plasmid were unable to grow. The *z*Δ*M15* deletion from the episomal β-galactosidase *z* gene was used as the basis of an assay for infection by mp8 or 9 since a functional β-galactosidase complex could be formed between the *lac z*-α peptide encoded in the phage (Messing *et al*, 1977) and the ω peptide encoded in the F-plasmid. In the presence of the inducer (IPTG), the chromogenic substrate XGal was hydrolysed by this complex to yield a blue dye. JM101 cells, whose growth was retarded by infection with M13 mp8 or mp9 (Hofschneider and Preuss, 1963), formed turbid plaques on a lawn of uniformly growing *E.coli* and were readily detected by their blue colour. Plaques formed after infection with recombinant M13 clones were colourless (or 'white') because the complementation ability of the M13 α peptide was destroyed by insertion of foreign DNA into the polyclonal linker, which lies near the amino terminal-encoding region of this gene (Messing and Vieira, 1982).

2.2.11.4 Restriction enzyme digestion and ligation of DNA

M13 RF vector DNA was prepared for ligation as described for plasmid DNA (section 2.2.9.3) but was not treated with CIP since non-recombinant clones were readily detected by the colour reaction. Foreign DNA was prepared as described in section 2.2.9.4. Small (<1 KB) foreign DNA restriction fragments were normally used for M13

cloning, since larger fragments were found to be heavily under-represented in subsequent recombinant M13 clones. The largest fragment successfully subcloned was ca. 4.0 KB, but all inserts greater than 1.5 KB resulted in retarded phage growth and extremely small white plaques. Small fragments were normally generated either by double digestion with hexanucleotide-recognising enzymes or with tetranucleotide-recognising enzymes. Fragments were ligated directly into the appropriate site(s) in the M13 polylinker except TaqI and MspI fragments which were cloned into the AccI site; Sau3A fragments which were cloned into BamHI site and AluI fragments which were cloned into the SmaI site.

150 ng of vector DNA was ligated with sufficient foreign DNA to contain a 3:1 molar excess of compatible foreign restriction fragments over vector. Assuming tetranucleotide enzymes cut every 256 BP on average and that linearised vector DNA is 7229 BP (Messing and Vieira, 1982) the mass of foreign DNA was (3 x 150 x 256)/7229 = 15.9 ng and similarly for hexanucleotide enzymes was 255 ng. When double digested fragments were cloned, the amount of vector DNA was doubled and for blunt-ended ligations both vector and foreign DNA concentrations were increased 10-fold.

Ligation mixtures contained:

1-2 µl DNA (as appropriate)

1 µl 10 x ligase buffer (section 2.2.9.1)

1-2 units T4 DNA ligase (2-4 units for blunt-ended ligations)

H_2O to 10 µl.

Ligations were allowed to proceed for 16 hr at 4-8°C.

2.2.11.5 Competent cells and transformation of JM101.

An 'overnight' (section 2.2.9.6) was grown from a single colony of JM101 which had been streaked out on a minimal agar plate. 1 ml of

this was used to inoculate a 100 ml L-broth culture which was grown to A_{600} = 0.2 in a 250 ml conical flask. Competent cells were prepared from this culture as described below and by Mandel and Higa (1970).

Cells were cooled on ice for 15 minutes then harvested by centrifugation at 1300 x g_{ave} and 4°C in 24 ml McCartney bottles, in an MSE Mistral 4L centrifuge. The cell pellet was resuspended in 0.5 ml/ml starting of culture of 50 mM $CaCl_2$ and after 20 minutes at 0°C cells were re-centrifuged as described above. This pellet was resuspended in 100 µl of ice cold 50 mM $CaCl_2$/ml of starting culture and cells were used for transformation (Humphreys et al, 1979) within 1-3 hrs of preparation.

Ligation mixtures (section 2.2.11.3) were diluted 10 or 100-fold with 50 mM $CaCl_2$ and 10 µl aliquots mixed with 200 µl competent cells in a sterile 6 ml metal capped glass tube. Transformations in plastic tubes were found to be less effective, probably because the heat shock (see below) is transmitted less efficiently. After 40 minutes on ice, cells and DNA were heated to 42°C for 2 minutes to effect transformation then allowed to cool to room temperature. 200 µl of plating cells (JM 101 diluted 1:100 into L-broth and grown at 37°C with shaking for 2-3 hours), 30 µl XGal and 20 µl IPTG were added and the mixture was plated onto 85 mm diameter minimal medium plates, using 3.5 ml of molten (45°C) BBL top-layer agar (section 2.1.5).

Plates were inverted and incubated at 37°C for 12-24 hr in order to visualise plaques. Normally, 0.5-1.5 x 10^6 plaques (100% blue) were obtained /µg of mp8 or mp9, and 10^4-10^5 plaques (10-20% white) were obtained /µg of vector when ligated to single-digested foreign DNA. The number of plaques fell approximately 500-fold (but were 90-100% white) when double digested vector was used, presumably

because the small fragment of polylinker excised between the two restriction sites was present in too low a concentration to allow efficient self ligation.

2.2.11.6 Plaque screening and purification.

2.2.11.6.1 Screening by plaque hybridisation. Plaques were normally transferred 'in situ' directly from the original transformation plates onto nitrocellulose by the Benton-Davis (Benton and Davis, 1977) technique (see below). Occasionally, when only small numbers of plaques were to be screened, an appropriate number of white plaques were picked from the transformation plate using a toothpick and transferred to a 100-place grid on a minimal plate, overlain with a top-layer containing mid-log phase JM 101. After overnight growth, this plate was then transferred to nitrocellulose in the same way. The main advantage of this method was easier subsequent identification of plaques of interest, because of the ordered grid used. However it suffered the disadvantages: (i) that only a few white plaques (500 maximum) could be reasonably screened; (ii) that very small plaques containing clones with large DNA inserts were discriminated against when picking visually; and (iii) that small in-frame insertions in the β-galactosidase α peptide do not always destroy its complementation activity (Gronenborn and Messing, 1978), and are therefore not recognised as recombinants by eye.

The abbreviated version of the Benton-Davis procedure used was as follows:

A circular piece of autoclaved (15 minutes, 121°C) nitrocellulose cut to fit the petri dish was lowered over the plate to be transferred, avoiding air bubbles. The filter was left in contact with the plate until it was uniformly damp and orientation marks had been made by stabbing with a sterile syringe needle. The nitrocellulose was peeled

off and transferred (plaque side upwards) to a pad of Whatman 3 MM paper soaked in denaturation buffer (section 2.2.5.1). DNA was allowed to denature for 2 minutes then the filter was immersed in neutralisation buffer (section 2.2.5.1) for 2-10 minutes. The filter was then rinsed in 2 x SSC, blotted dry and baked under vacuum at 80°C for 90 minutes. Filters could be stored indefinitely at 4°C after sealing in polythene bags. Up to 3 filter copies were made from each plate.

Filters were probed with ^{32}P-labelled DNA exactly as described in section 2.2.7.1. The probes used depended on the starting material for M13 cloning. Normally, for phage derived from sub-digestion of plasmid clones, one nitrocellulose copy from each plate was probed with the plasmid vector and another with the purified insert (section 2.2.4) after restriction with the appropriate enzyme(s). M13 clones containing vector, insert and vector-insert border derived fragments could therefore be identified.

Plaques hybridising to the appropriate probe(s) were identified by this means. If well separated from any other plaques, they were picked into 50 µl LTB and the same toothpick used to initiate a 1 ml phage miniprep (section 2.2.11.7). If contamination with other plaques was likely, the plaque of interest was picked into 50 µl of LTB and 200 µl of this phage suspension, after dilution 10^6-fold with LTB, was plated onto a minimal plate with 200 µl plating cells, 30 µl XGal, 20 µl IPTG and 3.5 ml BBL top layer agar, as described in section 2.2.11.4. After overnight incubation at 37°C, a single plaque well isolated from any others was picked and used to inoculate a phage miniprep (2.2.11.7).

2.2.11.6.2 Screening by restriction enzyme analysis. The bacterial cell pellet from the phage miniprep (section 2.2.11.7) could be used

as a substrate for an RF DNA miniprep by the method described in section 2.2.10.1. The DNA isolated by such minipreps could be treated with the appropriate restriction enzyme(s) and after gel electrophoresis (section 2.2.3), clones containing the desired insert size were readily identified. Estimates of the insert sizes in clones derived from tetranucleotide enzyme fragments cloned into a hexanucleotide enzyme site in M13 were made by double digestion with the hexanucleotide enzymes which recognised sites in the mp8 or mp9 polyclonal linker immediately adjacent to the cloning site used (e.g. BamHI + PstI for TaqI clones in the AccI site).

2.2.11.7 Phage and single-stranded (ss) DNA minipreps.

Phage particles were concentrated from the supernatant of an E.coli JM 101 culture infected with M13 using PEG 6000. ssDNA was isolated from these particles (Yamamoto et al, 1970; Schrieir and Cortese, 1979) and used as a template for DNA sequencing reactions (section 2.2.12) or for labelling by second strand synthesis (section 2.2.6.3). Up to 40 preparations were processed simultaneously by the following method:

Two ml of an 'overnight' culture of JM 101 was diluted into 50 ml of L-broth and grown in a 250 ml flask with shaking at 37°C to A_{650} = 0.2. One ml aliquots were dispensed into sterile 5 ml glass 'bijou' bottles and these were infected with M13 picked from a suitably purified plaque (section 2.2.11.5.1). The cultures were grown for 4¼ hr with vigorous shaking (250-300 rpm) in a New Brunswick orbital shaker at 37°C then cells were transferred to 1.5 ml polypropylene Sarstedt tubes and pelleted at 12000 x g_{max} for 5 minutes. Supernatants were poured into fresh tubes and 200 µl of PEG/NaCl added. After mixing and incubation at room temperature for 30 minutes, phage aggregates were collected by centrifugation at 12000 x g_{max} for

5 minutes. All traces of the PEG-containing supernatant were removed from the pellet, which was resuspended in 100 µl of T(0.1)E 80 and extracted with 50 µl of phenol (section 2.2.4.1). The aqueous phase was removed, made to 0.3 M in Na acetate (pH 5.5) then ssDNA precipitated at -80°C for 15 minutes with 2.5 volumes of ethanol.

DNA was pelleted for 10 minutes at 12000 x g_{max} then washed by resuspension in 1 ml of ethanol (chilled to -20°C) and re-centrifuged for 5 minutes. This DNA pellet was dessicated (in a Speedvac dessicator) then dissolved in 50 µl T(0.1)E 80 and precipitated again as above. After dessication, the final pellet of ssDNA (normally invisible) was dissolved in 50 µl of T(0.1)E 80 and stored at -20°C until required. 5 µl aliquots of the ssDNA preps were sometimes electrophoresed on minigels (section 2.2.3.3) and estimates of insert sizes made by comparison with the mobility of ss vector DNA.

2.2.12 DNA sequence analysis by the dideoxynucleotide chain termination reaction

The complementary strand elongation/termination method devised by Sanger et al (1977) for determination of DNA sequence from single stranded templates (Sanger et al, 1980) was used for all sequence analysis.

2.2.12.1 Buffers.

2.2.12.1.1 Termination mixes. These buffers contained d (A,G,C and T)TP with a base specific termination nucleotide (dd(A,G,C or T)TP). The ratios of dNTP/ddNTP varied from batch to batch of nucleotides and were re-optimised every time new stocks were acquired. For those used for most of this sequencing, the concentrations were:

Termination A : 5.22 mM Tris, 0.104 mM EDTA, 5.43 µM dATP, 109 µM dGTP, 109 µM dTTP, 75 µM ddATP, pH adjusted to 8.0 with 1 M HCl.

Termination G : 5.22 mM Tris, 0.104 mM EDTA, 109 µM dATP, 5.43 µM dGTP, 109 µM dTTP, 250 µM ddGTP, pH adjusted to 8.0 with 1 M HCl.

Termination C : 5.22 mM Tris, 0.104 mM EDTA, 74 µM dATP, 74 µM dGTP, 74 µM dTTP, 15 µM ddCTP, pH adjusted to 8.0 with 1 M HCl.

Termination T : 5.22 mM Tris, 0.104 mM EDTA, 109 µM dATP, 109 µM dGTP, 5.43 µM dTTP, 250 µM ddTTP, pH adjusted to 8.0 with 1 M HCl.

Termination mixes were stored in 10 µl aliquots at -80°C and were stable for at least 6 months. The dCTP concentration in each mix was brought to 1.16 µM by the addition of the α ^{32}P dCTP immediately before the sequencing reaction (section 2.2.12.2.2) was initiated.

2.2.12.1.2 Others

Primer : 2.5 µg/ml New England Biolabs' Synthetic Pentadecamer Primer ($^{5'}$TCCCAGTCACGACGT$^{3'}$, see also Sanger et al, 1980) in H_2O. Stored in 50 µl aliquots at -20°C.

10 x annealing buffer : 0.1 M Tris, 0.1 M $MgCl_2$, pH adjusted to 8.0 with 1 M HCl.

Chase buffer : 0.5 mM d(A,G,C and T)TP in H_2O. Stored at -20°C.

Sequencing gel loading buffer : 10 mM EDTA, 0.3% (w/v) bromophenol blue, 0.3% (w/v) xylene cyanol FF, 98% formamide. Stored at 4°C.

2.2.12.2 The sequencing reaction

2.2.12.2.1 Annealing. Eight µl of single stranded M13 template DNA (section 2.2.11.7) was mixed with 1 µl (2.5 ng) primer in 1 x annealing buffer in a 1.5 ml polypropylene Sarstedt tube. The tube was sealed tightly and heated to 60°C for 45 minutes in an oven then allowed to cool to room temperature over 15 minutes.

2.2.12.2.2 Chain termination. The annealed DNA was centrifuged briefly at 12000 x g_{max} to collect condensate then 2 µl aliquots transferred to the side of 4 x 1.5 ml Sarstedt tubes. 2 µl of DNA polymerase I (Klenow fragment), freshly diluted to 0.25 u/µl in 10 mM Tris HCl pH 8.0 was added to each tube in a separate spot from the DNA.

10 µl aliquots of the termination mixes were thawed and 0.5 µl (5 µCi) of α ^{32}P dCTP (section 2.1.4) was added to each one. When the ^{32}P was less than one-half life old, it was first diluted 1:1 with 40 µM dCTP. 2 µl of one termination mix, as appropriate, was added in a third spot to each tube which was then centrifuged briefly at 12000 x g_{max} to initiate the sequencing reaction.

2.2.12.2.3 Chase and denaturation of DNA. After 15 minutes at room temperature, 2 µl of chase buffer was added to each sequencing reaction, which were then centrifuged as above and allowed to stand a further 15 minutes at room temperature. Reactions were terminated by the addition of 4 µl loading buffer and DNA denatured by heating to 95°C followed by snap cooling on ice.

2.2.12.2.4 T-screening. When large numbers of randomly generated clones were to be sequenced, a prior screening was carried out by performing a T-termination reaction only, with each template. Identical clones and poor templates could therefore be identified before proceeding with the full sequencing reaction.

2.2.12.3 Sequencing gels.

Polyacrylamide gels were cast in moulds 400 mm long x 300 mm wide and 0.35 mm thick, and were used to separate ssDNA fragments generated by the sequencing reaction.

Sequencing gels contained:

5.7% w/v acrylamide

0.3% w/v NN' methylene bis acrylamide

0.06% w/v ammonium persulphate (AMPS)

0.079% v/v TEMED

8 M urea

in 1 x TBE (section 2.2.3.1).

Acrylamide and bis acrylamide were mixed and stored as a 10x concentrated stock in H_2O at -20°C. Acrylamide stored in solution at 4°C was found to deteriorate in 2-4 weeks leading to very poor resolution on gels. AMPS was made up as a 10% w/v stock in H_2O and discarded immediately after use. All components except TEMED were mixed thoroughly then warmed to room temperature to remove haze in the solution. After the addition of TEMED, gels were poured rapidly and a 49 or 85-tooth 0.35 mm comb was inserted to form loading wells. Gels normally set in 5-10 minutes but were allowed to polymerise fully for a further 1-2 hr before electrophoresis.

The gel was clamped to a vertical support with 800 ml of 1 x TBE in each tank. DNA samples from the sequencing reaction were applied using a drawn out melting point tube attached to a mouth aspirator, after washing out each well to disperse unpolymerised acrylamide and urea. Normally, only about ¼ of the sequencing mix (3 µl) was loaded since largervolumes were found to reduce resolution and cause overexposure of the autoradiograph.

DNA fragments were separated by electrophoresis at constant power (65 W) which maintained the gel temperature at about 50°C and therefore prevented self-annealing of DNA strands. The PD varied from 1100 to 1500 V and the current from 60 to 40 mA during this time. The run time for electrophoresis was determined by the vector cloning site used and the number of base pairs to be read. The bromophenol

blue dye was found to run just ahead of the HindIII site in mp8 and under these conditions reached the bottom of the gel in 2-2¼ hr. The EcoRI site reached the same point 30 minutes later. Sequences of 200-300 BP from the cloning site could be read from gels autoradiographed at this stage. After 7 hr electrophoresis, the first 150 BP of sequence was run off the bottom of the gel, but the readable sequence could be extended 150 BP beyond the end of the sequence from the short gel. Normally, a portion of the sample was loaded immediately and the gel run for 4½ hr when a further portion (after storage at -20°C) was loaded in an adjacent slot. Electrophoresis was continued for a further 2½ hr and the whole gel autoradiographed.

Following electrophoresis, plates were prized apart with a scalpel and the gel, still attached to one plate, immersed in 10% acetic acid for 15 minutes to fix the DNA, then rinsed briefly in distilled H_2O. The gel was blotted dry with paper towels then transferred to a large sheet of Whatman 3MM paper, covered in Saran Wrap, and dried on low heat with a Biowerk gel drier, for 1½ hr. Drying was found to increase the length of sequence which could be read from a single gel by at least 100 base pairs. Dried gels were exposed to 30 x 40 cm sheets of Fuji RX X-ray film in metal-backed folders at room temperature. Gels were covered with Saran Wrap for autoradiography because the drying process left a slightly tacky surface which otherwise adhered to the film. 12-16 hr exposures were normally sufficient to allow complete legibility of all bands.

2.2.13 Large scale preparation of bacteriophage M13 RF DNA

Large scale RF preparations of M13 clones were not normally necessary since sufficient double stranded DNA for most purposes could be made by the 'miniprep' method (section 2.2.11.6.2). However, a bulk preparation of mp8 RF DNA was carried out and this DNA was used

for subsequent M13 cloning. A method devised by D. Fisher (personal communication) was followed, as briefly described below, which employs standard techniques for the preparation of a high phage titre for infection of JM 101, followed by detergent lysis to release supercoiled RF DNA (see, for example, Tanaka and Weisblum, 1975, or Hines and Ray, 1980).

2.2.13.1 Buffers.

TE 85 : 50 mM Tris, 1 mM EDTA, pH adjusted to 8.5 with 1 M HCl.

TE 85 + sucrose : 50 mM Tris, 50 mM EDTA, 15% w/v sucrose, pH adjusted to 8.5 with 1 M HCl.

TE 85 + Triton : 50 mM Tris, 50 mM EDTA, 0.1% v/v Triton X-100, pH adjusted to 8.5 with 1 M HCl.

2.2.13.2 Growth of phage inoculum.

A stock of mp8 phage at 10^7 pfu/ml in LTB (section 2.2.11.1) was a gift of D. Finnegan, Department of Molecular Biology, University of Edinburgh. This was diluted 10^5-fold with LTB and 100 µl plated with 200 µl JM 101 plating cells, indicator and 3.5 ml BBL top layer agar (section 2.2.11.5). A single blue plaque was used to infect a 2 ml mid-log phase culture of JM 101 and this was grown up overnight at 37°C with shaking. This same blue plaque was used as a substrate for a single stranded DNA miniprep (section 2.2.11.7) and was sequenced (section 2.2.12) in order to verify the authenticity of the starting material.

The overnight culture was centrifuged at 12000 x g_{max} and 100, 10 and 1 µl aliquots of the supernatant (diluted 10^7-fold in L-broth) were plated in the normal manner to determine the titre. The remainder was stored at 4°C. Sufficient phage-containing supernatant was then used to infect a 50 ml early-mid log phase JM 101 culture (A_{600} = 0.3) at

a final concentration of 10^9 pfu/ml, which was grown for a further 16 hr at 37°C with shaking. The culture was clarified by centrifugation and the titre of the supernatant determined as previously.

2.2.13.3 Isolation of RF DNA.

A 500 ml culture of JM 101 was grown to OD_{600} = 1.0 then phage-containing supernatant (from section 2.2.13.2) was added to a final concentration of 10^{10} pfu/ml. The culture was grown for a further 2 hr at 37°C with vigorous shaking then cells were collected by centrifugation at 4,100 x g_{ave} for 15 minutes in a Sorvall GSA rotor at 4°C. Cells were resuspended in 50 ml ice-cold TE85, pelleted as before, and finally resuspended in a total volume of 18 ml TE85 + sucrose. 4.5 ml of lysozyme (5 mg/ml in TE85 + sucrose) was added and the mixture was incubated at 0°C for 10 minutes.

13.5 ml of TE85 + Triton was added and the temperature raised to 37°C for 15 minutes. The lysate was clarified by centrifugation at 77,000 x g_{ave} for 45 minutes in a Sorvall AH627 rotor at 4°C. To each ml of supernatant, 0.95 g CsCl and 59 µl of 10 mg/ml ethidium bromide was added then DNA was centrifuged to equilibrium in a Kontron TFT 65.13 rotor at 90,000 x g_{ave} for 45 hr at 15°C. The lower band containing supercoiled M13 RF DNA was removed as described in section 2.2.12.2.2. This DNA was diluted 4-fold in a solution containing 1.059 g/ml CsCl and 59 µl/ml of 10 mg/ml ethidium bromide and centrifuged to equilibrium as above but at 130,000 x g_{ave} and 20°C. The lower band was removed as before then DNA was purified and concentrated by organic extraction, dialysis and ethanol precipitation as described in section 2.2.10.2.2.

The yield of mp8 RF DNA was estimated to be 150 µg from a 500 ml starting culture, determined from A_{350}-A_{200} scans and by examination of restriction fragments stained with ethidium bromide after electrophoresis through 6% polyacrylamide gels (section 2.2.3.4).

2.2.14 Large scale preparation of bacteriophage λ DNA

DNA was isolated from phage λ carrying the cI857ts mutation (Lewin, 1977) and was used for the generation of molecular weight markers by restriction enzyme digestion. The fragment sizes produced by various enzymes are shown in Table 2.2.

Phage λ particles were a gift of M. Lockyer, Department of Molecular Biology, University of Edinburgh. These were used to infect a 1 l culture of E.coli C600 (Appleyard, 1954) at 2 x 10^8 pfu/ml. Phage particles were isolated from this culture and used to prepare λ DNA as described in Murray et al (1977) and Yamamoto et al (1970). The yield was 978 µg.

2.2.15 Isolation of petite yeast mtDNA

mtDNA was prepared from the cytoplasmic petite strains RP6 (Hensgens et al, 1979), LH30-A5 (Hensgens et al, 1983) and dD22 for use as probes for ATPase subunit 9, cytochrome oxidase subunitI/ ATPase subunit 6 and cytochrome b respectively. Yeast strains were gifts of P. Slonimski (RP6), L. Hensgens (LH30-A5) and T. Fox (dD22). After acquisition, RP6 was crossed (Deutsch et al, 1974) with the tester strain JS8-17 to check for retention of the O_1^R, marker as described by Slonimski and Tzagoloff (1976). LH30-A5 had been recently checked for its ability to restore wild-type phenotype to respiratory mutants in the S3 sector and was not re-verified.

Yeast cultures were grown as described in Sanders et al (1973) and protoplasts prepared using snail gut enzyme (Helicase, Industrie Biologique Francais; or Glusulase, Sigma) (Linnane and Lukins, 1975). Protoplasts were disrupted mechanically and mitochondria collected by differential centrifugation then lysed with detergent (4% w/v SDS, 2% w/v tri-iso propyl napthalene di sulphonate, 1% w/v para-amino salicylate, 6% v/v iso butanol). MtDNA was banded by either by

CsCl-DAPI (Fox, 1979) or NaI-ethidium bromide equilibrium gradient centrifugation. DNA was removed from gradients and purified as described in section 2.2.12.2.2. Digestion with restriction endonucleases and comparison of fragment sizes with published restriction maps confirmed the authenticity of the petite DNAs.

Table 2.2. Restriction enzyme fragment sizes (BP) generated by digestion of bacteriophage λcI857ts DNA.

Fragment number	Enzyme PstI[a,b]	HindIII[a,c]	EcoR1[a,c]	Uncut[a]
1	11,499	23,150	21,240	48,540
2	5,080	9,420	7,420	
3	4,749	6,560	5,810	
4	4,505	4,380	5,650	
5	2,840	2,320	4,880	
6	2,549	2,020	3,540	
7	2,454	560		
8	2,443	125		
9	2,140			
10	1,986			
11	1,700			
12	1,159			
13	1,093			
14	805			
15	516			
16	467			
17	448			
18	339			
19	265			
20	247			
21	216			
22	210			
23	200			
24	164			
25	150			
26	94			
27	87			
28	72			
29	15			

References:

[a] Sanger et al (1982)

[b] R. Hayward (personal communication)

[c] Philippsen et al (1978)

CHAPTER 3 CONSTRUCTION OF EcoRI CLONE LIBRARIES OF MAIZE MITOCHONDRIAL DNA

3.1 INTRODUCTION

A prerequisite for much of the experimental work described in Chapters 4 and 5 of this thesis was the preparation of collections of cloned mt DNA fragments. These clone 'libraries', containing a substantial proportion of the mitochondrial genome, were used for the isolation of fragments of DNA using smaller, or overlapping mt DNA fragments as hybridisation probes.

Two types of library were constructed:

1) A permanent, ordered library of mt DNA fragments cloned in a plasmid vector;
2) A temporary, non ordered library of fragments in an M13 phage vector. This library was probed once then discarded.

The methods used for the construction of these libraries are described below. Although the second type of library was not maintained, the approach used in its construction could be of more general use in the future for rapid analysis of mt DNA.

3.2 CONSTRUCTION OF AN ORDERED LIBRARY OF CLONED EcoRI FRAGMENTS OF MITOCHONDRIAL DNA ISOLATED FROM MALE FERTILE MAIZE.

3.2.1 Aims

The aim of this work was to establish a complete library of cloned mt DNA fragments of useful size (<10 KB) from male fertile (Normal, line B37N) maize. The library was maintained in an ordered array and clones containing DNA sequences of interest could be isolated from it by standard colony hybridisation techniques (Grunstein and Hogness, 1975; Gergen et al, 1979). The construction of this library made the repeated isolation and cloning of mt DNA restriction fragments eluted from gels unnecessary. The library also complemented similar

BamHI and HindIII libraries subsequently constructed in this laboratory. Thus, once a single restriction fragment containing a DNA sequence of interest had been isolated (as described in Chapter 4), overlapping clones could be obtained simply by hybridisation of this sequence to the appropriate clone library.

The restriction enzyme EcoRI was chosen for the construction of this library for a number of reasons:

1) Digestion of B37N mt DNA with this enzyme yields relatively small fragments (<13 KB, Fig. 3.1) which are easy to manipulate in subsequent investigations. In addition all fragments should be represented with roughly equal frequencies in the library since there are no exceptionally large fragments which would transform at lower efficiency, following ligation to the vector DNA (Maniatis et al, 1982).

2) The enzyme reliably and reproducibly digested mt DNA to completion.

3) Cloning vectors were available containing drug resistance markers which could be inactivated by DNA insertion at the EcoRI site. The vector pBR328 was chosen for the reasons outlined in Chapter 2, and also because this same vector could then be used for construction of the BamHI and HindIII libraries.

In order to determine the number of EcoRI clones required to produce a relatively complete library, it was necessary to estimate the number of DNA fragments generated by digestion of mt DNA with EcoRI. Assuming a genome complexity of 700 KB (Ward et al, 1981; Lonsdale et al, 1983; D. Lonsdale, pers. commun.), and that EcoRI cuts, on average, every 4^6 = 4096 BP, then about 170 fragments should be produced. However, since the G + C content of maize mt DNA is only 47% (Shah and Levings, 1974) the recognition site ($^{5'}$GAATTC) for EcoRI will occur every $1/[(0.235)^2.(0.265)^4]$, or 3672 BP. In this case, about 190 fragments would be generated (see Moore, 1983). The

probability,p, of any given point on the mt genome being excluded from a library of n clones is therefore $[(700 - 3.672)/700]^n$, i.e. $p = (0.99475)^n$ (see also Clarke and Carbon, 1978). Arbitrarily, a 10% probability of exclusion (i.e. 90% 'complete' library) was considered to be a realistic value, and this would entail the generation of $[\log_{10}(0.1)/\log_{10}(0.99475)]$, i.e. 438 clones.

3.2.2 Generation of pBR328 clones containing B37N mt DNA EcoRI fragments.

Mitochondrial DNA was isolated from etiolated maize coleptiles and 0.5 µg treated with EcoRI as described in Chapter 2. 0.2 µg of this DNA was electrophoresed on a 1% agarose minigel to assess the completion of the reaction. The remainder of the restriction digest was then heated to 65°C for 10 minutes and rapidly cooled on ice to inactivate the enzyme. 1 µg of pBR328 DNA was digested to completion with EcoRI, heat inactivated and dephosphorylated with calf intestinal phosphatase (CIP, Chapter 2). The vector DNA pellet, recovered after ethanol precipitation, was dissolved in TE80 at a final concentration of 50 ng/µl. The CIP step was included to reduce the fraction of transformant, but non-recombinant, colonies from ca. 90% to <10% (Efstradiatis et al, 1977), an important consideration in the construction of a library containing a large number of different cloned restriction fragments.

Approximately 83 ng of restricted mt DNA was ligated to 100 ng of EcoRI treated pBR328 DNA in a 20 µl reaction volume. 5 µl of the ligated DNA mixture was then used to transform 100 µl of competent E.coli HB101 cells, the remainder of the ligation was stored at -20°C. After dilution of the transformation mixture to 2 ml with L-broth followed by incubation at 37°C for 1 hr, 200 µl aliquots were spread on ten 85 mm diameter L-agar plates containing 10 µg/ml tetracycline (Tc). Higher concentration of the antibiotic have been shown to

select for clones carrying smaller DNA inserts (Gergen et al, 1979). Ampicillin (Ap) was not used because this was found to reduce considerably the number of transformants which were able to grow (see section 2.2.9.2).

After overnight incubation at 37°C, each plate contained 10 to 20 transformant (Tc^r) E.coli colonies. These clones were transferred, using sterile toothpicks, to identical 100 place grids on L-agar plus 100 µg/ml Ap and L-agar plus 100 µg/ml chloramphenicol (Cm) plates. Colonies displaying recombinant (Ap^r, Tc^r, Cm^s) phenotype could readily be identified after 16 hr incubation at 37°C. In order to generate sufficient recombinants to complete the library, the transformation was repeated a further three times using the frozen ligation mixture. A total of 419 recombinants were obtained so the library was estimated to contain 89% of the mt genome. Plasmid DNA was isolated from a variety of these recombinants by the method of Birnboim and Doly (1979) and treated with EcoRI (Fig. 3.1). Most (>75%) of the clones contained visible mt DNA inserts, the remainder were assumed to contain inserts too small to be detected by the gel system used.

3.2.3 Establishment of an ordered array of mt DNA clones.

The protocol devised by Gergen et al (1979) was followed for the organisation, preservation at -80°C and immobilisation on Whatman 541 paper of the 419 recombinant clones generated in the previous section. Briefly, these steps were as follows:

1) Recombinant E.coli colonies were transferred to a 6 x 8 grid on an L-agar plate containing 100 µg/ml Ap. The grid was arranged to coincide with the positioning of the wells in half of a Corning 96 well polystyrene microtitre dish. The colonies were allowed to grow at 37°C for 4 hr.

2) Using a 6 x 8-pronged replicating device, sterilised with

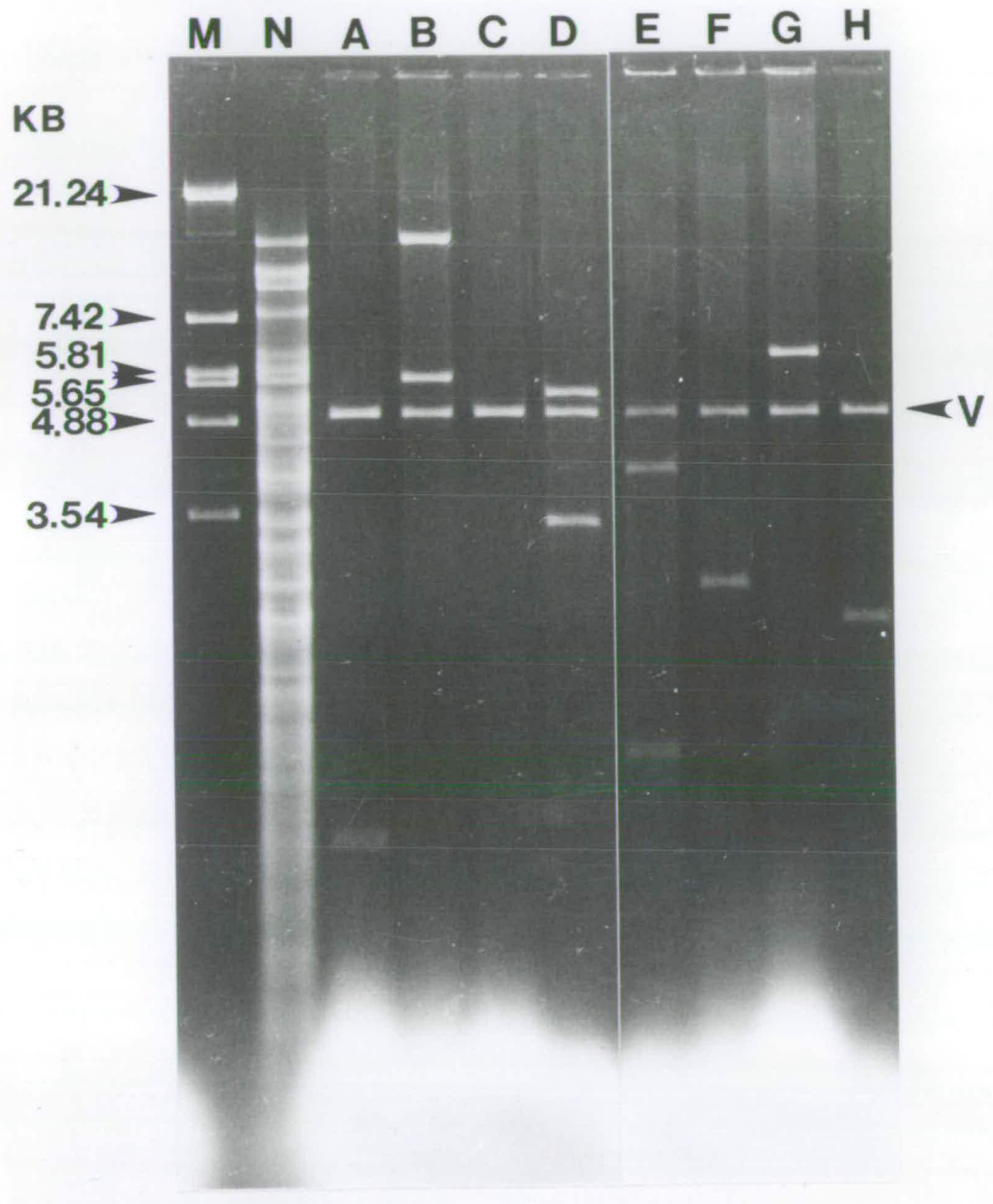


Fig 3.1 EcoRI digests of B37N mt DNA (lane N) and of representative clones from the EcoRI library (lanes A-H) Lane M = λ DNA digested with EcoRI. V indicates the vector (pBR328, = 4907 BP).

flaming ethanol, colonies were transferred simultaneously from these plates onto three duplicate L-agar plates (containing 100 µg/ml Ap) and allowed to grow for 16 hr at 37°C. The orientation of the colonies on each plate was designated as shown in Fig. 3.2.

3) The same replicating device was used to transfer inocula from the master grid colonies to wells, each containing 200 µl of freezing medium (see below), in a 96 place microtitre dish. The designation of the cell well co-ordinates corresponded to the designation given to each colony in the replica plates (see Fig. 3.2). Two 6 x 8 grids were transferred to each dish, and five dishes (numbered EN1 to EN5) were required for maintenance of the entire library.

Freezing medium contained L-broth supplemented with 100 µg/ml Ap, 4.4% w/v glycerol, 0.63% w/v K_2HPO_4, 0.18% w/v KH_2PO_4, 0.045% w/v Na_3 citrate, 0.009% w/v $MgSO_4.7H_2O$ and 0.09% $(NH_4)_2SO_4$. All components except antibiotic were mixed and autoclaved (121°C, 15 minutes) prior to use. Sterile antibiotic was added immediately prior to dispensation into the wells with a 12 nozzle Titertek pipetter. After inoculation, microtitre dishes were covered with sterile plastic lids and incubated at 37°C for 16 hr, in a sealed plastic box containing moistened paper towels to reduce evaporation of the medium. Following incubation, the medium became turbid due to growth of the E.coli. Dishes were then wrapped in tinfoil and stored at -80°C until required.

4) After 16 hr incubation, the replica colonies from section 2 were normally 6-7 mm in diameter. These were overlain with 5.6 x 7.3 cm sheets of autoclaved (121°C, 15 minutes) Whatman 541 paper previously marked with orientation co-ordinates as described above. Plates were then incubated for a further 2 hr at 37°C to allow cells to grow into and adhere to the paper.

5) The filters were transferred, colony side downwards, to plates

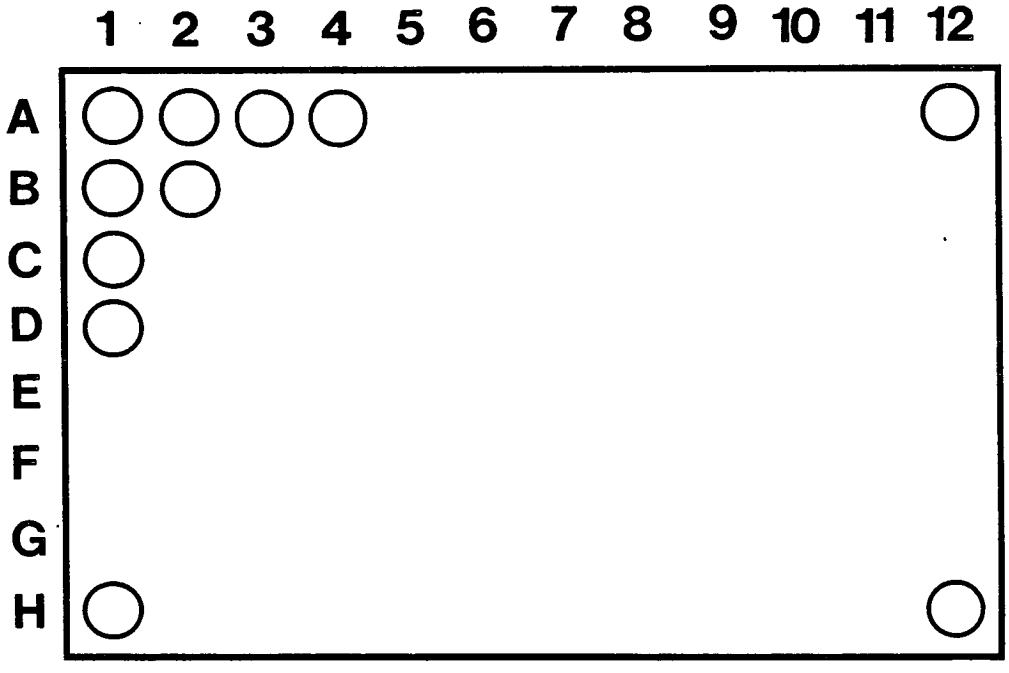


Fig 3.2 Plan view of a 96 well microtitre dish used for storage of the EcoRI library.

The 6 x 8 grids in which colonies were grown on 85 mm agar plates corresponded to half (A - H x 1 - 6 or A - H x 7 - 12) of the dish and were numbered accordingly.

containing L-agar supplemented with 250 µg/ml Cm. Plasmids were allowed to amplify for 24 hr at 37°C.

6) The colonies were lysed and the DNA denatured and fixed to the filter by successive passage through the following solutions:

(i) 0.5 M NaOH (2 x 5 minutes)

(ii) 0.5 M Tris HCl pH 7.4 (2 x 5 minutes)

(iii) 2 x SSC (2 x 5 minutes)

After a brief wash in 95% ethanol, filters were air dried then stored in sealed plastic petri dishes at 4°C.

BamHI and HindIII libraries in pBR328 have been constructed from B37N mt DNA exactly as described above except that Ap was the antibiotic used for selection (recombinants are Ap^r, Tc^s, Cm^r) and plasmid amplification was carried out in the presence of 100 µg/ml spectinomycin. The BamHI and HindIII libraries contain 172 and 648 clones respectively, and are therefore 68 and 97% 'complete'.

3.2.4 Detection of specific probe-homologous clones in the EcoRI library by colony hybridisation.

Filters, prepared as described above, were pre-hybridised for 1-2 hr at 37°C in 0.2 ml/(cm^2 of filter) of a buffer containing 50% formamide, 5 x SSC and 250 µg/ml thermally denatured, sonicated herring sperm DNA. The filters were hybridised with radioactive probe DNA, in 0.1 ml/cm^2 of the same buffer, for 16 hr at 37°C with gentle rocking, then washed, with shaking, four times in 1 ml/(cm^2 of filter) of 2 x SSC at room temperature. Colonies containing DNA sequences complementary to the probe were detected by autoradiography (12-48 hr) of the washed filters. Normally, 0.1-1 µg of probe DNA (specific activity 10^6-10^7 dpm/µg) was used in the hybridisation reaction, but colonies homologous to probes with a total activity of $<10^5$ dpm were also successfully detected after longer fluorographic exposures (section 2.2.7.3)

The DNA probes used were chosen to minimise spurious hybridisation signals due to vector:vector or vector:E.coli homology. Mt DNA fragments contained in ColE1-derived vectors could not be used because of homology to pBR328, used in the construction of the library. This effect could be reduced by addition of non-labelled competing vector DNA, but would be expensive in terms of the amount of DNA required. The most generally useful probes were nick-translated cloned mt DNA fragments, separated from the vector by restriction enzyme digestion followed by agarose gel electrophoresis and electroelution. An alternative approach was to use single stranded mt DNA fragments cloned in M13 and labelled by second strand synthesis. Separation from the vector was then unnecessary since M13 and pBR328 do not have sequence in common (except M13 mp9, which contains pBR322 sequences, see footnote to Messing and Vieira, 1982). Under the conditions described above, cross hybridisation of the β-galactosidase, α peptide gene in the M13 'mp' vectors (Messing et al, 1977) with the host E.coli chromosome (Bolivar and Backmann, 1979) did not contribute significantly to the background signal. Colonies containing plasmids with DNA sequences complementary to the probe hybridised several orders of magnitude more strongly than this background. The reasons are probably two-fold:

1) Second strand synthesis from the annealed primer does not normally completely copy the whole M13 molecule (Hu and Messing, 1982). Thus, sequences complementary to the lac region may not be included in the probe because these sequences are most distal (on the circular genome) from the point at which the primer binds.

2) Plasmid DNA sequences are probably amplified several hundred-fold compared to the single-copy chromosomal DNA (Clewell, 1972).

M13 clones could not be successfully used under 'nonstringent' conditions to probe the EcoRI library since all colonies were found to hybridise (not shown). This could result from weak second-order

hybridisation of labelled M13 DNA, via unlabelled molecules, to the E.coli chromosomal DNA (see section 4.3.2).

Hybridised DNA was removed from the filters, following autoradiography, by a repetition of the denaturation/neutralisation steps described above. Filters could then be reused in exactly the same way, after allowing at least 20 days to permit radioactive decay of any residual bound probe (Gergen et al, 1979).

Colonies of interest were recovered from the frozen replicas after allowing the microtitre dishes to thaw. A loopful of culture was removed from the relevant well and used to initiate larger scale bacterial cultures from which plasmid DNA could be isolated for more detailed analyses. It was found that the colonies could be repeatedly frozen and thawed with no appreciable loss of viability. In order to determine the grid co-ordinate of the hybridising colonies, the autoradiograph was overlain on the aligned filter. The colonies were readily located by their fluorescence when illuminated by 302 nm ultraviolet light.

3.3 CONSTRUCTION OF A NON-ORDERED EcoRI LIBRARY FROM THE MITOCHONDRIAL DNA OF MALE STERILE (B37T) MAIZE.

3.3.1 Aims

Previous experiments (Chapter 5) had shown that a 2.8 KB EcoRI mt DNA fragment from male sterile Texas (T) maize lines was adjacent in the mt genome to the 2.4 KB EcoRI fragment containing the COII (≡ mox-1) locus. This fragment was of interest because it apparently replaced a shorter fragment in the homologous position in the mt DNA of male fertile (N) lines. A specific clone was required, containing this portion of the T mitochondrial genome, in order to investigate in greater detail the nature of this reorganisation. A non-ordered M13 clone library was constructed to facilitate the isolation of this fragment for the following reasons:

1) The desired fragment was small enough to be stable when cloned in M13, whereas larger fragments (>3.5-4 KB) would not. The proportion of the total mt genome appearing in the library would therefore be substantially reduced, and enriched for the desired sequence.

2) An M13 library could be rapidly constructed at high plaque density, requiring only ten 85 mm diameter minimal agar plates. Plaques would be screened *in situ* without intermediate selection.

3) The library could be probed directly with a specific pBR328 clone without vector hybridisation.

4) A clone would be generated from which single stranded DNA, suitable for sequence analysis by the chain-termination method (Sanger *et al*, 1977), could easily be isolated.

3.3.2 Establishment and probing of the M13 T-mitochondrial DNA library

Mt DNA was isolated from etiolated coleoptiles of maize line B37T then digested with *EcoRI* and ligated to *EcoRI* digested M13 mp8 as described in Chapter 2. The ligation mixture (10 µl) was diluted ten fold and 10 µl aliquots were used to transform competent *E.coli* JM101 cells. It was found more efficient to dilute the ligation mixture prior to transformation rather than afterwards, since many more transformants were obtained in this way (see also Humphreys *et al*, 1979). Each transformation was plated in the normal way onto a single 85 mm diameter minimal medium plate.

Following overnight incubation at 37°C, each plate had developed approximately 800 blue plaques and 150 white plaques. The clone library was therefore estimated to contain 1500 recombinants. The probability of exclusion of any point on the mt genome is therefore $(0.99475)^{1625}$, which is <0.02%. Because of the size-selection by M13, the proportion will actually be much higher than this, but the probability of detecting a 2.8 KB fragment in this library is still high.

Each plate was replicated onto a sterile nitrocellulose disc, and these discs were probed with the homologous clone from N mt DNA, pZmH5 (Chapter 5), using the methods and conditions described in Chapter 2. The plates were meanwhile stored at 4°C until required. After hybridisation, the filters were fluorographed in order to detect plaques homologous to the probe (Fig. 3.3). In total, approximately 20 plaques hybridised to the probe. The autoradiogram was aligned with replica filters and the appropriate white plaques identified. Normally, these plaques were adequately separated from nearby plaques to make further purification unnecessary. Viable phage could be isolated from plates stored at 4°C for up to 3 months after transformation, provided efforts were made to prevent fungal and bacterial contamination. Thus, if desired, the libary could have been re-probed to detect other sequences. The plaques of interest were used to initiate single stranded and RF DNA preparations. The clone containing the desired 2.8 KB DNA insert (number 1573) was subsequently identified by sequence analysis and hybridisation data (see Chapter 5).

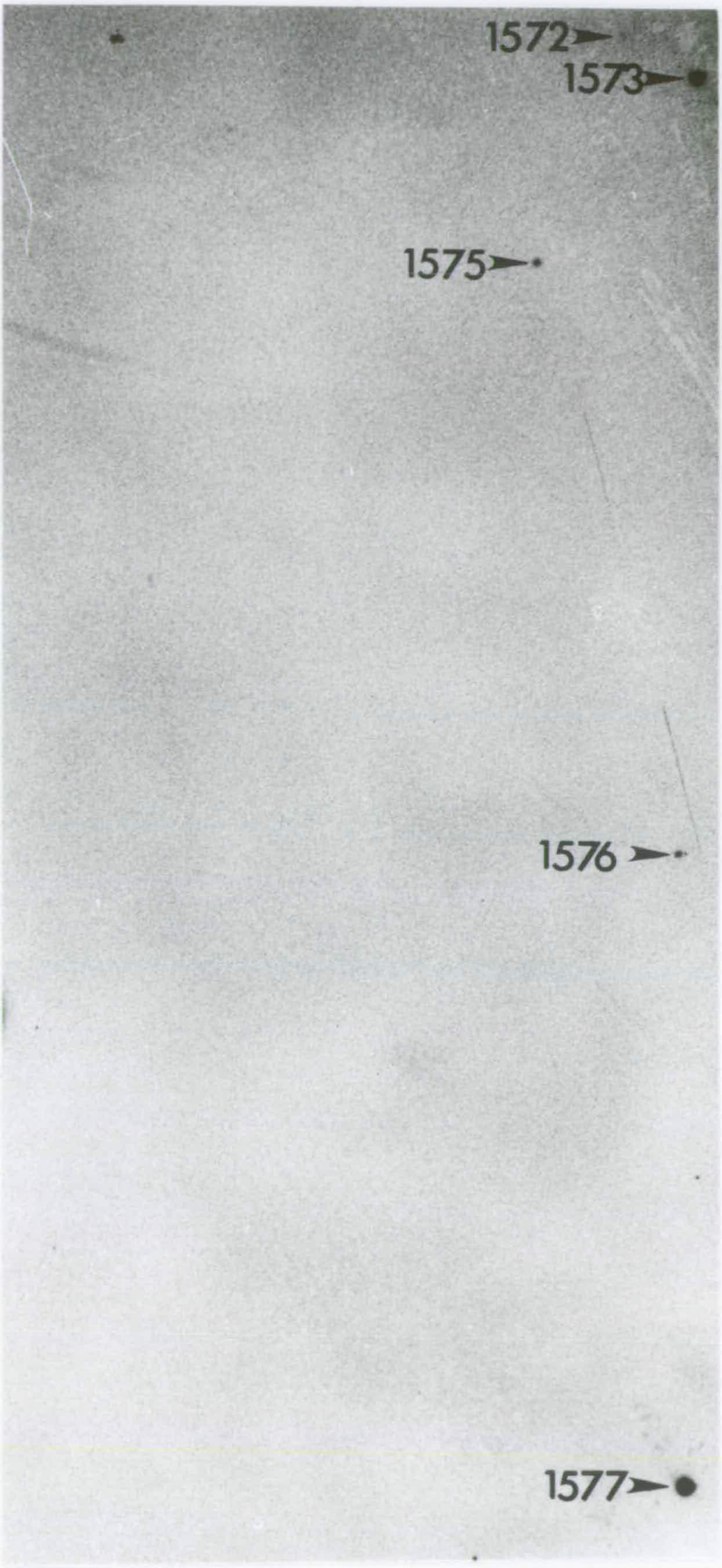
3.4 CONCLUSIONS

The construction of two types of clone library from maize mt DNA has been described in this chapter. Firstly, a permanent, ordered array of mt DNA clones in the plasmid vector pBR328 has been produced from the male fertile maize line B37N. This library is stored as a grid of frozen plasmid-containing E.coli colonies at -80°C. In addition, plasmid DNA from a replica grid has been immobilised on Whatman 541 paper and can be repeatedly used for colony hybridisation experiments. Using such an approach, clones containing DNA fragments homologous to both COB (Chapter 4) and COI (P. Isaac, pers. commun.) have been isolated from this library.

The second type of library was constructed using the bacteriophage

Fig 3.3 In situ hybridisation of pZmH5 to the B37T EcoRI library.

The hybridising plaques shown in this figure were designated 1572, 1573, 1575, 1576 and 1577 (arrowed). Clone 1573 contained the sequence of interest.


1572
1573
1575
1576
1577

M13 mp8 cloning vector. This library was constructed specifically to isolate a single mt DNA fragment from the male sterile maize line B37T. Once isolated, the library was discarded. A replacement library could, however, be established with ease if similar requirements arose in the future.

Each library has its own particular merits, but used in combination, they provide a bank of mt DNA fragments from which novel sequences can be isolated and analysed with rapidity. The sequence analysis of mitochondrial genes could, in the future, be facilitated by the construction of a large library of small mt DNA fragments cloned in M13. For example, a 99% complete Sau3A library would require approximately 12,500 recombinants. This library could be probed, under nonstringent conditions, with an appropriate heterologous gene sequence which could be cloned in any plasmid vector. It would not be necessary to remove vector DNA sequences from the probe since sequences capable of cross-hybridising with the vector should not be present in the library.

M13 clones identified in this way could be sequenced directly and also labelled by second strand synthesis then used to probe the permanent plasmid mt DNA library. Longer DNA fragments would therefore be identified which could be used to extend sequence analysis obtained from the smaller M13 clone. The sequence of any mitochondrial gene for which 'heterologous' probes were available could therefore be built up with rapidity and without the need at any time to purify probe sequences from the plasmid vector containing them.

CHAPTER 4 THE APOCYTOCHROME b GENE IN MAIZE MITOCHONDRIA

4.1 INTRODUCTION

The first sections of this chapter describe the isolation and sequence analysis of the apocytochrome b (COB) gene in maize (Zea mays, L.) mitochondria. Predictions from this sequence, and the precise mapping of this gene on the maize mitochondrial (mt) DNA restriction map, are discussed in subsequent sections.

4.1.1 Rationale

Apocytochrome b is a key polypeptide in the ATP-generating electron transport chain of the inner mitochondrial membrane (see Slater, 1983 for review). The gene encoding this protein (COB) has been shown by genetic and biochemical analyses (Haid et al, 1979; Anderson et al, 1981; see also Chapter 1) to reside in the mt genome of all eukaryotes studied to date. The COB gene has been isolated and sequenced from a variety of fungi (Saccharomyces cerevisiae, Nobrega and Tzagoloff, 1980; Aspergillus nidulans, Waring et al, 1981; Neurospora crassa, Citterich et al, 1983; and Kluyveromyces lactis, M. de Haan and L. Grivell, pers. commun.) and mammals (Homo sapiens, Anderson et al, 1981; Mus musculus, Bibb et al, 1981; Rattus rattus, Görtz and Feldmann, 1982; Bos taurus, Anderson et al, 1982), but not any higher plant. Sequence analysis of this gene in these organisms has revealed a wide diversity in its organisation and structure, from a complex arrangement of introns in Saccharomyces cerevisiae (Lazowska et al, 1980) coupled with an equally complex pattern of transcription (Van Ommen et al, 1980), to the relatively simple genes found in all animals studied so far, which lack any intervening DNA sequences (Borst and Grivell, 1981). Sequence analysis of the maize COB gene would considerably increase the body of data available on protein-coding genes in plant mitochondria, at present limited to

the cytochrome oxidase subunit II gene (Fox and Leaver, 1981). It would also allow comparisons of gene structure, codon usage and polypeptide conformation both *within* the cytochrome *b* s and, for the first time, *between* different plant mitochondrial genes, to be made. Furthermore, in the absence of a linked transcription-translation system for plant mt DNA (see De Ronde *et al*, 1980 and Leaver and Gray, 1982; also V. Jones, pers. commun.) and a set of well characterised non-lethal mitochondrial mutations (Leaver and Gray, 1982), DNA sequencing (coupled with RNA transcript analysis), is the only method available to prove that a particular gene actually resides in maize mt DNA. This is especially important in the case of cytochrome *b* since antibodies raised against the yeast polypeptide consistently fail to immunoprecipitate any maize mt protein (C.J. Leaver, pers. commun.) despite success with 'heterologous' cytochrome oxidase subunits I and II antibodies (Forde and Leaver, 1974).

Sequence analysis of the maize *COB* gene was undertaken primarily for the reasons described above, but an additional consideration was the availability of a wide variety of well characterised DNA hybridisation 'probes' from other organisms. These could be used under 'non stringent' conditions to detect the putative *COB* region in maize mt DNA, prior to its isolation and sequence analysis. A number of such 'probes' were used, containing sequences from several mt genes, but in almost all experiments (see section 4.2.1.2), the strongest hybridisation signals were obtained from *COB* probes. This proved to be a compelling final reason for the selection of the *COB* gene for further study.

4.1.2 The structure and function of cytochrome b

The *b* type cytochromes are a collection of protohaem IX-containing polypeptides and are identified by an α band of the pyridine-ferrohaemochrome

in alkali at 556-558 nm (see Lemberg and Barrett, 1973, for review). They are widely distributed in nature and are found in animals, plants, fungi, and bacteria. Many different b-type cytochromes have been identified displaying a variety of molecular weights, solubilities and redox potentials. All are involved in electron transport processes through changes in the oxidation number of the haem iron (Keilin, 1925). Of particular interest are the cytochromes b which are found in the mitochondria of the eukaryotes (Green and Hatefi, 1961) and in many bacteria (see for example Keilin, 1927 or Hauska et al, 1983, for review). These cytochromes are involved in a variety of redox reactions, (Reiske, 1973), for example Saccharomyces mitochondria contain a cytochrome b_2 which is associated with L-lactic dehydrogenase (Bach et al, 1946). Plant mitochondria probably also contain cytochromes b_{559} (Bendall, 1968) and b_{555} (Storey, 1969) of uncertain function. However, the b-cytochromes of interest in this work are those normally co-isolated with cytochrome c_1 in the 'cytochrome bc_1 complex' (Hatefi et al, 1962) and are involved in electron transport during oxidative phosphorylation.

The 'cytochrome bc_1 complex' is 'complex III' of the mitochondrial inner membrane, more correctly known as Ubiquinol: cytochrome c oxidoreductase (EC 1.10.2.2) (Saraste, 1983). In fungal and mammalian mitochondria, this complex contains 7-9 different subunits (Mendel-Hartig and Nelson, 1981; Weiss and Kolb, 1979; Kreike, 1982) in which cytochrome b (subunit IV) is probably present as a single subunit per monomeric complex (Kreike et al, 1979) although many workers suggest a stoichiometry of 2 (Hauska et al, 1983). Several spectral forms of cytochrome b are found in complex III (b_{558}, b_{566}, b_{562} (low potential) and b_{562} (high potential); Slater, 1983) although it appears that only a single cyt b polypeptide exists,

containing two protohaems (Slater, 1981; Saraste, 1983; Widger et al, 1983). When cytochrome b is isolated, the spectral differences between the b species disappear (von Jagow et al, 1978). Apocytochrome b (i.e. the polypeptide to which protohaem groups are attached in cytochrome b) is a product of translation by mitochondrial ribosomes (Krieke et al, 1978), but the other subunits of complex III appear to be encoded in nuclear DNA, synthesised on cytosolic ribosomes and imported (Nelson and Schatz, 1979; Tienze et al, 1982). In view of the fixed stoichiometries of the complex III subunits, rigorous mechanisms must exist to ensure that expression of mitochondrial and nuclear genes are co-ordinated. At present no satisfactory model to explain this co-ordination has been proposed. Sequence analysis of the genes involved combined with biochemical analyses of the polypeptides will be required in the formulation of a complete model.

The role of complex III in mitochondrial electron transport has been ascertained using a number of biochemical and biophysical methods (see Yu and Yu, 1981). Electrons are transferred from ubiquinone ('Q') to cytochrome c (the penultimate member of the respiratory chain) via the cytochrome b (Chance, 1974), iron-sulphur protein and cytochrome c_1 subunits of complex III (Mitchell, 1976). Cytochrome b appears to be involved in Q-cycle reduction/oxidation of ubisemiquinone and thereby achieves vectorial transfer of electrons from the outer side of the inner mitochondrial membrane to the inner (matrix) side (Slater, 1983). These findings are consistent with a proposed model (Widger, et al, 1983; Saraste, 1983; Wikström and Saraste, 1983) for the secondary and tertiary structure of cytochrome b in which the protein is envisaged to cross the inner mitochondrial in nine α-helical segments, with the two protohaem groups being held between helices II and V (see section 4.2.3.3.2 for fuller discussion). Vectorial transport of

electrons across the membrane in complex III could explain how, according to the chemiosmotic model (Mitchell, 1961), ATP synthesis is coupled to electron transport in this complex (see Wikström et al, 1981 for review). It must be noted that these models have been evolved using data from mammalian and fungal mitochondria and as yet the isolation of intact complex III from any higher plant has not been reported in the literature. It would therefore be desirable to elicit more information on this complex in higher plants in order to determine whether the same models are applicable. Sequence analysis of the COB gene provides a rapid and relatively easy means of gathering such data. If the gene were shown to encode a radically different polypeptide, this would suggest that vectorial electron transport could not occur by the proposed mechanisms in higher plant mitochondria. Furthermore, isolation of the COB gene from maize mt DNA would be the first formal proof that mitochondrically-encoded subunits form part of a higher plant complex III. Once this step has been taken, further experiments can be designed to investigate how nuclear and mitochondrial genomes interact in higher plants to bring about the precise assembly of a functional complex III, and how these interactions may differ from those in other eukaryotes.

4.2 ISOLATION AND SEQUENCE ANALYSIS OF THE MAIZE COB GENE

4.2.1 Identification and cloning of the COB gene

4.2.1.1 DNA hybridisation probes.

The following probes were used to identify the COB gene in maize mt DNA:

dD22: a cytoplasmic petite derived from a strain of S.cerevisiae containing the whole of a 'long' COB locus, with a repeat size of 11 KB (T.D. Fox, pers. commun., see also fig. 4.1). This yeast strain was a gift from T.D. Fox (Cornell University, USA).

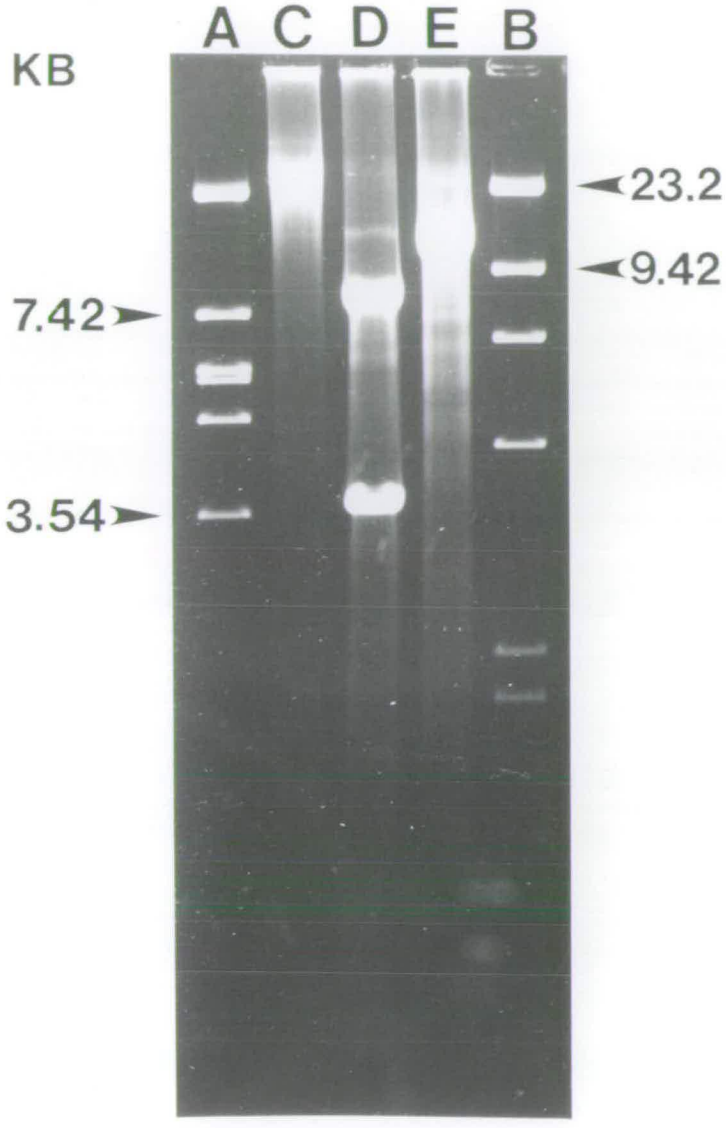


Fig 4.1 Restriction enzyme digestion of mt DNA from the yeast petite strain dD22, containing the 'long' COB gene.

Cleavage with HindIII (lane E) reveals a repeat length of ca 11KB. Digestion with EcoRI (lane D) shows the repeat unit contains two sites separated by 7.5 and 3.6KB. Lanes A and B are λDNA cut with EcoRI and HindIII respectively. See Lazowska et al (1980) and Nobrega and Tzagoloff (1980) for further details.

Lane C (undigested)

M13 B6: an M13 mp2 clone containing nucleotides <15,100 to 15,361 of the bovine (Bos taurus) COB gene (the gene spans nucleotides 14,514 at the amino terminus to 15,653 at the carboxy terminus; Anderson et al, 1982).

M13 T13: an M13 mp2 clone containing nucleotides 14,568 to <15,000 of the bovine COB gene. Both B6 and T13 were gifts from S. Anderson, MRC Unit, Cambridge .

M13 BglII-4: an M13 mp6 clone of a 1.6 KB BglII fragment from Aspergillus nidulans mt DNA containing nucleotides 24,000 to 25,600 (approximately). The amino-terminal exon of this gene spans approximately nucleotides 24,500 to 25,100 (Waring et al, 1981). This clone was a gift from R.B. Waring (UMIST, Manchester).

M13-3.8: an M13 mp7 clone of a 750 BP MboI fragment from the COB gene of K. lactis, and was a gift of M. de Haan (University of Amsterdam, the Netherlands). The gene spans nucleotides 112 to 1269 in this organism, and the clone covers nucleotides 209 to 958 (M. de Haan and L.A. Grivell, pers. commun.).

DNA hybridisation probes for genes other than COB were as follows:

pKL41, pKL106 and pKL114 are pBR322 clones containing portions of the S. cerevisiae cytochrome oxidase, subunit I gene (COI) (Hensgens el al, 1983) and were gifts of L.A.M. Hensgens (University of Amsterdam, the Netherlands).

RP6: a petite yeast strain whose mt genome covers the whole of the ATPase, subunit 9 (oli-1) gene (Hensgens et al, 1979). The strain was a gift of P. Slonimski (Centre de Genetique Moleculaire du CNRS, Gif-sur-Yvette, France).

LH30-A5: a yeast petite containing the COI and ATPase, subunit 6 (oli-3) genes (Hensgens et al, 1983), a gift of L.A.M. Hensgens. Hybridisation experiments using this probe were not initially successful and are not described here.

DNA was prepared from these plasmid or M13 clones and from the yeast petites as described in Chapter 2, then labelled with ^{32}P by nick-translation or second strand synthesis.

4.2.1.2 'Heterologous' hybridisation to maize mt DNA.

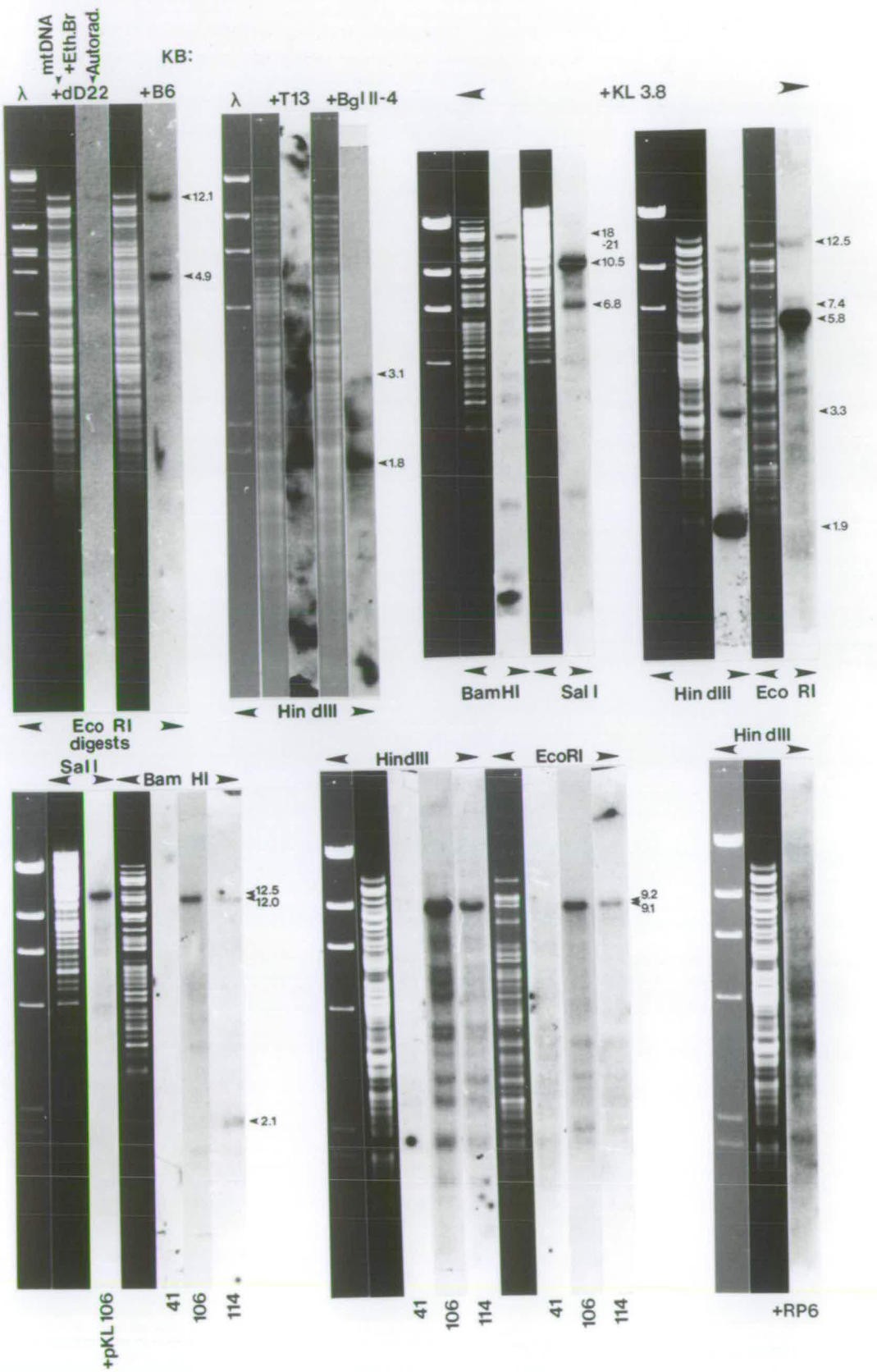
Mt DNA (strain B37N) was prepared from etiolated maize coleoptiles, digested with a variety of restriction endonucleases, electrophoresed through 1% agarose gels and transferred to nitrocellulose (see Chapter 2). Fig. 4.2 shows hybridisation of various labelled DNA probes to such mt DNA blots under non-stringent (Tm-45°C) conditions. The results are summarised below:

1) Despite several attempts, the RP6 DNA did not hybridise to maize mt DNA. The reasons are unknown, since maize mt DNA is thought to encode a DCCD binding protein analogous to ATPase subunit 9 (Hack and Leaver, in preparation; also Leaver and Gray, 1982). A possible factor could be the high (67%) A + T content of the gene (Hensgens et al, 1979) and the possible destabilising effect this would have on the formation of heteroduplexes. In future attempts to identify this gene in maize, other probes, for example from the subunit 9 gene in Neurospora nuclear DNA (Sebald et al, 1979) may have to be used.

2) The three yeast COI clones all hybridise to the same main EcoRI, HindIII and BamHI fragments. Since COB probes identified smaller, more manageable restriction fragments, and in view of the reasons outlined in section 4.1.1., isolation and sequence analysis of this gene was not pursued. The maize COI gene has recently been sequenced in this laboratory by Dr P. Isaac (Isaac et al, in preparation). From this hybridisation data alone, it was not possible to deduce whether the maize COI gene contained any intervening sequence, since summation of the hybridising restriction fragment sizes in any of the

Fig 4.2 Identification of maize mt DNA BamHI, SalI, HindIII and EcoRI fragments containing the COB and COI genes by heterologous hybridisation.

See text for origins of the probes used. dD22, B6, T13, Bgl II-4 and K1 3.8 are COB probes, pkL41, pkL106 and pkL114 are COI probes. RP6 is an ATPase 9 probe. Note lack of hybridisation to pkL41 and RP6.


mtDNA
+Eth.Br
-Autorad.
KB:
λ
+dD22
+B6
12.1
4.9
Eco RI
digests
λ
+T13
+Bgl II-4
3.1
1.8
Hin dIII
+KL 3.8
18
-21
10.5
6.8
BamHI
Sal I
12.5
7.4
5.8
3.3
1.9
Hin dIII
Eco RI
Sal I
Bam HI
12.5
12.0
2.1
+pKL 106
41
106
114
HindIII
EcoRI
9.2
9.1
41
106
114
41
106
114
Hin dIII
+RP6

digests yields a total size in excess of 10 KB, which would be large enough to accommodate even the longest form of the yeast COI gene (Hensgens et al, 1983a).

3) All of the COB probes used hybridised more strongly to a 1.8 KB HindIII fragment than to any other restriction fragment (particularly noticeable with the K.lactis clone). It therefore seemed likely that this molecule contained a large portion of the maize COB gene so efforts were directed towards its isolation and characterisation.

4.2.1.3 Cloning the 1.8 KB HindIII fragment from maize mt DNA

Six micrograms of maize mt DNA, prepared from the line B37N, were digested with HindIII and electrophoresed on a 1% agarose gel which was stained and photographed as described in Chapter 2. A 5 mm section immediately ahead of the 2.02 KB λ marker fragment was cut from the gel and DNA (estimated to be 100-200 ng) was isolated from this segment by electro elution then dissolved in 10 µl of TE80. These restriction fragments were ligated to 200 ng of HindIII digested pBR328 (not pre-treated with CIP) in a 20 µl reaction volume. An aliquot (5 µl) of the mixture was used to transform competent E.coli cells. After adjustment of the total volume to 1 ml with L-broth, 50, 100 and 250 µl portions were plated onto L-agar containing 100 µg/ml ampicillin and incubated overnight at 37°C.

One hundred transformant colonies were picked onto a selector plate containing 20 µg/ml tetracycline and five displayed recombinant (Ap^r Tc^s) phenotype. Plasmid DNA was prepared from these by the 'miniprep' method (section 2.2.10.1), restricted with HindIII and the DNA fragments separated on a 1% agarose gel (Fig. 4.3a). All the clones contained inserts of approximately 1.8 KB, and were designated pZmH1730, 1940, 1970, 1790 and 1840 according to the exact insert size in base pairs. This gel was blotted to nitrocellulose and hybridised

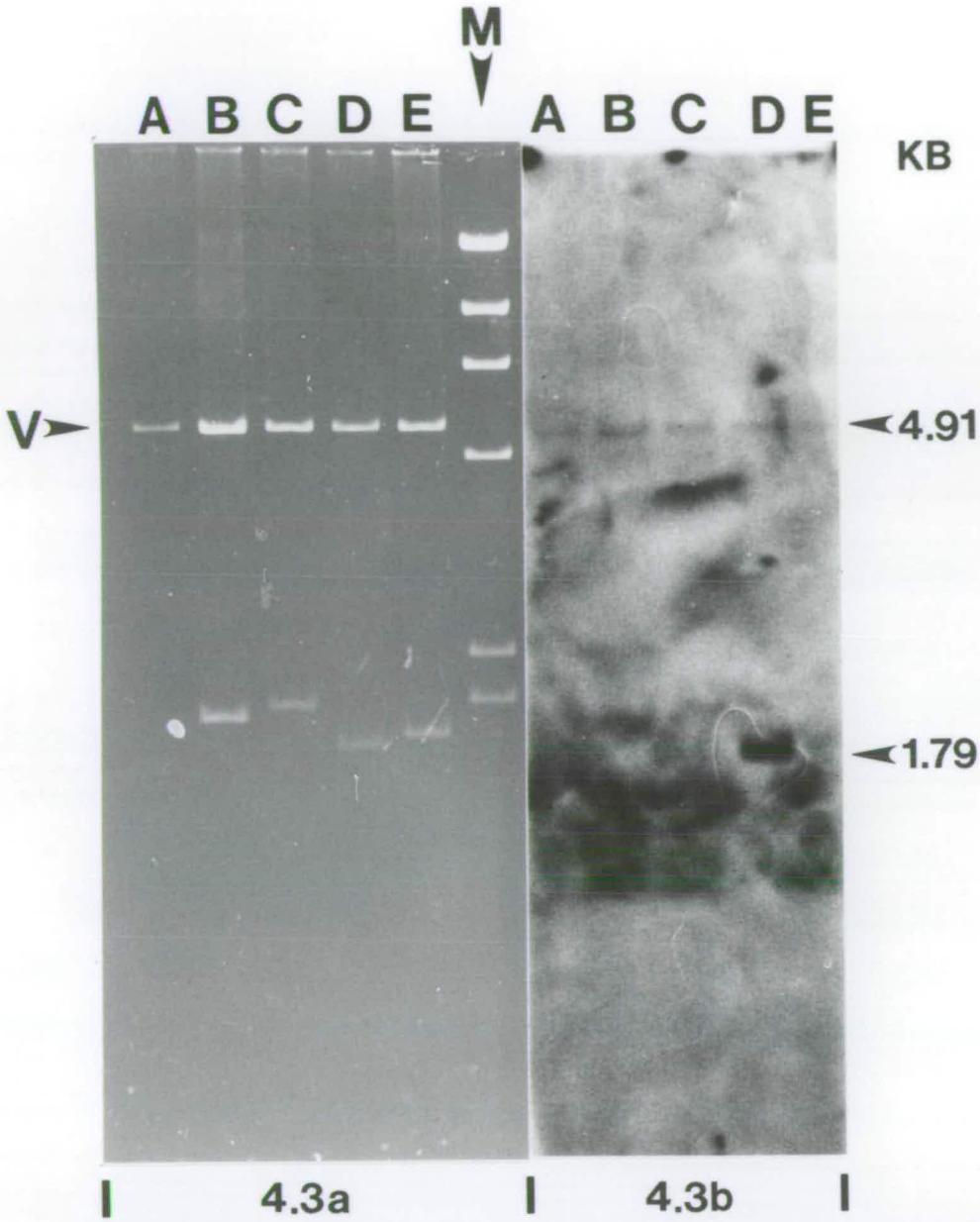


Fig 4.3 Identification of the clone containing the COB - homologous HindIII fragment using the bovine probe B6.

Tracks A to E contain HindIII digested plasmid DNA from clones pZmH1730, 1940, 1970, 1790 and 1840 respectively.

4.3a Ethidium bromide stained agarose gel. M indicates marker lane (λDNA digested with HindIII, V indicates the vector (pBR328).

4.3b Fluorogram of the same DNA after transfer to nitrocellulose and hybridisation under nonstringent conditions (T_m-46°C) with nick-translated B6. Note specific hybridisation to the insert in pZmH1790 (arrowed).
The dark background arises from the incorporation of 5% dextran sulphate in the hybridisation buffer.

with 1 µg of M13 B6 (^{32}P-labelled to 5 x 10^6 dpm/µg) under non-stringent conditions in order to detect which clone, if any, contained the maize COB gene. Fig. 4.3b shows the autoradiograph of this filter after hybridisation. Despite a high 'background' probably caused by the incorporation of 5% dextran sulphate in the hybridisation buffer, it is clear that the insert in pZmH1790 is specifically homologous to the probe. This clone therefore was selected for further study as described below.

4.2.1.4 Restriction mapping of pZmH1790

Restriction maps of pZmH1790 were constructed using a variety of enzymes, in order to guide subsequent sequence analysis. Fig. 4.4 shows that of ten hexanucleotide-recognising enzymes used to restrict pZmH1790, only BglII cuts the plasmid in the insert sequence. This enzyme reduces the insert size to ca. 1.65 KB, the site must therefore lie some 140 BP from one end. Analysis of HindIII + BglII digested pZmH1790 DNA on a 6% polyacrylamide gel more accurately sized the smaller HindIII-BglII fragment at 62 BP (Fig. 4.5). Double digestion with EcoR1 and BglII (not shown) demonstrated that the BglII site was located 1.3 KB from the single EcoR1 site in the vector.

In order to map the insert in pZmH1790 more extensively, both pZmH1790 and its HindIII insert prepared by electroelution were restricted with Sau3A, TaqI and MspI. These enzymes were used because they produce fragments with protruding ends compatible with cloning sites in M13 mp8 and were useful for subsequent sequence analysis. Fig. 4.6a shows representative polyacrylamide electrophoretograms used for the construction of the restriction map (Fig. 4.6b). HaeIII and HhaI sites are also indicated but it was not possible to produce a consistent map of AluI sites. Sequence analysis later showed this was probably due to a close coincidence of several AluI recognition sites with those recognised by TaqI and Sau3A.

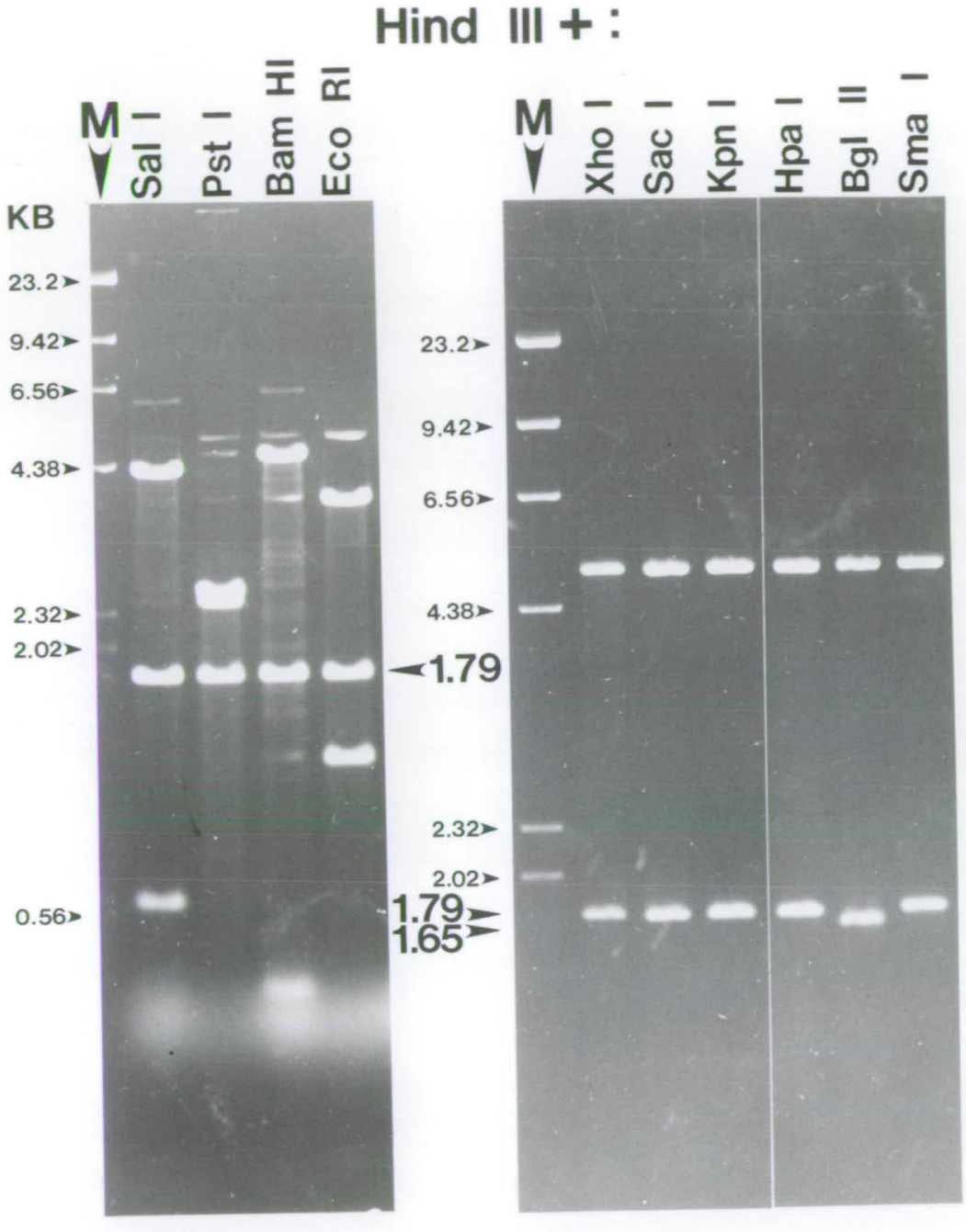


Fig 4.4 Restriction mapping pZmH1790 with hexanucleotide-recognising enzymes.

pZmH1790 DNA was double digested with HindIII plus the ten enzymes shown. Note that only BglII reduces the size of the 1.79KB HindIII-bounded mt DNA insert.

Left panel: 1.0% agarose gel; right panel 0.8%

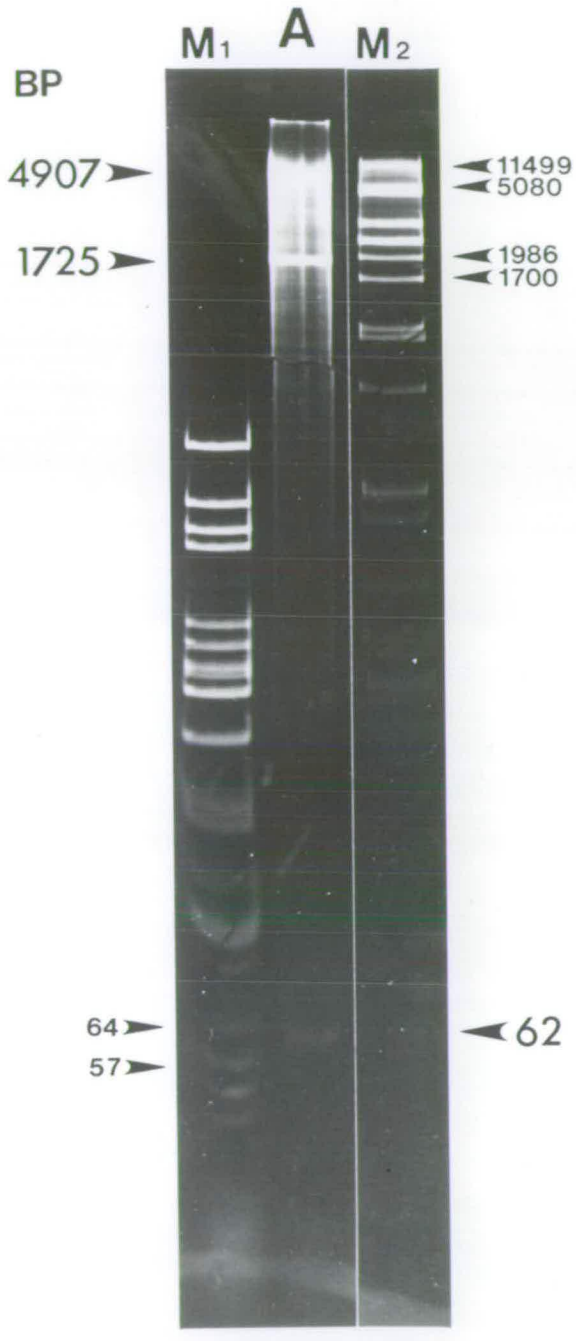


Fig 4.5 Mapping the BglII site in pZmH1790.

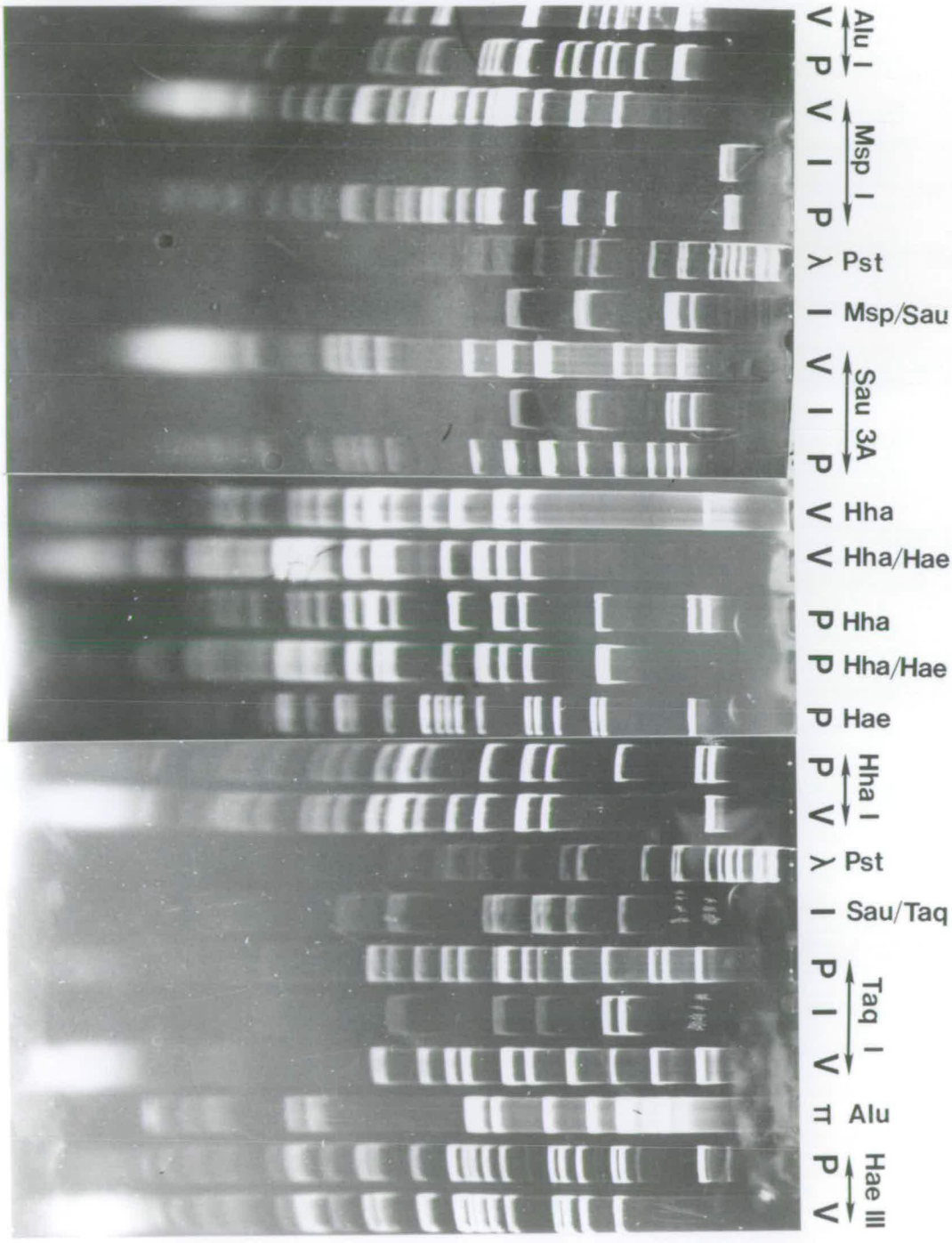
pZmH1790 DNA was double digested with HindIII and BglII (lane A). Comparison with marker DNA fragments (M_1 = pBR 328 + HaeIII, sizes from Prentki et al, 1981; M_2 = λ + PstI, sizes from table 2.2) shows the small BglII - HindIII fragment has a mobility equivalent to ca.62BP.
6% polyacrylamide gel

Fig 4.6 Restriction mapping pZmH1790 with tetranucleotide recognising enzymes.

4.6a Representative ethidium bromide stained polyacrylamide electrophoretograms used in the construction of the map. Fragment sizes were estimated by comparison with pBR322, pBR328 and λDNA digests.

Abbreviations: V = pBR328 DNA; P = pZmH1790 DNA; I = purified HindIII insert DNA from pZmH1790; π = pBR322 DNA.

4.6b (Next page)
Restriction map of Sau3A(S), TaqI(T), HaeIII (He), HhaI (Hh), AluI (A), MspI (M), HindIII (H) and BglII (B) sites in and around the 1.79 KB mt DNA insert in pZmH1790.


V P V I P λ I V I P V V P P P P λ I π P V
Alu I
Msp I
Pst
Msp/Sau
Sau 3A
Hha
Hha/Hae
Hha
Hha/Hae
Hae
Hha I
Pst
Sau/Taq
Taq I
Alu
Hae III

a

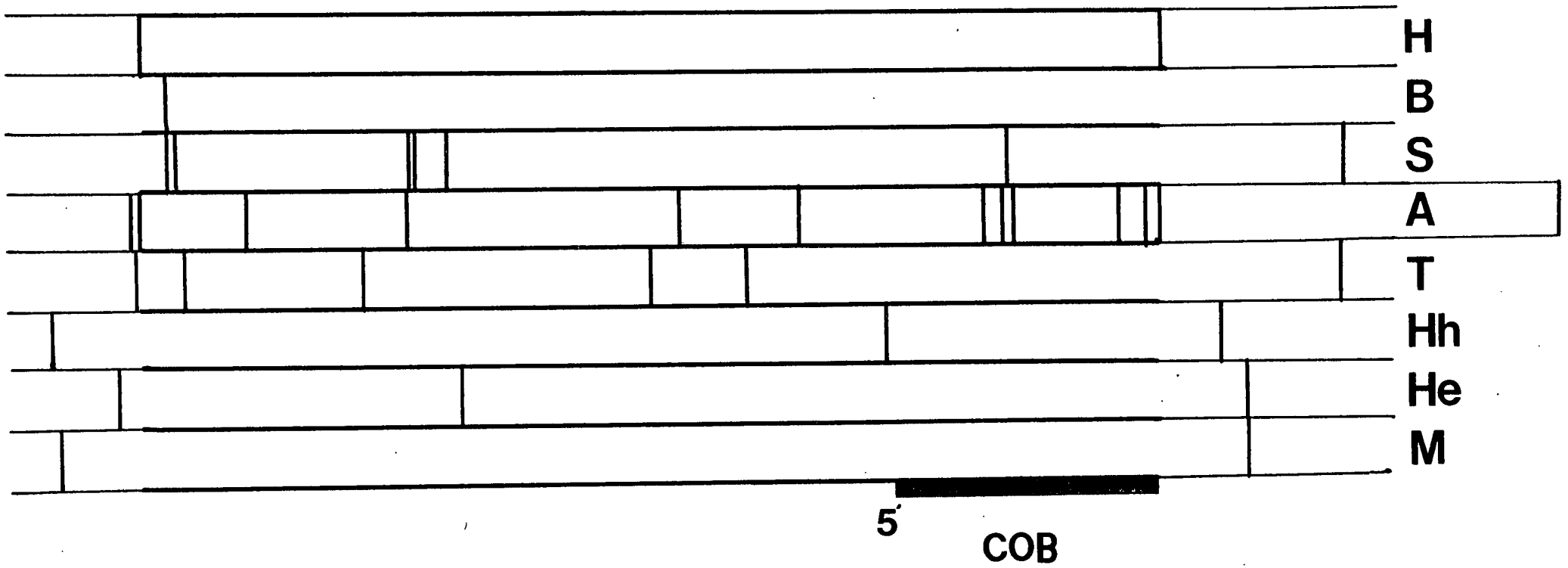
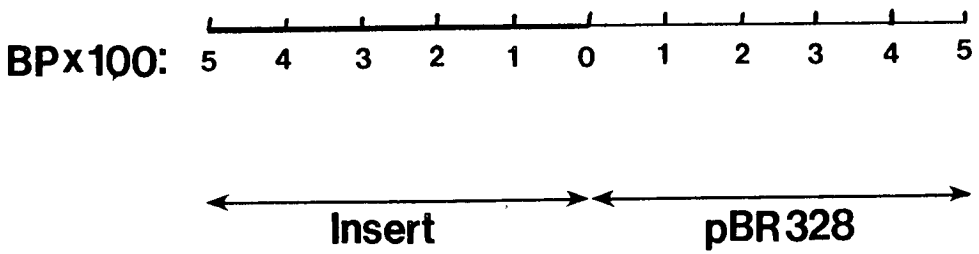


Fig 4.6b



The restriction map shows that almost all the potential sites for M13 subcloning are widely separated, the lack of all MspI sites being especially remarkable. Sequence analysis of pZmH1790, and mapping of DNA sequences surrounding it in the mt genome, has shown (see below) that this restriction site 'asymmetry' is due more to chance than to any unusual sequence usage. The main consequence is that M13 sequencing was made more difficult by the inconvenient location of suitable restriction enzyme sites for subcloning.

4.2.2 Sequence analysis of the COB gene

4.2.2.1 Sequence analysis of HindIII subclones of pZmH1790.

pZmH1790 and M13 mp8 RF DNA were restricted with HindIII and ligated as described in Chapter 2. After transformation of E.coli JM101, six white plaques designated clones 98-103 were selected and phage DNA was transferred to nitrocellulose by the method of Benton and Davis (1977). These clones were screened by hybridisation with nick-translated pBR328 and pZmH1790 insert DNA. The autoradiograph (Fig. 4.7) shows that all six clones contain DNA homologous to the HindIII insert in pZmH1790, but not to pBR328. Four clones were picked at random and sequenced by the dideoxy chain termination reaction of Sanger et al (1977). Two classes of sequence were obtained, corresponding to opposite cloning orientations of the insert in the vector. Clones 98, 99 and 103 yielded identical sequences and contained a BglII site ($^{5'}$AGATCT) precisely 62 BP from the HindIII site ($^{5'}$AAGCTT) (Fig. 4. 8a). However, translation of this DNA sequence in all three reading frames and in both orientations yielded no amino acid sequence with discernible homology to that predicted from any published COB DNA sequence.

The DNA sequence of clone 101 (Fig. 4.8b) was translated in a similar manner. A reading frame was identified specifying a polypeptide

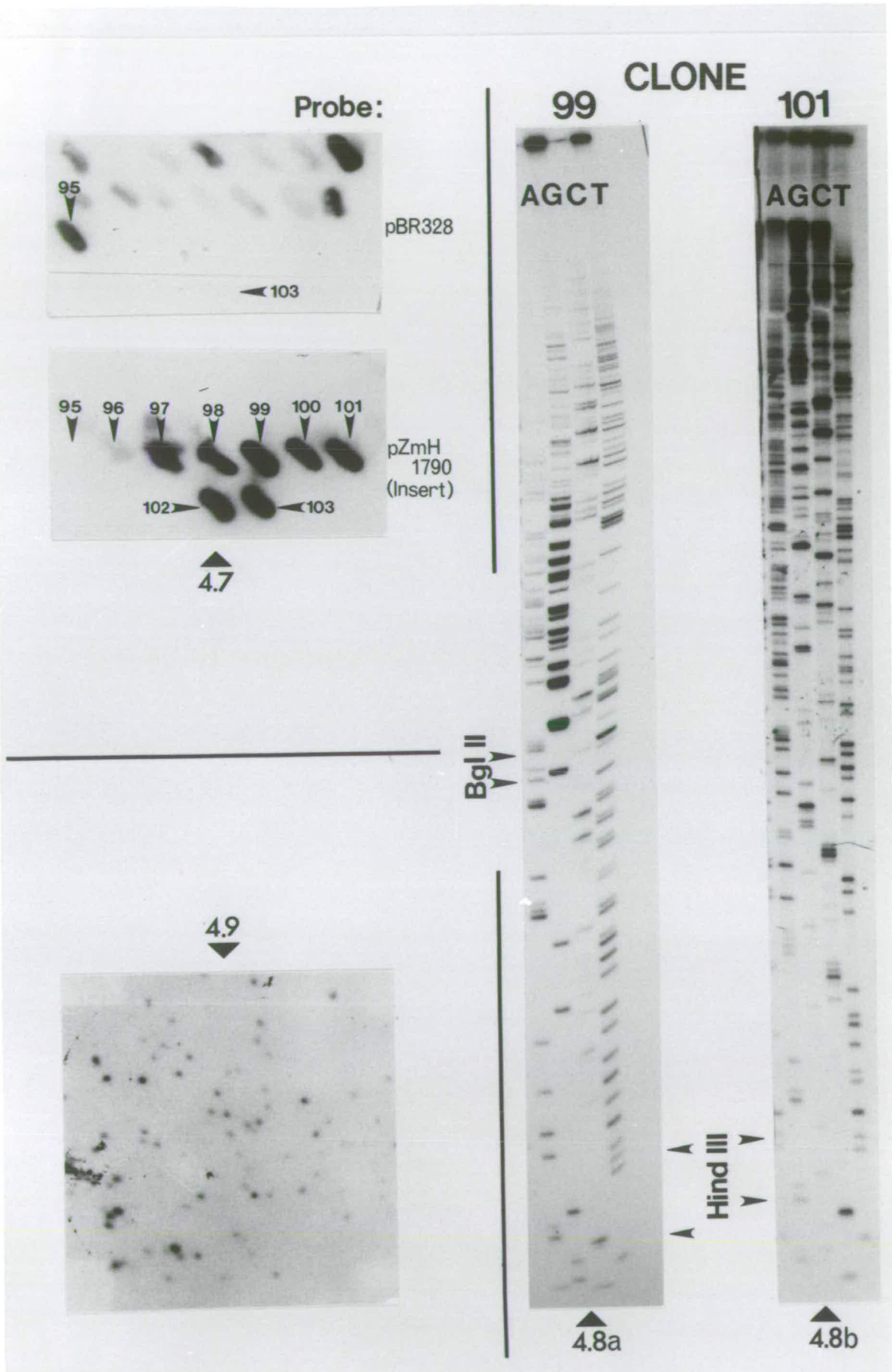
Fig 4.7 Identification of recombinant M13 mp8 clones containing the 1.79KB HindIII-bounded mt DNA insert by plaque hybridisation.

Fig 4.8 Sequence analysis from the HindIII sites bordering the 1.79 KB mt DNA fragment.

4.8a Clone 99, showing the BglII site 62BP from the HindIII site.

4.8b Clone 101, which contains the same insert as clone 99 but in the opposite orientation. When read from bottom to top, the autoradiograph yields sequence complementary to that of the 5' portion of the maize COB gene.

Fig 4.9 Autoradiograph of an in situ plaque hybridisation experiment used for the identification of Sau3A subclones of pZmH1790. The probe was pZmH1790 (insert) DNA.


CLONE
Probe:
99
101
AGCT
AGCT
95
pBR328
103
95 96 97 98 99 100 101
pZmH
1790
(Insert)
102
103
4.7
Bgl II
4.9
Hind III
4.8a
4.8b

with high sequence homology to the amino terminal regions of the fungal and mammalian COB gene products. The following conclusions could therefore be drawn from these preliminary sequencing experiments.

1) The 5' end of the maize COB gene is located within pZmH1790, approximately 450 BP from the nearest HindIII site (assuming no intervening sequences and that the coding sequence is the same length as in other organisms studied to date).

2) This HindIII site falls within the gene, thus more than half of the predicted gene lies outside the clone. This was surprising since the HindIII fragment cloned in pZmH1790 hybridised much more strongly to heterologous probe DNAs than did any other HindIII fragment. Possible explanations are discussed in section 4.2.3.1.

3) The gene contained no introns in the 300 BP region sequenced. Interestingly, this region contains introns I_1 and I_2 in a yeast 'long' strain (Lazowska et al, 1980) and intron 1 in N.crassa (Citterich et al, 1983).

4) The single BglII site in pZmH1790 is located 62 BP from the HindIII site and lies outside the COB coding region. This unique site is useful in determining the orientation of COB with respect to the cloning vector.

4.2.2.2 Sequence analysis of tetranucleotide-recognising enzyme subclones of pZmH1790.

The complete sequence of the HindIII insert in pZmH1790 was determined in order to verify the location of the 5' terminus of the gene and to determine the amino-terminal sequence of the predicted apocytochrome b polypeptide. Tetranucleotide-recognising enzymes were used to produce small, overlapping subfragments which were cloned in M13 mp8. Analysis of these clones allowed internal sequences to be determined which could not be reached by sequence analysis from the HindIII ends. The lack of MspI sites in the pZmH1790 insert limited

the generation of cohesive-ended sub-fragments compatible with hexanucleotide cloning sites in M13 mp8 to the restriction enzymes Sau3A and TaqI. Blunt-ended fragments produced by digestion with AluI were also cloned into the SmaI site of mp8 but under the ligation conditions used, insertion of multiple DNA fragments occurred with high frequency. This complicated the analysis of plaque hybridisation data. Furthermore, TaqI proved difficult to inactivate so fewer clones were obtained with this enzyme. Most of the sequence of pZmH1790 was therefore determined from Sau3A subclones but AluI and TaqI subclones provied useful confirmatory and overlapping sequence data.

M13 cloning and DNA sequence analysis was performed as described in Chapter 2. Recombinant clones carrying the desired inserts were detected by plaque hybridisation which was carried out either after picking selected plaques onto a grid, or in situ directly after transformation. The latter approach was particularly fruitful since when picking plaques by eye, one class of Sau3A insert (opposite orientations of this fragment are represented in clones 7 and 885) was very heavily over-represented. In situ hybridisation allowed less frequently occurring clones to be detected with ease. Fig. 4.9 is an autoradiograph showing the results of a typical in situ hybridisation experiment. The cloning strategy for determination of pZmH1790 sequences is depicted in Fig. 4.18 and the complete HindIII insert sequence is shown in Fig. 4.17.

A number of features emerge from the complete sequence analysis pZmH1790:

1) The length of the HindIII insert in the clone is 1,787 BP which is in remarkably good agreement with the predicted size of 1.79 KB.

2) The deduced restriction map is similar in general terms to that constructed from multiple restriction enzyme digestion of pZmH1790. The main omission is the Sau3A site at position 304 (Fig. 4.17).

3) An open reading frame extends from an ATG codon at nucleotide 1554 to the HindIII site at nucleotide 2015. On the basis of the homology between the predicted amino acid sequence and that of apocytochrome b in other eukaryotes (see Fig. 4.19), this open reading frame probably encodes the maize COB gene.

4) This open reading frame appears to be uninterrupted by introns although homology with other apocytochrome b sequences is poor at the amino terminus, so the possibility that an intron occurs in this region cannot be excluded. The ATG codon starting at nucleotide 1554 would specify a polypeptide with an amino-terminal extension of about six residues when aligned with other apocytochrome b sequences.

4.2.2.3 Characterisation of further COB clones.

In order to determine the DNA sequence of the 3' portion of the COB gene, clones overlapping pZmH1790 were isolated and then used to construct suitable 'downstream' subclones in M13.

'Heterologous' hybridisation with various COB probes (Fig. 4.2) to BamHI and EcoR1 digested mt DNA identified COB homologous fragments of 21 KB and 5.8 KB respectively. The EcoR1 (Chapter 3) and BamHI (P. Isaac) libraries were therefore probed with nick-translated pZmH1790 insert DNA (prepared by electroelution) in order to detect clones containing these DNA fragments.

After hybridisation under the 'stringent' conditions described in Chapter 3, filters were autoradiographed and two clones, designated pbB1 and pbB2 were isolated from the BamHI library (Fig. 4.10). Similarly, one clone (pbE4) in the EcoR1 library hybridised. Plasmid

Fig 4.10 Identification of BamH1 clones containing the maize COB gene.

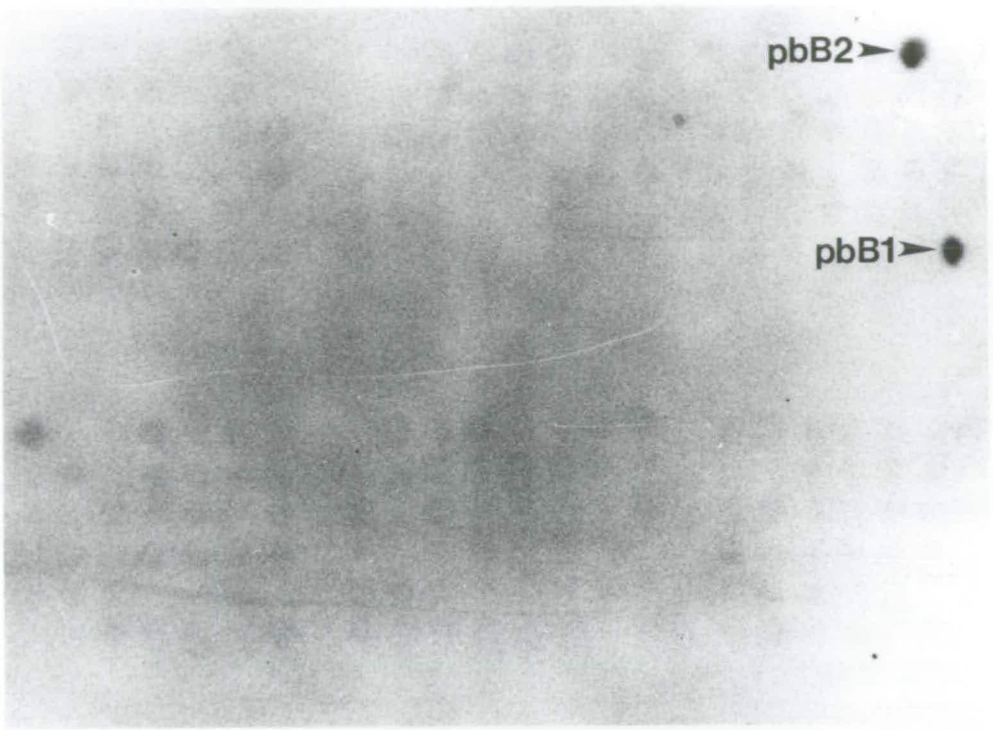
Replica filters of the BamH1 library were probed with nick-translated pZmH1790 (insert) DNA and autoradiographed. Two strongly hybridising clones (pbB1 and pbB2) were identified. See Chapter 3 for details.

Fig 4.11 Characterisation of pbB1, pbB2 and pbE4.

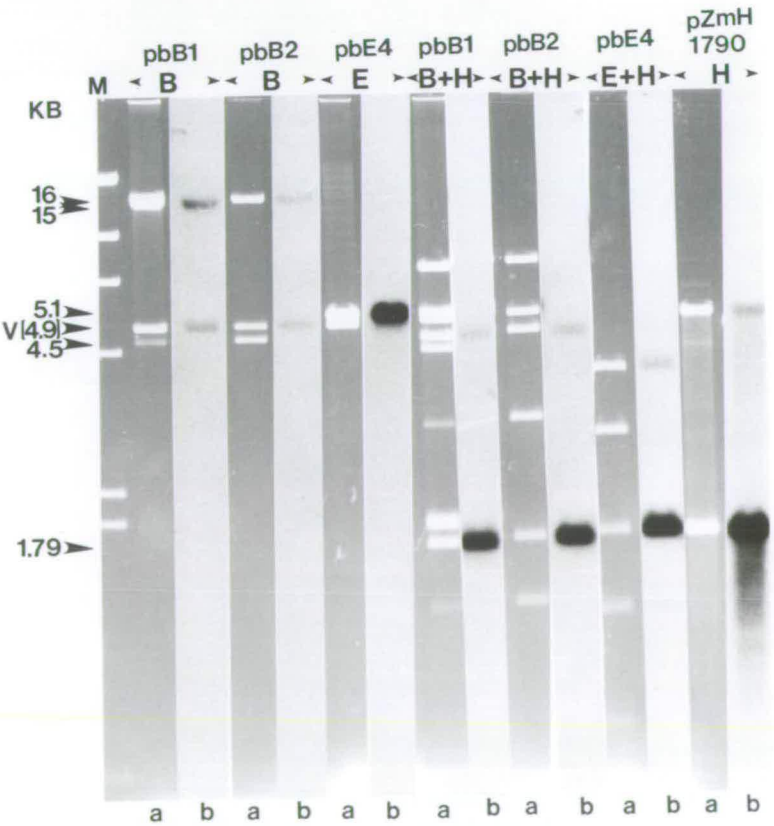
Digestion of pbB1 and pbB2 reveals the presence of a 16KB insert in each case. Additionally, pbB1 and pbB2 contain second inserts of 15 and 4.5KB respectively. The faint bands in pbB1 arise from cross-contamination with pbB2 DNA. Digestion of pbE4 with EcoRI shows it contains a single insert of 5.1KB. Double digestion of pbB1 and pbB2 with BamH1 and HindIII, and of pbE4 with EcoRI and HindIII shows that all three contain a 1.79KB HindIII fragment which co-migrates with the insert in pZmH1790.

4.11a Ethidium bromide electrophoretograms.

4.11b Autoradiographs of the DNA after transfer to nitrocellulose and hybridisation with pZmH1790 (insert) DNA. The faintly hybridising fragments are vector-derived and are identified because the probe contained a small amount of pBR328 DNA.



4.10



4.11

B = Bam HI
E = Eco RI
H = Hin dIII

DNA was prepared from these clones, restricted with the appropriate enzyme and electrophoresed on a 1% agarose gel to assess insert size. Fig. 4.11a shows that pbB1 contains a double insert (16 KB and 15 KB) as does pbB2 (16 kb and 4.5 KB). The larger fragment in each case was shown by hybridisation (Fig. 4.11b) to contain COB homologous sequences and is assumed to be identical in pbB1 and pbB2. This fragment is smaller than that identified by 'heterologous' hybridisation of COB probes to BamHI digested mt DNA (21 KB) indicating either errors in sizing or DNA rearrangements in the clone. Digestion of pbE4 with EcoR1 revealed that this clone contained a single 5.1 KB fragment homologous to COB (Fig. 4.11), in reasonable agreement with heterologous hybridisation data. Digestion of these clones with HindIII showed all three contained a 1.8 KB fragment which co-migrated with, and hybridised to, the HindIII insert in pZmH1790 (Fig. 4.11). Hence the DNA rearrangements, if any, in pbB1 and pbB2 are likely to occur some distance from the COB gene.

Restriction maps of pbB2 and pbE4 were constructed (Fig. 4.12) by multiple restriction enzyme digestions using HindIII, EcoR1 and BamHI. Additionally, BglII sites in pbE4 were mapped so the orientation of the 1.8 KB HindIII fragment in this clone could be ascertained. These results showed that the smaller (ca. 0.66 KB) EcoR1-HindIII fragment should contain the 3' portion of the COB gene.

4.2.2.4 Sequence analysis of the small EcoR1-HindIII subfragment of pbE4.

In order to obtain DNA sequence from the 3' region of COB, pbE4 was double digested with EcoR1 and HindIII and ligated to double digested M13 mp8 and M13 mp9. Ligation mixtures were used to transform E.coli JM101 and were plated out as described in Chapter 2. Sixty white plaques resulting from ligation of DNA into mp8 (designated clones 605-664) were picked onto a grid and two replica copies were

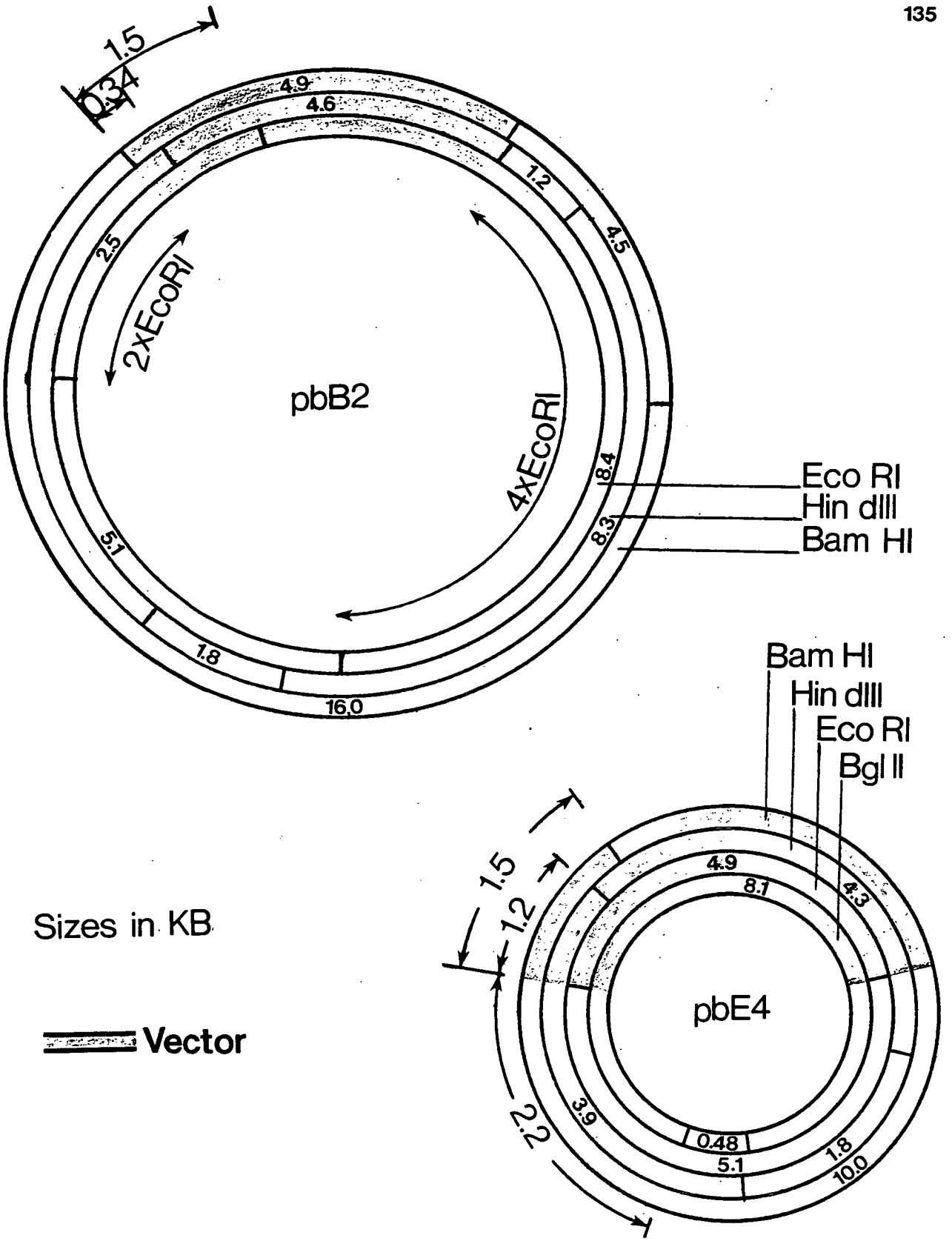


Fig 4.12 Restriction maps of pbB2 and pbE4.

Plasmid DNA was digested with EcoRI, BamHI and HindIII and with combinations of these three enzymes. Fragments were separated by agarose gel electrophoresis and sized by comparison with λDNA marker fragments (not shown), allowing the maps to be constructed.

transferred to nitrocellulose. Plaque hybridisations were performed with these filters using nick-translated pZmH1790 and pBR328 as probes. Three classes of plaque were observed (Fig. 4.13):

1) Those which hybridised to both probes and are therefore derived from EcoR1-HindIII fragments of pBR328;

2) Those which hybridised strongly to pZmH1790 but not to pBR328. Digestion of the RF DNA of one member of this class (clone 652) with EcoR1 and HindIII showed that it contained a 2.2 KB DNA insert corresponding to the larger insert-derived fragment of pbE4 (Fig. 4.14). Sequence analysis of this clone is included in Fig. 4.17;

3) Those which hybridised weakly to pZmH1790 but not pBR328. Digestion of RF DNA (not shown) indicated these clones contained the small EcoR1-HindIII fragment of pbE4. Sequence analysis of one such clone (number 640) confirmed that an open reading frame, specifying a polypeptide homologous to apocytochrome b in other eukaryotes, ran from the HindIII site. The sequence continued directly from the point at which the COB reading frame in pZmH1790 was truncated by the HindIII site.

Sub clones resulting from ligation of DNA into mp9 could not be screened by plaque hybridisation since all clones were found to hybridise to pBR328-derived probes (not shown). This was probably due to a section of pBR322 DNA included in mp9 during its construction (see footnote to Messing and Viera, 1982). Twenty four white plaques (clones 977-1000) were therefore picked and RF DNA was prepared from each. Clones containing the desired EcoR1-HindIII insert (e.g. number 979) were identified (Fig. 4.15) after restriction and electrophoresis in 1% agarose minigels. Sequence analysis of the corresponding single stranded DNA revealed, in the complementary strand, an open reading frame strongly homologous to COB in other eukaryotes.

Fig 4.13 Analysis of M13 mp8 recombinant clones containing EcoRI - HindIII fragments of pbE4.

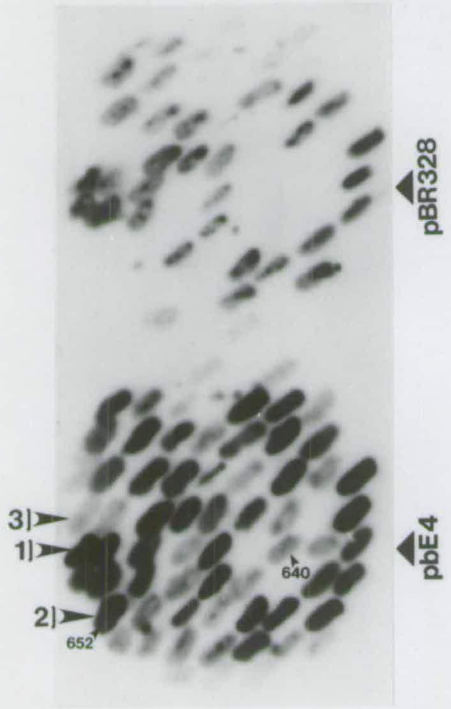
Three types of recombinant were observed on the basis of differential hybridisation with nick-translated pBR328 and pbE4 DNA. Three examples, designated 1), 2) and 3) are shown in the autoradiograph and correspond to the classes described in the text (section 4.2.2.4).

Fig 4.14 Analysis of M13 clone 652.

The ethidium bromide stained agarose gel shows that this recombinant contains the larger (2.2 KB) insert - derived EcoRI - HindIII subfragment of pbE4.

Fig 4.15 Analysis of M13 mp9 subclones of pbE4.

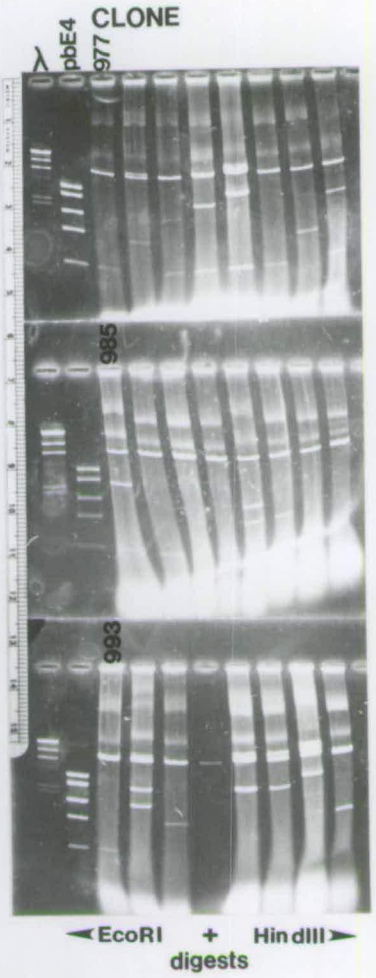
RF DNA was prepared by the 'miniprep' method, digested with EcoRI + HindIII and electrophoresed on 1% agarose minigels. Comparison with EcoRI/HindIII digested pbE4 shows that several clones (eg no. 979) contain the small mt DNA subfragment.


PROBE:
pBR328
pbE4
3
1
2
640
652

4.13


λ Hin dIII
652
pbE4
KB
3.7
2.2
1.8
1.2
0.8
Eco RI
+
Hin dIII

4.14


CLONE
λ
pbE4
977
985
993
EcoRI + Hin dIII
digests

4.15

The following deductions were made from sequence analysis of clones 640 and 979:

1) An intron in the COB coding sequence between the HindIII and EcoR1 sites was extremely unlikely since the distance between conserved predicted amino acids near these sites was, in other organisms, almost exactly the same as the length of polypeptide which would be encoded by the small maize mt DNA EcoR1-HindIII fragment.

2) No in-frame termination codons were found up to the EcoR1 site so further experiments were required to isolate clones extending beyond this point.

4.2.2.5 Sequence analysis of MspI subclones of pbE4

In order to verify sequence data from clones 640 and 979, pbE4 was digested with MspI and subcloned into the AccI site of M13 mp8. Clones which hybridised to pbE4 but not pBR328 were sequenced. Two (numbers 564 and 572) yielded sequence which overlapped with that determined from clones 640 and 979 and also resolved sequence ambiguities in the centre of the fragment (see Fig. 4.17). Sequence analysis of the small HindIII-EcoR1 fragment downstream from the 1,787 BP HindIII fragment cloned in pZmH1790 was therefore completed. This fragment was shown to be 680 BP long and contain a continuous open reading frame homologous to COB from other organisms. The fragment did not, however, extend to the 3' terminus of the gene.

4.2.2.6 Sequence analysis of Sau3A subclones of pbB2.

The smallest available clone likely to contain the complete 3' region of COB was pbB2. Consideration of the restriction sites near the 3' terminus predicted from earlier sequence analysis (above) indicated that a Sau3A fragment starting 113 BP from the EcoR1 site would span this site and extend DNA sequence analysis beyond it,

Fig 4.16 Identification of M13 mp8 Sau3A subclones of pbB2 containing the 3' terminus of COB.

Approximately 500 white plaques were screened in situ after transfer to nitrocellulose, with nick-translated DNA containing the small EcoRI - HindIII fragment of pbE4. 80 plaques contained sequences homologous to the probe.

possibly to the end of the gene. However, since pbB2 is a large (ca. 23 KB) clone, many Sau3A subclones would be generated, only one of which would contain the desired sequence. In order to overcome this difficulty, approximately 500 white plaques were generated by transformation of E.coli JM101 with Sau3A digested pbB2 ligated to BamHI digested M13 mp8. These plaques were screened 'in situ' using the nick translated 680 BP EcoR1-HindIII fragment of pbE4, isolated by electroelution, as a hybridisation probe. Of the 80 plaques which hybridised (Fig. 4.16), 29 (designated 1421-1449) were sequenced. Most contained sequences from within the EcoR1-HindIII fragment and therefore yielded useful confirmatory data. One subclone (number 1445) started at a Sau3A site outside the previously determined sequence, spanned the EcoR1 site and continued to the first internal Sau3A site. This clone contained an open reading frame continuous with that interrupted in previous sequences by the EcoR1 site and also an in-frame termination codon ($^{5'}$TAG). This triplet was found at the position expected from comparison of the predicted amino acid sequence specified by this open reading frame with that predicted from the fungal and mammalian COB genes. It is likely therefore that clone 1445 contains the 3' terminus of the maize COB gene (the possible existence of an intron in this region is discussed in section 4.2.3.4.2). If this is the case, then the DNA sequence analysis of the maize COB gene is complete. Fig. 4.17 shows a compilation of all the continuous sequence gathered in sections 4.2.2.1-4.2.2.6 and Fig. 4.18 summarises the sequencing strategy and restriction map used. The predictions from this sequence analysis are discussed below.

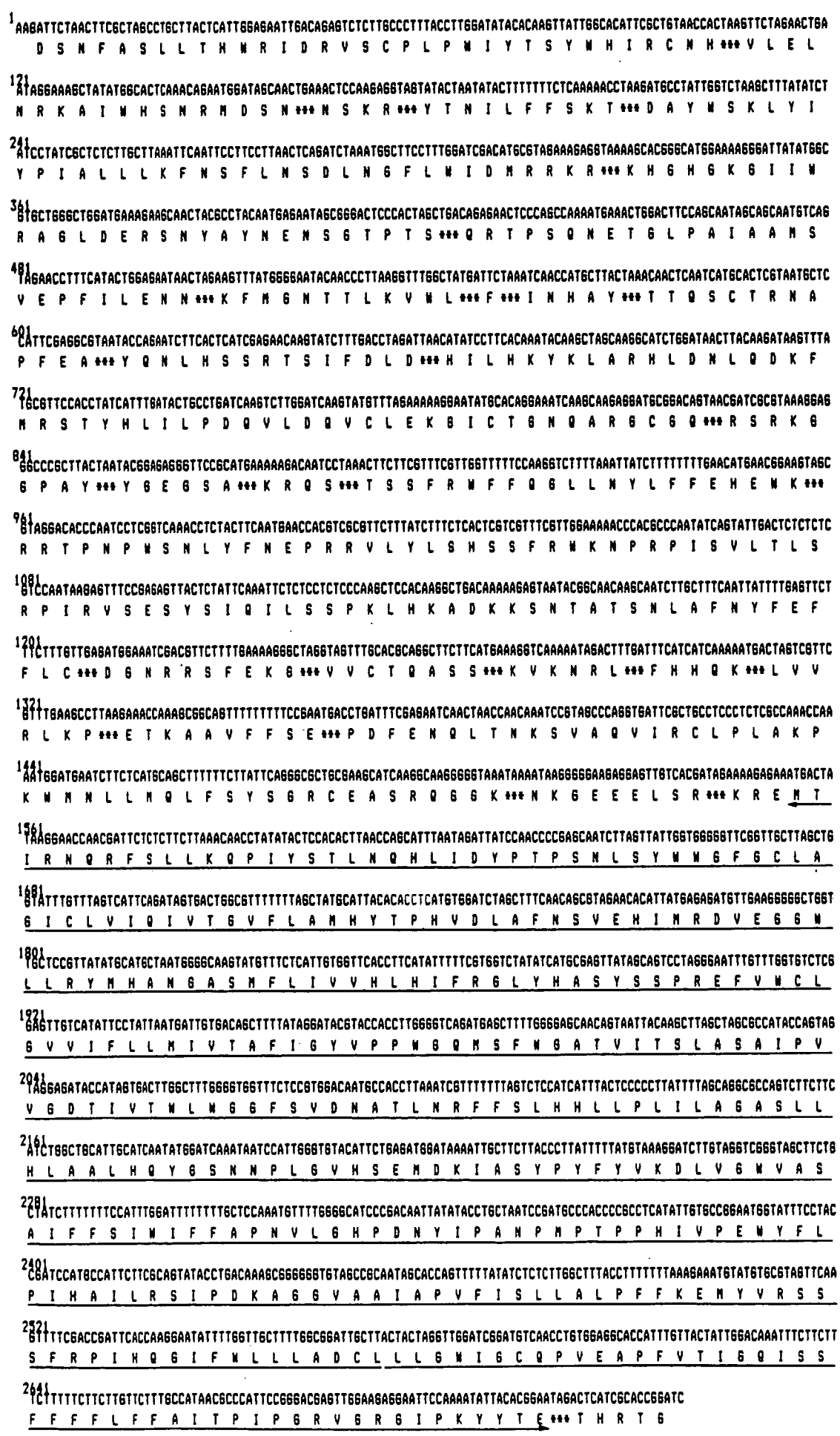
4.2.3 Discussion

4.2.3.1 DNA and amino acid sequence comparisons.

Sequence analysis of the maize COB gene has allowed an amino acid

Fig 4.17 Complete DNA sequence analysis of the maize COB gene and flanking regions. The COB coding sequence is underlined.

COBFRAG



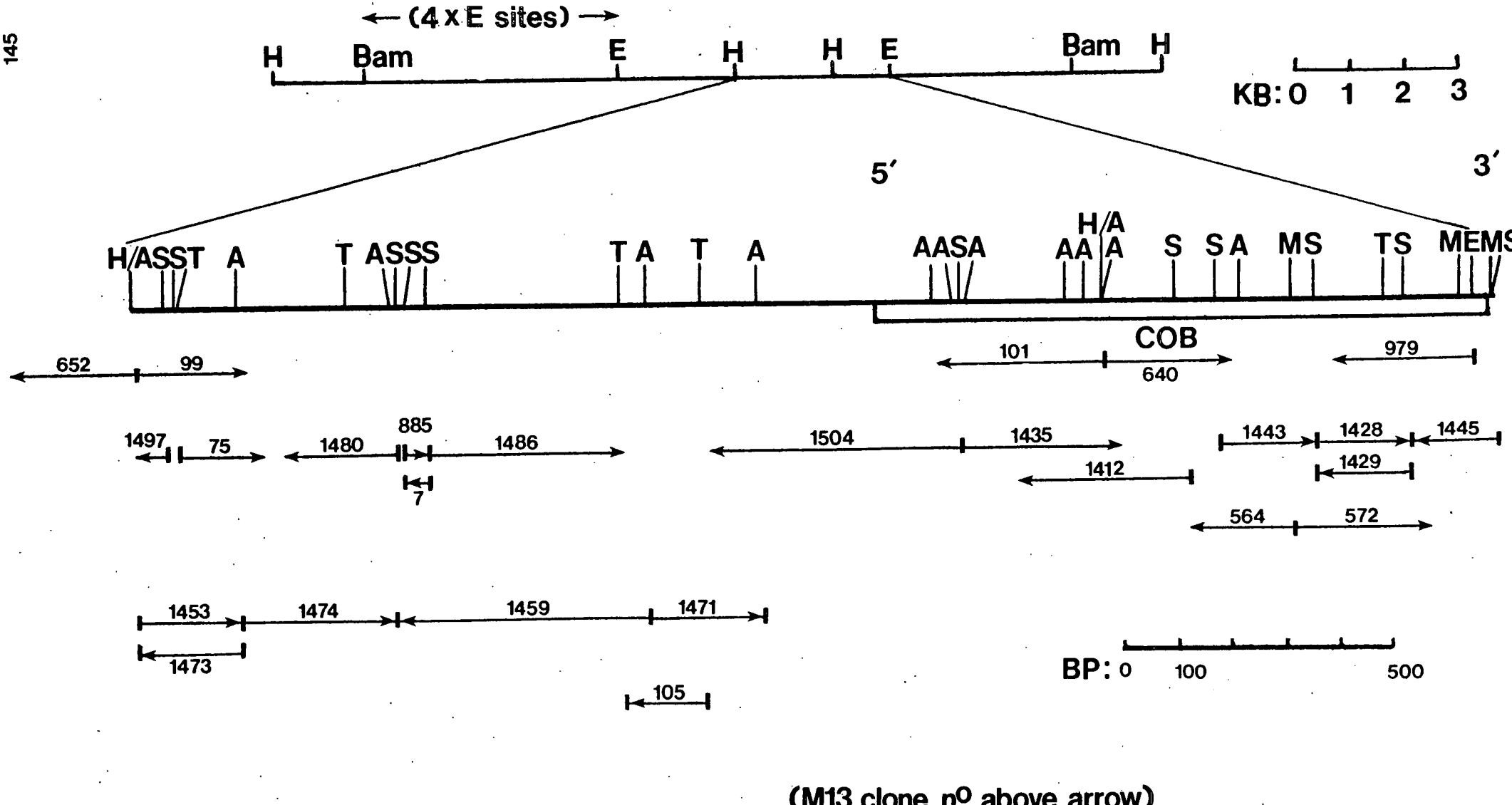


Fig 4.18 Restriction map of the maize COB gene and flanking sequences. Abbreviations are as in Fig 4.6b with the addition of E (= EcoRI).

Arrows indicate the extent of sequence analysis from each restriction site, shown as a short vertical bar.

sequence for the maize apocytochome b polypeptide to be predicted, using the codon assignment suggested by Fox and Leaver (1981) (see also section 4.2.3.2). This sequence is compared with all other known or predicted apocytochrome b sequences in two ways:

1) The amino acid sequences are aligned maximising homology between them (Fig. 4.19); highly conserved sequences can then be detected visually.

2) Fig. 4.20 shows a stepwise comparison of the average number of conserved amino acids in successive blocks of three along the polypeptide, so conserved regions can be identified graphically. Averaging over three residues reduces 'noise' and allows more major trends to be discerned. Superimposed on this is a similar graph depicting the number of amino acids in all other sequences which are identical to the maize amino acid at each position in the polypeptide.

The data in Figs. 4.19 and 4.20 show that several areas throughout the apocytochrome b polypeptide are more highly conserved than others, in particular regions near the maize amino acid residues 56-60, 143-154, 172-177, 277-280 and 286-292. These regions are probably important in the maintenance of higher-order structures within the polypeptide or are involved in the electron transport activities of cytochrome b. The amino acid homology profile including all polypeptides is very similar to the pattern of homology of all other apocytochrome b sequences to the maize sequence (Fig. 4.20). The most significant discrepancies between the two occur near the carboxy terminus of the polypeptide where residues which are conserved in all the other apocytochrome b sequences, are not found in the maize polypeptide. The most striking region includes residues 372-385 where despite an approximate average of 5.5 conserved amino

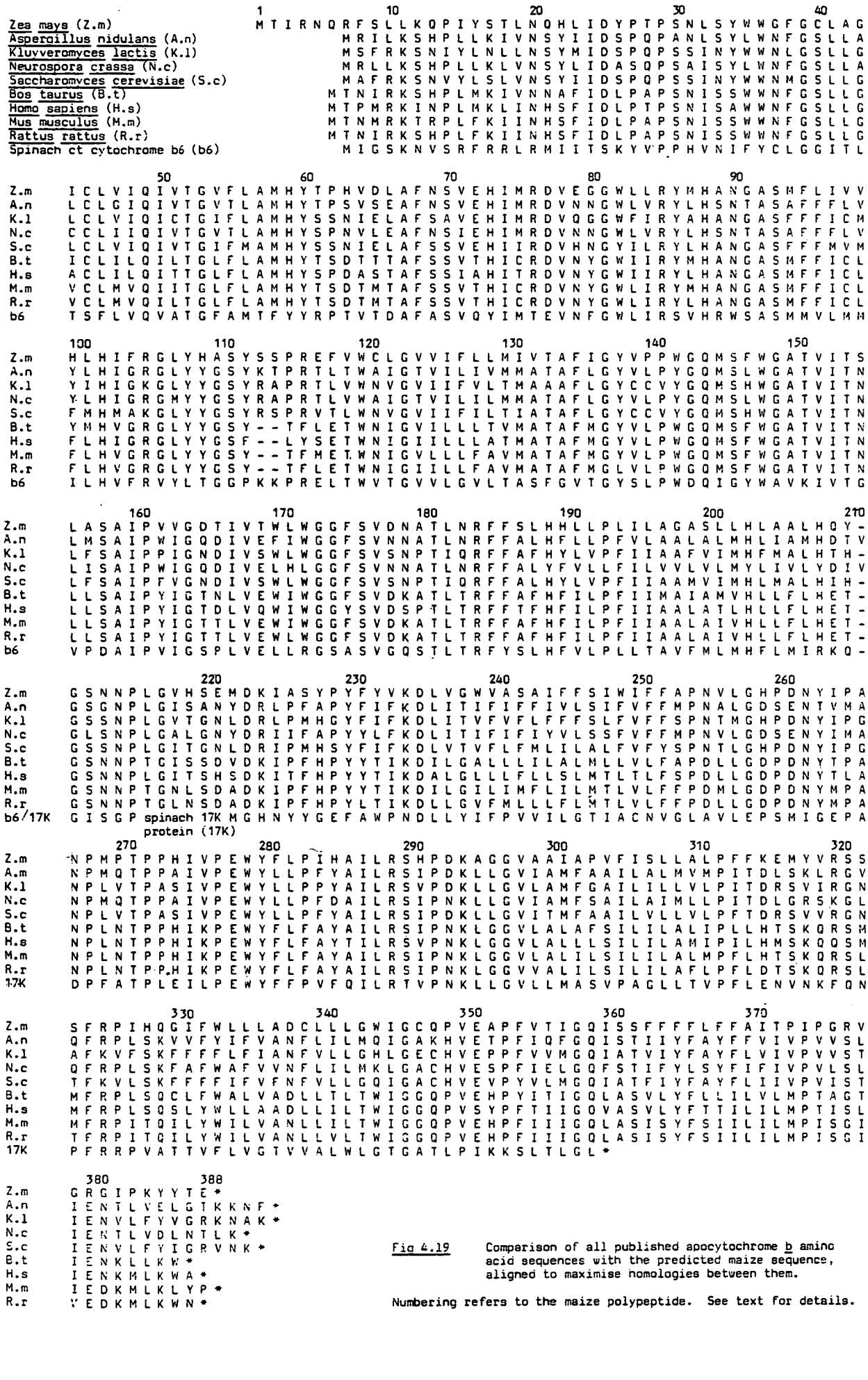


Fig 4.19 Comparison of all published apocytochrome b amino acid sequences with the predicted maize sequence, aligned to maximise homologies between them.

Numbering refers to the maize polypeptide. See text for details.

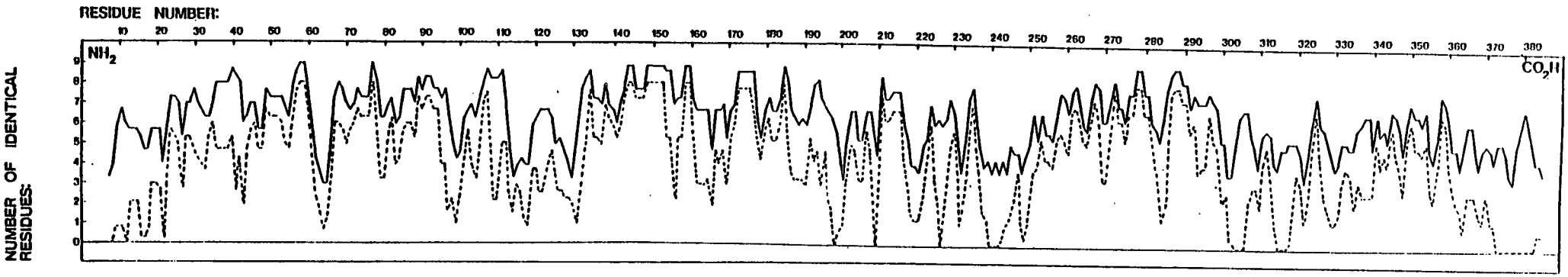


Fig 4.20 Comparison of amino acid sequence homologies within apocytochrome b.

Upper (solid) line: homologies between all sequences (ie number of identical residues at each point).

Lower (dashed) line: homologies to the maize sequence (ie number of sequences containing the same amino acid as the maize sequence at each point).

Each point is averaged over three residues to minimise random fluctuation. Note divergence between upper and lower lines at the amino and carboxy termini.

acids between all nine polypeptides, the maize polypeptide shows no homology at all to any of these sequences. This region falls in an area of low DNA sequence conservation, for example the section of the maize COB gene encoding amino acids 360-388 shares only 24.7% identity with the human, and 32.2% with the yeast COB sequence in the equivalent region. Both of these figures are close to the 25% value predicted for two completely unrelated DNAs. In the same region, the yeast and human DNA sequences share a 51.8% homology. The equivalent figures for DNA homologies over the previous 90 nucleotides are 56.7, 52.2 and 45.6% between maize and man, maize and yeast and man and yeast respectively.

The significance of this area of low DNA and amino acid conservation is not clear. It is interesting that an extremely T-rich sequence (18 nucleotides in 24), encoding a long tract of phenylalanine residues in positions 362-369, occurs near the start of this region. Such sequences have been found in other COB genes (for example yeast; Nobrega and Tzagoloff, 1980) although not at this position. It is tempting to speculate that this region of low homology does not in fact form part of the maize COB gene and is an intron, with the T-rich sequence in some way being involved in its processing. This possibility is discussed in section 4.2.3.4.2. It should be noted that if a single amino acid gap is introduced into the alignment of the maize sequence between residues 362 and 363, amino acid homology for the remainder of the gene is slightly increased. Perhaps the most important consequence of this would be the alignment of the proline residue (374) with that occurring in all other apocytochrome b sequences at this point.

The amino terminal section of the maize polypeptide also displays low homology to that of other cytochromes b. However, the degree of

Table 4.1 Sequence homologies between maize mitochondrial genes and their counterparts in S. cerevisiae and H. sapiens.

		GENE		
		COB	COI	COII
DNA homology to:	S. cerevisiae	52	58	56
	H. sapiens	47	58	52
Predicted amino acid homology to:	S. cerevisiae	48	60	47
	H. sapiens	47	67	40

Values are percentages.

conservation between all sequences is low in this region so the divergence probably simply reflects a lack of strong structural and functional constraints on the maintenance of any particular sequence.

Table 4.1 summarises DNA and predicted amino acid sequence homologies between maize mitochondrial genes and their counterparts in man and yeast. The COB gene and the protein it encodes do not appear from this data to have strong affinities to either the fungi or the mammals. This is also generally true for the COI and COII genes. Attempts to deduce evolutionary relationships between the mitochondrial genes on the basis of sequence comparisons alone are therefore likely to be of limited value. A more useful approach could lie in an examination of gross gene structure. For example, the amino acid sequences derived from the COB genes of various eukaryotes can be grouped according to the presence or absence of a short gap (at residues 113 and 114) which is included to maximise sequence homology between the polypeptides. Fig. 4.19 shows that maize and all the fungi contain two extra amino acids at this position which are absent in the mammals. The fungi in turn are subdivided by the presence in Neurospora and Aspergillus of a single amino acid insertion between residues 210 and 211. These findings could suggest a slightly closer evolutionary relationship between higher plant and fungal mitochondria than between those of plants and mammals. This would be consistent with known or assumed phylogenetic relationships but much more sequence data will be required before parameters such as these can be used as quantitative indices of evolutionary divergence.

Recent studies (Widger et al, 1983; Hauska et al, 1983) on the chloroplast b_6f complex, which appears to be functionally equivalent to the mitochondrial bc_1 complex (Nelson and Neumann, 1972), have

revealed that the amino acid sequence of the cytochrome $\underline{b}_6$ subunit in spinach has significant homology to that of fungal and mammalian apocytochrome $\underline{b}$. Fig. 4.19 includes the sequence of this subunit from spinach, aligned to maximise homology with the mitochondrial sequences. Additionally, the figure shows an alignment of the amino acid sequence of a second component of the $\underline{b}_6\underline{f}$ complex, the m_r = 17,000 polypeptide (known as the 17 K protein), as determined by Heinemeyer and Herrmann (1983, in preparation. See also Widger *et al*, 1983). Both cytochrome $\underline{b}_6$ and the 17 K polypeptide are encoded in the chloroplast genome (Widger *et al*, 1983; also J. Gray (pers. commun.)). Three conclusions can be drawn from this comparison:

1) The maize apocytochrome $\underline{b}$ sequence has homology with both the $\underline{b}_6$ and 17 K protein sequences. The $\underline{b}_6$ sequence corresponds to the 5' region of the maize gene, the 17 K protein to the 3' region with a gap of six amino acids between the two. The homology is not significantly higher than between the fungal or mammalian apocytochromes $\underline{b}$ and these proteins. This is perhaps not surprising since there is no reason to suppose that chloroplasts should be any more closely related, in terms of the DNA sequences they contain, to higher plant mitochondria than to those from fungi or mammals.

2) Widger *et al* (1983) identify four 'invariant' histidine residues which are found in all the $\underline{b}$ cytochromes examined. They propose a model in which these amino acids function as ligands for protohaem binding (see section 4.2.3.3.2). Maize apocytochrome $\underline{b}$ appears to contain all these residues in exactly the predicted positions, hence the same model could be applied to explain protohaem binding in this higher plant polypeptide.

3) The maize apocytochrome $\underline{b}$ polypeptide contains a methionine residue which coincides exactly with the position of the initiator

methionine in the aligned 17 K protein. None of the other mitochondrial apocytochromes *b* contain a methionine at this position. Whether this is purely coincidence or of more fundamental genetic significance is unknown. It is interesting, however, that the b_6 and 17 K protein genes are adjacent on the chloroplast genome (R. Herrmann, unpublished observation, see also Hauska *et al*, 1983). An inference could be that these two sequences arose from a single progenitor gene, such as encodes the mitochondrial apocytochrome *b*, which was divided by a single intron. The exons could then have devolved into the two separate genes now found adjacent to each other in chloroplast DNA. The existence of a methionine residue at the appropriate position in the maize polypeptide could support this speculation, but perhaps argues more strongly in favour of the converse, i.e. a fusion combining the functions encoded by two separate genes into a single cistron. The maize methionine residue could therefore represent a 'relic' of the progenitor gene. These suggestions are purely speculative at the present time and are difficult to prove. A more comprehensive analysis of mitochondrial *COB* genes from representatives of difference phyla may help to trace the common ancestor of the chloroplast and mitochondrial genes, assuming one exists (or existed). It would also be interesting to study cytochrome *b* genes in aerobic bacteria, a number of which contain respiratory complexes very similar to bc_1 or b_6f (Haddock and Jones, 1977). The electron transporting cytochrome b_{562} from *E.coli* (Itagaki and Hager, 1968) has no sequence homology with the mitochondrial cytochrome *b*, however.

A consideration of DNA sequence homologies reveals why all 'heterologous' *COB* probes hybridise more strongly to the 1.8 KB *Hin*dIII

fragment than to any other HindIII fragment from maize mt DNA, despite the fact that the former contains less than half of the gene. The maize COB gene, from its predicted 5' terminus to the internal HindIII site displays 55%, 57% and 48% sequence homology to the S.cerevisiae, K.lactis and H.sapiens sequences respectively, in this region. The equivalent homologies from the HindIII site to the 3' terminus are 50, 48 and 47%. Between S.cerevisiae and H.sapiens, the sequence homologies in the equivalent 5' and 3' regions are 53 and 48%. 'Heterologous' COB probes which contain DNA sequence equivalent to maize sequences on either side of the HindIII site will therefore hybridise more strongly to the 1.8 KB fragment than to the adjacent fragment. This is particularly marked with the K.lactis clone M13-3.8. The 347 nucleotides from the start (near the 5' terminus of the gene) of the clone to the maize HindIII site display an average 62% homology to the maize sequence whereas the 403 nucleotides from the HindIII site to the end of the cloned fragment are only 52% identical. This relatively small difference in sequence homology is reflected in very differential hybridisation (Fig. 4.2), showing that hybridisation intensity is strongly sequence dependent, with a rapid cut off below 50-55% sequence identity. Under the conditions used, it seems unlikely that DNA fragments with <45% average sequence homology to the probe could be detected. It is possible that short stretches of much higher homology may be required to generate a strong signal.

4.2.3.2 Codon usage.

'Nonstandard' codon usage appears to be a feature of all mitochondrial genomes studied to date. Specifically, these unusual mitochondrial codon assignments are:

1) In mammals (Barrell et al, 1979) and fungi (Fox, 1979b) UGA encodes

Table 4.2 Codon usage in the maize COB gene

First letter	Second letter: A				Second letter: G				Second letter: C				Second letter: U				Third letter
A	AAA	(K)	5		AGA	(R)	2		ACA	(T)	5		AUA	(I)	14	A	
	AAG	(K)	1		AGG	(R)	2		ACG	(T)	2		AUG	(M)	10	G	
	AAC	(N)	3		AGC	(S)	6		ACC	(T)	4		AUC	(I)	3	C	
	AAU	(N)	9		AGU	(S)	9		ACU	(T)	4		AUU	(I)	18	U	
G	GAA	(E)	6		GGA	(G)	11		GCA	(A)	7		GUA	(V)	10	A	
	GAG	(E)	2		GGG	(G)	5		GCG	(A)	3		GUG	(V)	9	G	
	GAC	(D)	3		GGC	(G)	3		GCC	(A)	6		GUC	(V)	2	C	
	GAU	(D)	7		GGU	(G)	10		GCU	(A)	12		GUU	(V)	9	U	
C	CAA	(Q)	7		CGA	(R)	3		CCA	(P)	3		CUA	(L)	6	A	
	CAG	(Q)	3		CGG	(W)	1		CCG	(P)	7		CUG	(L)	1	G	
	CAC	(H)	2		CGC	(R)	1		CCC	(P)	4		CUC	(L)	6	C	
	CAU	(H	14		CGU	(R)	4		CCU	(P)	10		CUU	(L)	12	U	
U	UAA	(*)	0		UGA	(*)	0		UCA	(S)	2		UUA	(L)	11	A	
	UAG	(*)	1		UGG	(W)	12		UCG	(S)	0		UUG	(L)	8	G	
	UAC	(Y)	5		UGC	(C)	2		UCC	(S)	3		UUC	(F)	10	C	
	UAU	(Y)	12		UGU	(C)	3		UCU	(S)	3		UUU	(F)	21	U	

(One letter amino acid translations are shown in brackets.)

tryptophan and is not a termination codon. The 'standard' tryptophan codon UGG is rarely used.

2) In the yeasts (Li and Tzagoloff, 1979), but not in Aspergillus, Neurospora or the mammals, the CUN family encode threonine rather than leucine.

3) The standard isoleucine codon AUA encodes methionine in mammals and yeast (Anderson et al, 1981; Hudspeth et al, 1982). In addition, the triplet AUU (encoding isoleucine) is occasionally used as an initiator codon in mammals (Anderson et al, 1981),

4) In mammals, the triplets AGA and AGG are avoided, and are probably used as termination codons (Barrell et al, 1980).

The sequence analysis of the maize COB gene, and comparison with that of COII (Fox and Leaver, 1981) and COI (P. Isaac, pers. commun.) has allowed the following conclusions to be drawn on codon usage in maize mitochondria (codon usage in the maize COB is shown in Table 4.2):

1) Twelve UGG triplets occur in the COB coding sequence. Eleven of these are located in positions where the amino acid tryptophan is conserved in at least four of the eight other apocytochrome b sequences which have been determined. In all these 71 cases, the tryptophan residue in the other mitochondrial protein is specified by the codon UGA. This strongly suggests that UGG is the major tryptophan codon in maize mt DNA and that UGA could be a termination codon. No plant mt gene terminated by UGA has yet been sequenced, however. These findings support the similar conclusions drawn from COI and COII sequence analyses, and analysis of the Oenothera COII gene (Hiesel and Brennicke, 1983) has indicated a similar codon usage in this species.

2) A single CGG codon occurs in the gene, encoding amino acid residue 239. This has been translated as tryptophan in view of the findings of Fox and Leaver (1981). This region of the polypeptide is rather poorly conserved and no tryptophan residues occur in other

apocytochrome b polypeptides at this point. However, the 'hydropathy' of tryptophan is -0.9 (Kyte and Doolittle, 1982) which more closely resembles that of the other amino acids at this point (hydropathies are 1.8 to 4.5) than does that of arginine (-4.5), the amino acid predicted from the standard genetic code.

3) No highly conserved amino acids in the predicted apocytochrome b sequences from the other eukaryotes are consistently substituted by a different amino acid in the maize sequence. Hence it appears that with the exception of CGG encoding tryptophan, maize mt DNA uses a 'standard' genetic code.

4) The codon usage table shows a bias in favour of triplets ending in U (41.7%). A, G and C are used less frequently in this position (24.9, 17.2 and 16.2% respectively). The nucleotide usage (in all positions) for the COB gene is U = 35%, A = 24%, G = 21% and C = 21%. Similar trends were noted in COII (the fractions for U, A, G and C in the third nucleotide position are 37.5, 28.7, 15.6 and 18.2%), COI (35.2, 26.6, 16.9 and 21.3%) and also COII in Oenothera (38.6, 29.7, 13.9 and 17.8%). This resembles neither the fungi, where U and A predominate (Waring et al, 1981) nor the mammals, where A and C predominate (Anderson et al, 1981). This consistent feature could prove useful in the identification of genuine protein-encoding unidentified reading frames (URFs) in maize mt DNA.

The overall G + C content of the gene is 41.2%, which is lower than that of maize COII (48.2%) but considerably higher than that of COB in S.cerevisiae (25.4%, Nobrega and Tzagoloff, 1980).

5) The gene is terminated by the codon UAG. Most of the other sequenced COB genes end with UAA, except Bos taurus, which uses AGA. COII is also terminated by UAA in maize.

6) The predicted M_r of the maize apocytochrome b polypeptide, based on this codon usage, is 42,868.

4.2.3.3 Higher order structures in maize apocytochrome b.

4.2.3.3.1 Identification of hydrophobic domains. In order to investigate the organisation of maize apocytochrome b in the inner mitochondrial membrane, a 'hydropathy' profile was constructed using the indices of Kyte and Doolittle (1982). Hydropathy values for blocks of eleven amino acids were calculated and the summed value plotted above the position of the central amino acid in the block (Fig. 4.21). The block was successively displaced by one amino acid throughout the length of the polypeptide from amino to carboxy termini. Taking summed values for blocks in this way emphasises the presence of 'domains' with predominantly hydrophobic (positive hydropathy) or hydrophilic (negative hydropathy) characteristics, but reduces the effects of short term variation.

A drawback of this approach is that the hydropathy values determined do not necessarily allow the secondary structures (β pleated sheet or α helix) of the domains to be predicted. In addition, quantitative measurements of free energy changes during the transfer of the peptide from an aqueous to a non-aqueous environment cannot be made. Nevertheless, this plot highlights nine or ten hydrophobic 'domains' in the polypeptide, which are separated from each other by more hydrophilic regions. Each of these domains contains approximately 20-30 amino acids. If the polypeptide assumes an α helical structure in a hydrophobic environment to maximise internal hydrogen bonding (Chothia, 1976) and if each residue advances the helix by 0.15 nm (Dickerson and Geis, 1969) then each hydrophobic domain would form an α helix 3-4.5 nm long. This is similar to values accepted for the thickness of the lipophilic phase of a normal biological membrane (Tanford, 1978), suggesting that the maize apocytochrome b could span the inner mitochondrial membrane in nine α helical regions.

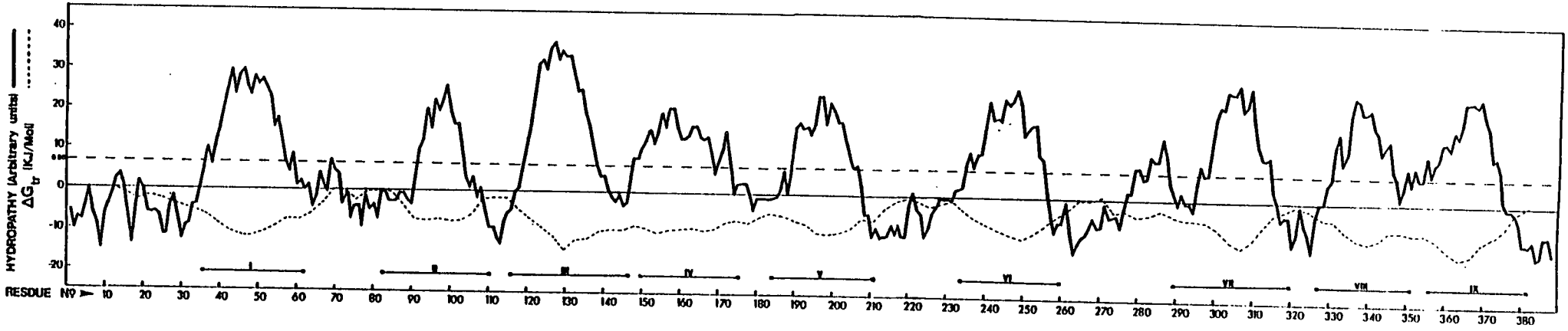


Fig 4.21 Hydrophobicity profile of the predicted maize apocytochrome b polypeptide.

Solid line: Hydropathy index of Kyte and Doolittle (1982). Hydrophobic domains (marked I - IX) have values >0. Averaged over blocks of 11 residues.

Dashed line: Hydrophobicity according to von Heijne and Blomberg (1979). Values are ΔG_{tr} for transfer of an amino acid from an aqueous to a lipophilic environment. Hydrophobic domains have values <0. Averaged over blocks of 21 residues.

Saraste and Wikstrom (1983) and Widger et al (1983) have conducted similar studies on apocytochrome b in a variety of fungi and mammals. Both groups identified nine hydrophobic domains which are marked I to IX in Fig. 4.21. These domains correspond closely with the putative hydrophobic regions of the maize polypeptide. Even at a detailed level, the fine structure of the hydropathy profile is remarkably well conserved in mammals, yeast and maize. The maxima of the hydrophobic peaks are almost identical in all organisms, except that in maize domain VI reaches only 27 units whereas Saraste (1983) finds that in the yeast and bovine polypeptides it is nearer 40 units. This is unlikely to be reflected in any significant difference between the polypeptides, however. The only major discrepancy between the maize and the bovine and yeast profiles is in the calculated value of the mean hydropathy of the eleven-amino acid peptides used in the construction of the plot. Saraste calculates a value of ca. -6 for the bovine and yeast polypeptides, yet in maize the equivalent value is +6.56. It seems that Saraste may have used different criteria for the calculation of this value, since such a large divergence, if genuine, would imply the plant and the fungal and bovine cytochromes b had considerable structural differences.

The otherwise very good correlation between the hydropathy profiles shows that while many amino acid substitutions have occurred in maize, they probably have only minor effects on the secondary (and tertiary) structure of the protein. It is interesting that domain IX is conserved in maize, despite the fact that amino acid homology to both the yeast and bovine polypeptides is <18% over most of this region (residues 360-388). This suggests that these putative membrane-spanning domains are of genuine functional significance and that selection pressures operate to maintain them, even in the face of primary sequence divergence.

Fig. 4.21 also shows the results of a second type of hydrophobicity analysis, using the method developed by Von Heijne and Blomberg (1979). This analysis was carried out by Professor L. Grivell using the computing facilities at the University of Amsterdam. The protocol attempts to calculate the free energy change for the transfer (ΔG_{tr}) of an amino acid in a randomly coiled polypeptide in water, to an α helical polypeptide in a lipophilic phase. The method was originally devised to explain co-translational transfer of proteins across membranes, but has also been used to locate hydrophobic regions in membrane-spanning polypeptides. It has the advantage that quantitative estimates of free energy changes can be made.

The plot is inverted with respect to the hydropathy plot, i.e. hydrophobic regions have $\Delta G_{tr}<0$ and hydrophilic regions have $\Delta G_{tr}>0$. The graph has fewer minor fluctuations because average values for blocks of 20 amino acids, rather than 11, were plotted and because successive blocks were displaced by three residues rather than one. The profile shows the presence of hydrophobic domains, and these coincide closely with domains I to IX identified by the hydropathy plot. The only possible exception is domain IV which is not clearly associated with a minimum in the hydrophobicity profile. Since two different hydrophobicity analyses yield very similar results, it seems probable that the membrane-spanning model for apocytochrome b may be essentially correct. Furthermore, hydropathy plots of the maize, bovine, human and Saccharomyces cerevisiae proteins by the Kyte-Doolittle approach (Saraste, 1983; Widger et al, 1983) or of maize, bovine, Kluyveromyces lactis and Aspergillus nidulans by the Von Heijne-Blomberg method (L. Grivell, pers. commun.) show these domains are conserved in all the polypeptides analysed. Hence, it seems likely that most of the amino acid substitutions in the

predicted maize apocytochrome b sequence are silent in terms of their effect on the secondary structure of the protein.

A short relatively hydrophilic region is predicted by both models to exist in the apocytochrome b polypeptide before the first 'transmembrane' sector. In maize, this region contains about 35 amino acids, in most other apocytochromes b, it is about 29 residues long. The significance, if any, of this region to the assembly of a newly synthesised polypeptide in the membrane is uncertain. Cytochrome oxidase, subunit I in Neurospora (Burger et al, 1982) (and probably Aspergillus; T. Brown, pers. commun.) is believed to be synthesised as a precursor of higher molecular weight, containing an amino terminal extension which is removed by post-translation processing. Similarly, in S.cerevisiae (Sevarino and Poyton, 1980; Pratje et al, 1983), N.crassa (Machleidt and Werner, 1979; van den Boogaard et al, 1982) and probably Z.mays (Fox and Leaver, 1981), the precursor to cytochrome oxidase, subunit II has an amino terminal extension of ca. 15 residues. It is possible that the equivalent region could be processed from the Z.mays apocytochrome b but as yet there is no experimental evidence to support this premise. Proof that it occurs in maize awaits identification of the polypeptide and protein sequence analysis. This identification would certainly be complicated by post-translational processing, if it occurs. Furthermore, in yeast, cytochrome b (predicted M_r from DNA sequence analysis = 44,000; Nobrega and Tzagoloff, 1980) displays a mobility on SDS-polyacrylamide gels equivalent to a protein of M_r between 26,000 and 37,000 depending on the method used for its isolation (Kreike, 1982). Hence, if cytochrome b displays similar behaviour in maize, attempts to identify it in an electrophoretogram of in vitro synthesised mitochondrial polypeptides on the basis of gel mobility alone, are likely to be very unreliable.

4.2.3.3.2 Protohaem binding sites. Recent work (T'sai and Palmer, 1982) has shown that a single cytochrome b subunit of the cytochrome bc_1 complex probably binds two protohaems. The amino acids which act as ligands to these prosthetic groups are not known, but Carter et al (1981) have suggested that histidine residues could be involved. Saraste (1983) and Widger et al (1983) have independently proposed a model in which pairs of 'invariant' (i.e. found in all apocytochrome b sequences) histidine residues separated by 13 amino acids in hydrophobic domains II and V, could act as axial ligands. Widger et al draw additional support for the model from the finding that chloroplast cytochrome b_6, which also binds two protohaem groups, contains histidine residues in exactly these positions. The protohaem groups are inferred to lie between the two trans-membrane α helices (domains II and V), with the haem planes roughly perpendicular to the membrane.

This model is consistent with electron paramagnetic resonance data (Erecinska et al, 1978) but depends on observed 'conservation' of amino acids for its formulation. The maize cytochrome b contains all four 'conserved' histidine residues (at positions 88, 102, 189 and 203; Fig. 4.19), so a similar model could be applied to this polypeptide. However, the amino acids at positions 189 and 203 are not conserved in Neurospora (Citterich et al, 1983) both being replaced by tyrosine residues. It is unlikely that a different mechanism of protohaem binding occurs in this polypeptide, particularly since cytochrome b in Aspergillus, a fairly closely related eumycete, contains histidine residues in these 'conserved' positions. Hence, the haem-binding model may need modification since it is unlikely that the tyrosine residues in Neurospora could act as axial ligands in exactly the same way as the imidazolium function of histidine.

A further criticism of the model is that an 'invariant' pair of phenylalanine residues, near the maize positions 95 and 96 is postulated to be involved in transfer of electrons between the two protohaem groups. In maize, the second of these 'conserved' residues is replaced by leucine (position 96) so it is not clear whether or not a similar mechanism of electron flow could operate.

The model proposed for protohaem binding in cytochrome b is rather preliminary and is not yet backed by solid experimental evidence. The maize protein appears to contain the requisite amino acids in the predicted positions, so if the model is proven, it could function in the proposed manner. Both Saraste and Widger et al attach importance to 'invariant' lysine, arginine or histidine residues at positions 85, 105, 184 and 208 in the formation of salt bridges linking the propionate side-chains of the protohaem groups to the apocytochrome. The apocytochrome b sequence from maize contains these residues in the predicted positions, and would appear, therefore to possess all the features of a biologically functional cytochrome b.

4.2.3.4 Does the maize COB gene contain introns?

The discovery of a 794 BP intervening sequence in the maize COII gene (Fox and Leaver, 1981) set a precedent for the existence of introns in plant mitochondrial genes. In view of this finding, and since many fungal COB genes contain large numbers of introns, the presence of introns in the maize COB gene would not have been unexpected. However, the sequence analysis of this gene revealed no obvious intron(s), and initiator AUG and termination codons are found in positions close to those predicted from alignment with other COB gene sequences. Introns are only possible towards the 5' and 3' ends of the gene where sequence homology of the maize gene to other COB genes is low. These possibilities are examined below.

Table 4.3 DNA and predicted amino acid sequences at the 5' terminus of several COB genes (See text for references)

Splice? (between N and M of the Maize 'spliced' start)

Maize 'spliced' start								M	F	R	K	R	N	M	H	R	K	S
								ATG	TTT	AGA	AAA	AGG	AAT	ATG	CAC	AGG	AAA	TCA
K.lactis start							M	S	F	R	K	S	N	I	Y	L	N	L
							ATG	TCA	TTT	AGA	AAA	TCT	AAT	ATT	TAT	TTA	AAT	TTA
S.cerevisiae start							M	A	F	R	K	S	N	V	Y	L	K	L
							ATG	GCA	TTT	AGA	AAA	TCA	AAT	GTG	TAT	TTA	AAG	TTA
A.nidulans start							M	R	I	L	K	S	M	P	L	L	K	I
							ATG	AGA	ATT	TTA	AAA	AGT	CAT	CCT	TTA	CTA	AAA	ATA
N.crassa start							M	R	L	L	K	S	M	P	L	L	K	L
							ATG	AGA	TTA	TTA	AAA	AGT	CAT	CCT	TTA	TTA	AAA	TTA
B.taurus start						M	T	N	I	R	K	S	H	P	L	I	K	I
						ATG	ACT	AAC	ATT	CGA	AAG	TCC	CAC	CCA	CTA	ATA	AAA	ATT
H.sapiens start						M	T	P	I	R	K	I	N	P	L	I	K	L
						ATG	ACC	CCA	ATA	CGC	AAA	ATT	AAC	CCC	CTA	ATA	AAA	TTA
R.rattus start						M	T	N	I	R	K	S	H	P	L	F	K	I
						ATG	ACA	AAC	ATC	CGA	AAA	TCT	CAC	CCC	CTA	TTC	AAA	ATC
M.musculus start						M	T	N	I	R	K	T	H	P	L	F	K	I
						ATG	ACA	AAC	ATA	CGA	AAA	ACA	CAC	CCA	TTA	TTT	AAA	ATT
Maize non-spliced start	M	T	I	R	N	Q	R	F	S	L	L	K	Q	P	I	Y	S	T
	ATG	ACT	ATA	AGG	AAC	CAA	CGA	TTC	TCT	CTT	CTT	AAA	CAA	CCT	ATA	TAC	TCC	ACA

Splice? (between CAA and CCT of the Maize non-spliced start)

4.2.3.4.1 Introns near the 5' terminus. The AUG codon proposed to initiate the open reading frame encoding the maize COB gene would specify a polypeptide with an amino terminal extension of five or six amino acids when compared to the aligned sequences of other known apocytochromes b (Fig. 4.19). The sequence of the first nineteen residues is different from the aligned sequences, and a highly conserved arginine (or leucine)-lysine pair at positions 10 and 11 is missing. The DNA sequence in this region is also dissimilar from all other COB sequences, with only 22 and 17% homology to the human and yeast sequences respectively, between maize amino acid co-ordinates 6 (for the human comparison) or 7 (yeast comparison) and 18. The equivalent homology between the human and yeast sequences is 44%.

If the 5' end of the maize COB gene is separated from the remainder of the gene by an intervening DNA sequence, then this intron must finish before residue 20, and probably before residue 14, since amino acid homology is good after this point. A sequence capable of encoding the amino terminal portion of the polypeptide does occur, starting with an AUG codon at nucleotide 770 in Fig. 4.17. The feature identifying this as a potential 'start' point is the phenylalanine-arginine-cysteine tripeptide encoded after this codon which is highly conserved in the yeasts and (to a lesser extent) in the other fungi and mammals.

Table 4.3 compares the nucleotide sequences near the two potential initiation codons in maize with the 5' ends of other COB genes. The putative upstream exon, if joined to the rest of the COB gene so as to maximise amino acid homology with the other predicted apocytochrome b sequences, would encode a polypeptide with an initiation codon very close to that in these other proteins. The DNA homology of the first 33 nucleotides of this sequence to the 5' ends of

both the human (45%) and yeast (55%) genes is considerably higher than the corresponding homology if the in-frame AUG codon represents the start of the gene.

The 5' end of the maize COB gene could therefore reside on a separate exon from the rest of the gene. The arguments in its favour are:

1) Non-homologous positioning of the initiation codon when compared to the other aligned COB sequences.

2) A very low DNA sequence homology between the first 57 nucleotides of the gene and sequences from the 5' termini of other COB genes.

3) The presence of a 36 nucleotide sequence upstream from the putative 'in-frame' initiation codon, which could encode a short peptide with an initiator methionine almost coincident with that of other predicted apocytochrome b sequences, and with relatively good sequence homology to the first residues in these other polypeptides. The most 'likely' splice point between the two exons would join the sixth residue of the first 'exon' to the fourteenth residue after the 'in-frame' AUG codon preceding the second 'exon', with a 806 BP intron separating them.

4) The complex pattern of transcription of the maize COB gene revealed by 'Northern' blot analysis (see section 4.3).

5) A short sequence ($^{5'}$CAACCU) at the 3' end of the proposed intron which is identical to a sequence in the corresponding position at the end of the intron in maize COII.

The major arguments against the presence of an intron are:

1) The length of the proposed first exon (18 nucleotides), although the second exon of the 'long' yeast COB gene is only 14 nucleotides long (Lazowska et al, 1980).

2) The presence of an in-frame AUG codon in the proposed intron which,

although not exactly coincident with the initiation codon in other sequenced COB genes, is not unreasonably displaced and is not followed by in-frame termination codons. This AUG codon is preceded closely by termination codons in all three reading frames.

3) The absence of a good ribosome binding/translation initiation point preceding the AUG codon in the proposed first exon (see section 4.4.2).

4) The alignment of a leucine-lysine pair in the maize polypeptide with the 'conserved' leucine/arginine-lysine pair at residues 11 and 12, if the maize sequence is displaced one position to the 'left'.

The presence of an intron near the 5' end of the maize COB gene is unproven but on teleological grounds seems unlikely. Further investigations, for example, RNA:DNA hybridisation and electron microscopy, would be required in order to establish its existence with more certainty.

4.2.3.4.2 Introns near the 3' terminus. The poor homology of the 3' end of the maize COB gene to other COB genes has been discussed in section 4.2.3.1. It was suggested that this region may not encode the carboxy terminus of apocytochrome b, the true coding sequence being located at some other point in the genome and separated from the remainder of the gene by an intron of undetermined length. Extensive sequence data 3' to the COB gene has not been obtained, so the proposed 3' exon could not be located. The existence of an intron in this region is considered unlikely for the following reasons:

1) An in-frame termination codon (UAG) is located in a position coincident with those of aligned COB gene sequences, and no other in-frame termination codon precedes it. The use of UAG is slightly

unusual - all other COB genes terminate with UAA and AGA. Similarly translation of COII from maize, and of COII from Oenothera (Hiesel and Brennicke, 1983) is terminated by a UAA codon.

2) The hydropathy profile (Fig. 4.2.1) of the carboxy terminal region of maize apocytochrome b closely resembles that of all other apocytochromes b which have been studied (section 4.2.3.3.1). Despite a low overall amino acid sequence conservation, it therefore seems probable that this portion of the protein could form a membrane traversing α-helix in exactly the manner proposed for the other apocytochrome b (Saraste, 1983; Widger et al, 1983).

It is therefore unlikely that an intron is found in the 3' region of the COB gene. However, in the absence of 'downstream' sequence data, it has not been possible to identify a coding sequence which could form the exon specifying the carboxy terminal of apocytochrome b. Until such data has been obtained, and confirmed by DNA:RNA hybridisation experiments, the matter will remain unresolved.

4.2.3.5 Potential coding regions in COB flanking sequences.

Several small open reading frames (ORFs) were found in an examination of the entire sequenced region (2737 BP) around the COB gene. Discounting possible ORFs internal to COB and on the same DNA strand, and assuming TGA is a termination codon, the longest ORF is one of 62 amino acids. This ORF is contained in the COB coding region, but is on the opposite DNA strand (nucleotides 2375 to 2190 in Fig. 4.17). It would be unprecedented for one mitochondrial gene to lie entirely within the coding sequence of another, although cases of genes with some overlap (e.g. URF A6L and ATPase 6 in human mt DNA; Anderson et al, 1981) have been reported. In many viruses, for example, SV40 (Fiers et al, 1978) and ØX174 (Sanger et al, 1977) two

genes are encoded by the same sequence of DNA, translated in overlapping reading frames. It is unlikely that this ORF represents a genuine gene for the following reasons:

1) The reading frame is closed by several TAA codons in this region of both the human and yeast COB genes. It seems unlikely that the maize mitochondrial sequence would specify a novel polypeptide in this area.

2) The occurrence of A, G, C and T in the third triplet position is 26, 23, 26 and 25% respectively. This does not differ significantly from a random distribution and certainly does not reflect the pattern found in the other sequenced maize mitochondrial genes known to encode polypeptides.

3) The predicted amino acid sequence shares no detectable homology with that of any other sequenced mitochondrial protein.

Only two other ORFs could encode polypeptides of greater than 50 residues. One of these is on the DNA strand complementary to that encoding COB and does not overlap it. This ORF lies between nucleotides 873 and 721 (Fig. 4.17), but a consideration of third triplet position nucleotides again reveals an essentially random distribution (A= 22%, G = 24%, C = 26%, T = 28%) indicating that it probably does not encode a polypeptide.

The third ORF is found on the same DNA strand as COB, between nucleotides 292 and 465. This ORF has an asymmetric third triplet position usage (A = 41%, G = 26%, C = 19%, T = 14%) which differs from that in the maize COB and COII genes (section 4.2.3.2) but is seemingly non-random. Fig. 4.22 is a hydropathy plot for this polypeptide, prepared exactly as described for apocytochrome b. The mean hydropathy of the eleven amino acid peptides used in this construction is -2.55 indicating a generally hydrophilic protein. These data do not

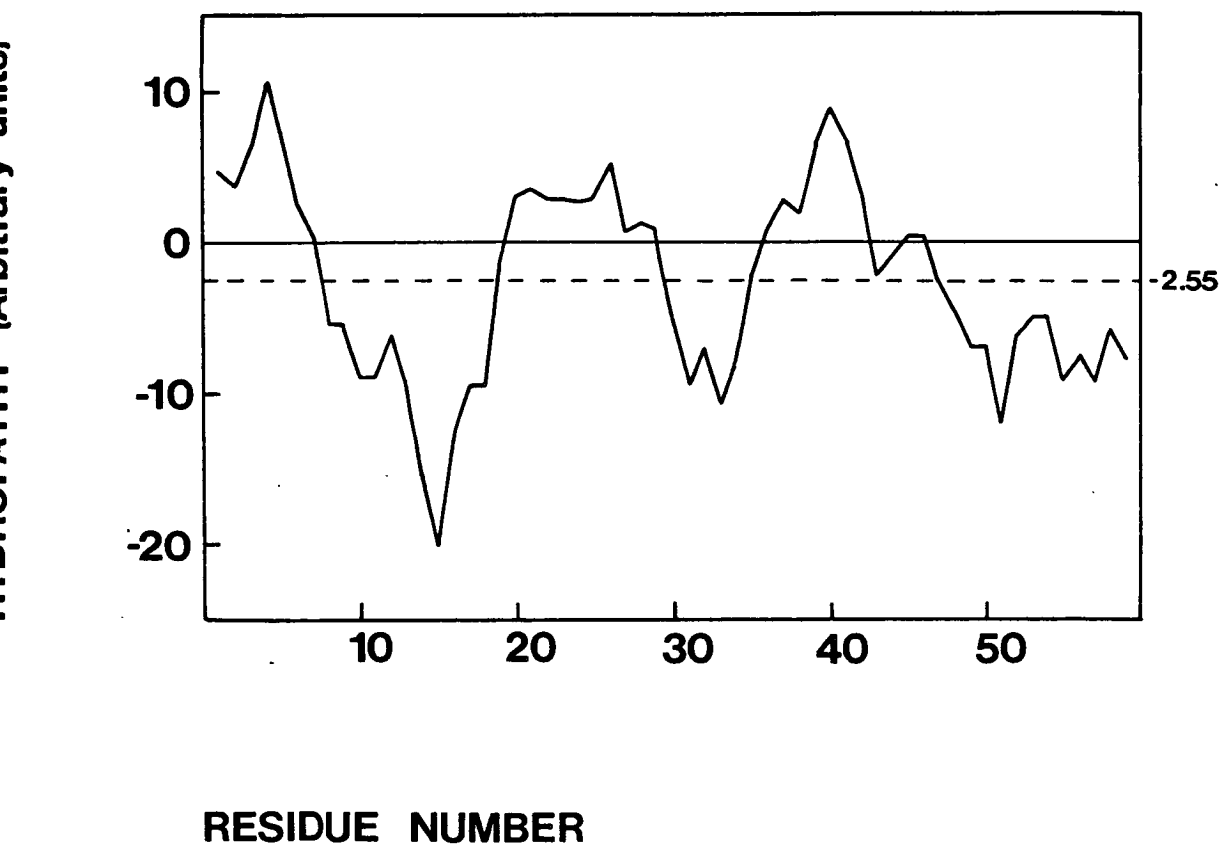


Fig 4.22 Hydrophobicity profile of the short polypeptide predicted from an ORF sequence 5' to COB.

Values, averaged over 11 residues, were calculated as described by Kyte and Doolittle (1982).

prove that this ORF is an actively expressed gene however. Amino acid sequence analysis of all proteins near this size (M_r ca. 6000) synthesised *in vitro* by isolated maize mitochondria would be required to established this. The presence of signals for transcription initiation and ribosome binding must also be demonstrated (section 4.4.4). The predicted amino acid sequence of this protein does not resemble that of any published mitochondrial protein or of any predicted ORF product.

Several other short ORFs are found 5' to the maize *COB* gene, ranging in size from 60 to 102 BP. Whether these represent genuine genes or are simply chance open reading frames is not easy to prove. These ORFs are much shorter than those found upstream from *COB* in *Aspergillus* (Brown *et al*, 1983). A computer search (using a program developed by R. Staden) of the regions flanking the *COB* gene also revealed no sequences likely to encode tRNAs (L. Grivell, pers. commun.) nor any potential hairpin loops of length greater than 12 BP.

4.3 TRANSCRIPTION OF THE *COB* REGION IN MAIZE

Transcript analysis of the *COB* gene was undertaken in collaboration with V. Jones (Department of Botany, University of Edinburgh), with the following aims:

1) To demonstrate the presence of *COB*-specific RNAs in maize mitochondria and therefore to show that the gene is expressed *in vivo*.

2) To investigate patterns of transcription from the *COB* gene, and to formulate a model for the expression of this gene.

4.3.1 Construction and use of an internal COB clone (pZmEH680) for transcript analysis

Preliminary experiments (not shown) revealed that numerous mt RNA species hybridised to the semi-internal *COB* clone pZmH1790. In order

to carry out more specific analyses, the EcoR1 clone pbE4 was used as a substrate to generate an internal probe for COB.

When double-digested with EcoR1 and HindIII, pbE4 yields five fragments (Fig. 4.23). The smallest of these (680 BP) is an EcoR1-HindIII fragment which contains the mt DNA sequence from the HindIII site at the end of the 1787 BP fragment cloned in pZmH1790, to the EcoR1 site near the 3' terminus of the COB gene (i.e. nucleotides 2016 to 2695 in Fig. 4.17). This fragment has been shown by sequence analysis to be completely internal to the COB gene. In order to clone this fragment, pbE4 (200 ng) was digested with EcoR1 plus HindIII, and after phenol extraction and ethanol precipitation, the DNA was re-ligated (see Chapter 2). This ligation mixture was used to transform E.coli HB101 and recombinant colonies having Ap^r, Tc^s, Cm^s phenotype were selected and plasmid DNA was prepared. Double digestion with EcoR1 and HindIII revealed that several of these clones contained the 680 BP insert and one of these was designated pZmEH680 (Fig. 4.23).

Mt RNA was prepared from etiolated maize seedlings (hybrid line 3541) as described in Chapter 2 and a 50 µg sample was loaded into a single well, 5 cm wide, in a 1.3% w/v agarose, 17% w/v formaldehyde gel. Following electrophoresis, RNA was transferred from the gel to nitrocellulose using the method of Thomas (1980). Fig. 4.28a shows part of this gel, including the molecular weight markers, stained with acridine orange. The section of the gel which was blotted was not stained prior to transfer. The hybridisation pattern of nick-translated pZmEH680 to a strip cut from this blot is shown in Fig. 4.28b.

These results identify a major transcript of 2.2 Kbases, an additional transcript of 4.3 K bases and numerous smaller RNA species which hybridise to the COB-specific probe. These findings are reproducible when different RNA denaturation and electrophoresis

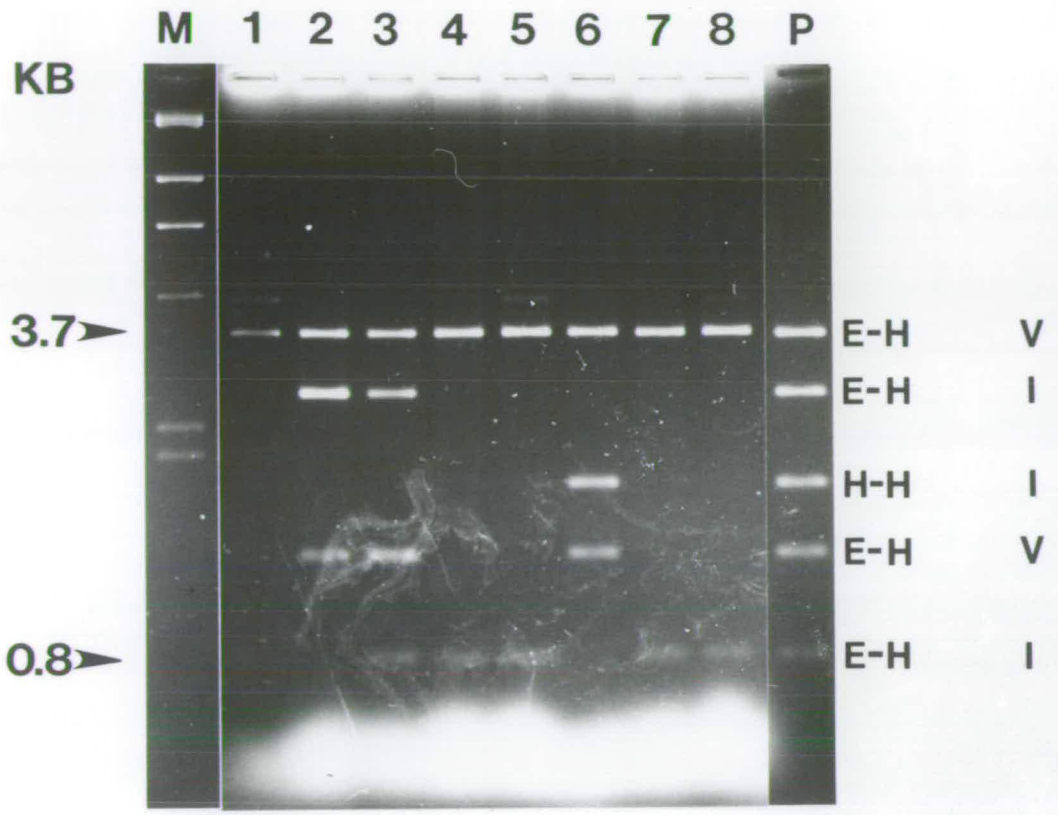


Fig 4.23 Identification of pZmEH680.

Plasmid DNA was prepared from Ap^r, Tc^s, Cm^s subclones of EcoRI - HindIII digested pbE4 and cut with EcoRI and HindIII. DNA fragments were separated by electrophoresis on a 1% agarose gel (lanes 1 - 8). One clone (track 4) contained a single DNA insert which co-migrated with the small EcoRI - HindIII fragment of pbE4 (lane P) and was designated pZmEH680. λDNA digested with HindIII formed molecular weight markers (Lane M).

Abbreviations: EcoRI (E); HindIII (H); vector-derived (V); insert-derived (I).

procedures (e.g. 2% w/v agarose-urea gels) are used, and also if RNA is transferred to DBM paper rather than nitrocellulose (not shown). Under the hybridisation conditions used, nick-translated pBR328 did not anneal to mt RNA. It therefore seems likely that these transcripts are genuine products of the COB gene.

4.3.2 The use of single stranded (M13) internal COB clones for transcript analysis

Transcript analysis using single-stranded M13 clones was carried out for two reasons:

1) To probe mt RNA with a non-pBR328 derived clone, thereby reducing further the possibility that spurious hybridisation was being identified due to vector homology.

2) To detect transcripts (if any) from both the COB encoding and non-encoding DNA strands, using strand-specific probes.

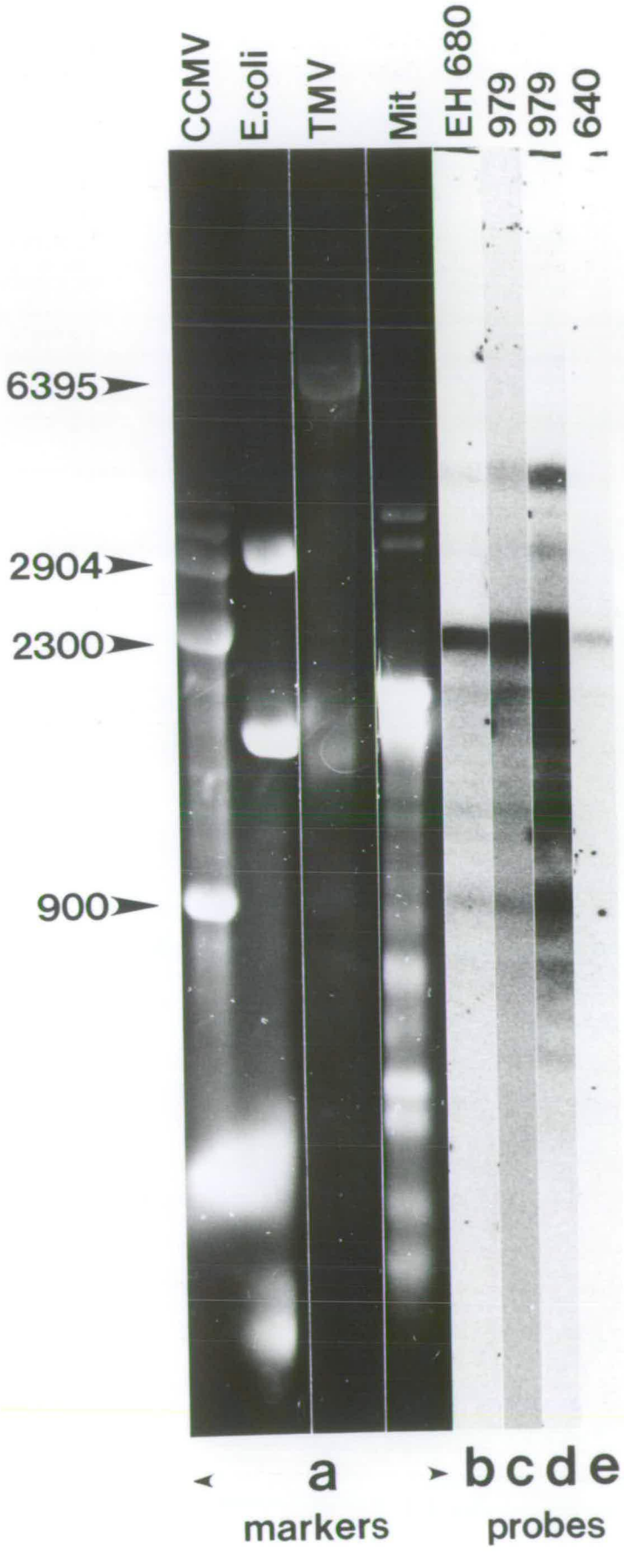
The EcoR1-HindIII clones designated numbers 640 and 979, which contain the same mt DNA fragment as pZmEH680, were used for the production of single stranded, radioactive DNA probes complementary to the mt RNA insert. The probes were ^{32}P labelled to high specific activity (5-10 x 10^6 dpm/µg) by second strand synthesis using the Klenow fragment of E.coli DNA polymerase I (section 2.2.6.3). Clone 640 contains mt DNA inserted in the M13 vector mp8 and the sequence of the newly synthesised DNA strand from this template is identical to that of the COB mRNA. Hence, the probe should only hybridise to transcripts from the non-coding DNA strand. Clone 979 contains the same DNA fragment in M13 mp9, the insert is therefore in the opposite orientation and the probe should hybridise to COB transcripts.

Parallel strips were cut from the RNA blot shown in Fig. 4.24a and hybridised with these probes. The transcript pattern identified by clone 979 (Fig. 4.24) is identical to that obtained with nick-

Fig 4.24 Transcripts of the maize COB gene.

4.24a Acridine orange-stained electrophoretogram of cowpea chlorotic mottle virus (CCMV), E.coli, tobacco mosaic virus (TMV) and maize mitochondrial (mit) RNAs. See chapter 2 for size estimations.

4.24b,c,d,e
mt RNA transferred to nitrocellulose and hybridised with ^{32}P labelled pZmEH680(b), M13 clone 979 (c,d) and M13 clone 640(e). After hybridisation, filters were autoradiographed for 16 hr (b,c) or 72 hr (d,e).


CCMV
E.coli
TMV
Mit
EH 680
979
979
640
6395
2904
2300
900
a
b c d e
markers
probes

translated pZmEH680, showing that all the transcripts which hybridise are complementary to the COB coding strand. This also indicates that neither M13 nor pBR328 vector homology makes any significant contribution to the hybridisation pattern. A longer exposure (Fig. 4.24d) revealed the presence of an extremely large (>9 Kbases) transcript which is homologous to the COB probe.

Clone 640, with the same specific and total activity as 979, hybridised only weakly to the mt RNA blots (Fig. 4.24e). Furthermore, the hybridising RNA species co-migrated with those detected using the M13 640 and pZmEH680 clones. It therefore seems unlikely that these are genuine transcripts from the non-coding DNA strand. Rather, this apparent homology probably arises because the non-labelled single-stranded template DNA hybridises to the RNA and M13-homologous sequences in the radioactive probe, generated by second strand synthesis beyond the M13 EcoR1 site, anneal to their complements in these hybridised molecules. The level of hybridisation in this case is low, presumably because a second-order reaction, rather than a first order reaction as for directly complementary sequences, is required to bind the radioactive probe to the blotted RNA. If transcripts did arise directly from this strand, they would probably be detected as more strongly hybridising species superimposed on the fainter background. The lack of such homology indicates that, in the tissue used as the source of RNA in these experiments, the ORF detected in the non-COB encoding DNA strand (section 4.2.3.5) is almost certainly not transcribed.

4.3.3. Discussion

Transcript analysis of the maize COB gene has shown:

1) Mitochondrial RNA hybridises to COB-specific DNA probes, indicating that this gene is being activity transcribed in the tissue studied.

2) The hybridisation probes identify a complex pattern of transcripts homologous to the COB gene.

3) The major transcript, presumably the mature mRNA, is much larger (ca. 2.2 Kbases) than the gene which encodes it (1164 bases).

Two models could be proposed to explain the presence of multiple transcripts:

1) Transcription initiation and termination could occur at several points throughout the COB region, giving rise to a set of overlapping RNAs. The relative abundance of these transcripts would depend both on the potencies of the competing promoter/terminator sequences and on the stabilities of the various RNAs in the mitochondrion.

2) A single large RNA species (perhaps the >9KB molecule) could be transcribed between a single promoter and terminator then processed via intermediates of lower molecular weight to a mature message. The abundance of the different COB-homologous transcripts would then depend on the rates at which the various processing steps were completed.

The latter model is perhaps preferable on teleological grounds although neither model is supported by firm experimental evidence. It should be possible to distinguish the two possibilities by S1-nuclease 'protection' experiments (Berk and Sharpe, 1977). In this way, the DNA sequences at the points where transcripts are initiated or terminated can be determined. The first model would predict that all the initiation points would have sequence homology if acting as promoters for a single RNA polymerase. Similar arguments apply to the terminator sequences. The second model, in contrast, would not necessarily predict any sequence homology at the transcript termini since all could be generated from recognition by different processing enzymes. Promoter sequences for plant mitochondrial RNA transcription

have not yet been identified. Sequences with similarities to the E.coli Pribnow box (Pribnow, 1975) and '-35' sequence (Rosenberg and Court, 1979) are present approximately 800 BP 5' from the initiation codon of the COI gene (P. Isaac, pers. commun.). The significance of this finding remains to be ascertained since similar sequences have not been located upstream from COB. However, transcription of COI appears altogether less complex than that of COB, so putative promoters of the latter gene may well be differently organised and located.

The additional coding capacity of the large 'mature mRNA' could be occupied by sequences encoding other genes, i.e. the COB mRNA could be part of a polycistronic message. Such messages are found in yeast mitochondria (e.g. Macreadie et al, 1983) and also in those of mammals (Ojala et al, 1981). However, this seems unlikely to be the case for the maize COB gene since long ORFs are not found immediately 5' to the COB gene. Whilst the possibility of 'downstream' ORFs cannot be excluded because of the lack of sequence data, it does not seem that stable polycistronic mRNAs are common in maize mitochondria since the 3' and 5' regions of the maize COI (P. Isaac, pers. commun.) and COII (Fox and Leaver, 1981) genes do not contain long ORFs yet the 'mature mRNAs' in both cases are significantly longer than the genes which encode them.* The closest approximation to a polycistronic 'message' in higher plant mitochondria is the 18S-5S rRNA locus in wheat which is believed to be expressed as a single transcriptional unit (Gray and Spencer, 1983) and subsequently processed. It is more probable however that the excess length of the mature mRNA is occupied by nontranslated RNA sequences of structural rather than coding importance. This would be similar to the situation in yeast where the COB mRNA contains a 5' untranslated leader sequence of 950 nucleotides (Bonitz et al, 1982; Christianson et al, 1983).

*(But see also Chapter 5).

The other higher molecular weight RNAs could, however, be polycistronic precursors which are processed to yield the mature mRNAs for several genes. This would be more consistent with the single transcript/processing model than with the multiple initiation and termination model. This is because in the latter model, the higher molecular weight, low-abundance RNA species are proposed to be products of rare initiation/termination events and are therefore unlikely to contain the coding sequences of important structural genes. The 'polycistronic precursor' model would be proven if a long open reading frame, close to COB and on the same DNA strand were found which hybridised to the same precursor RNAs but to a different mature mRNA.

The complex pattern of transcription of the COB gene resembles that of mitochondrial genes which are split by introns, for example the S.cerevisiae COB (Schweyen et al, 1982) and COI (Hensgens et al, 1983) loci and also the Z.mays COII locus (Fox and Leaver, 1981). The maize COI gene, which is not divided by an intron (P. Isaac, pers. commun.) displays a much simpler transcript pattern, with only a single RNA species hybridising (V. Jones, pers. commun.). An obvious conclusion is that the maize COB gene could contain an intron (see section 4.2.3.4) which was not detected by DNA sequence analysis. This is not the only explanation consistent with the data, however. The COB gene could be co-transcribed with a nearby ORF to yield a polycistronic message which was subsequently processed. The mRNA from COI gene, on the other hand, may never form part of such a precursor transcript and would not require processing. Thus, similar mechanisms of transcription could operate in both instances, the differences in the RNA hybridisation patterns being due solely to the amount of processing required to produce the mature mRNA. This would be similar to the situation in yeast where probes from the COII gene,

which is not split, hybridise to at least ten RNA species (Tzagoloff, 1982; see also Grivell (1983) for a review of RNA splicing in yeast). If such mRNA processing does occur, this would provide an additional level at which control of the expression of the maize COB gene could occur. Thus, when studying RNA levels during mitochondrial biogenesis, it will be important to make qualitative as well as quantitative measurements of the RNA species present.

A completely different model could be proposed to account for the transcript pattern of the COB gene. If this locus lay in a region of genome instability, then a heterogeneous series of mitochondrial genomes could exist in the tissue studied. Different genome populations might possess different DNA sequences flanking the gene. Thus, even if transcription of the COB gene was initiated and terminated at only a single point in each population, the RNA isolated from this tissue would contain a mixed set of COB transcripts. Whilst this could be the case for maize COII (D. Lonsdale, pers. commun.), however, the mt DNA restriction map around COB is well established so this gene is unlikely to reside in a region of mt DNA instability.

RNA hybridisation experiments have therefore demonstrated that the maize COB gene is transcribed in a complex manner. This type of transcription is not a universal feature of maize mt genes, however. A single model, involving differential RNA processing has been proposed which could account for all the patterns of transcription observed. More experimental data, for example DNA sequence and RNA transcript analyses of regions 5' and 3' to the COB gene, coupled with nuclease S1 protection experiments will be required in order to verify this model or to propose a more accurate one.

4.4 INITIATION OF RNA TRANSLATION IN HIGHER PLANT MITOCHONDRIA

4.4.1 Introduction

The mature mRNAs transcribed from all plant mitochondrial protein-coding genes so far studied are considerably longer than the genes which encode them. Assuming that at least some of this extra sequence is 5' to the initiation codon and that statistically this sequence contains AUG codons which are not initiation codons, then some mechanism must exist to confer specificity on the AUG codon which is used to initiate translation of the gene.

In human mitochondria, the mature mRNAs do not appear to contain 5' and 3' non-translated regions and translation simply starts at the beginning of the mRNA (Montoya et al, 1981). Similarly, translation of eukaryotic cytoplasmic mRNAs, which often have short 5' leader sequences, starts at the first AUG codon encountered by the ribosome (Kozak, 1978), and a 'scanning' mechanism involving the 40S ribosomal subunit has been proposed to account for this (Kozak, 1978). At present, the signals involved in translation/initiation of fungal mitochondrial mRNAs are not known. Li et al (1982) have proposed that a sequence $^{5'}$AAAUUCUAUA occurring near the 3' end of the S.cerevisiae 15S rRNA could basepair with a 4 to 10 nucleotide complementary sequence found between 8 and 116 nucleotides upstream from the initiation codons of all known S.cerevisiae genes. The importance of this region remains to be assessed since although functionally equivalent to the bacterial Shine-Dalgarno sequence (see below) its length and positioning vary considerably.

In the eubacteria (Rosenberg and Court, 1979) and also probably in chloroplasts (Whitfield and Bottomley, 1983), RNA transcription normally starts a considerable distance (50-500 nucleotides) upstream from the AUG initiation codon. Specificity of translation initiation is conferred on this codon by a sequence, between three and nine

nucleotides long and three to eleven nucleotides 'upstream' from it, which is complementary to a sequence near the 3' end of the 16S rRNA (Shine and Dalgarno, 1974; Steitz and Jakes, 1975). This region is commonly known as the Shine-Dalgarno (or SD) sequence.

4.4.2 Ribosome binding sites in plant mitochondrial mRNAs

The complete sequence of the gene encoding the wheat mitochondrial 18S rRNA has been determined recently (Spencer et al, 1983), as has the sequence of the portion of the gene encoding the 3' end of this rRNA in maize mt DNA (Chao et al, 1983). Comparisons with the E.coli 16S rRNA sequence (Brosius et al, 1978) have revealed remarkable primary and secondary sequence homologies (Spencer et al, 1983). These higher plant 18S rRNAs are much more similar to the E.coli 16S rRNA than are the functionally equivalent 15S fungal (Köchel and Küntzel, 1981) or the 12S mammalian mt rRNAs (Zwieb et al, 1981). It therefore seems possible that translation of higher plant mitochondrial mRNAs could have some prokaryotic features.

The nucleotide sequences at the 3' termini of the E.coli 16S and the wheat and maize 18S rRNAs are compared in Fig. 4.25. The SD region of the E.coli sequence is indicated. It appears that whilst the sequences are in general very well conserved, part of the SD sequence ($^{5'}$CACCUCCU) is replaced by a novel sequence ($^{5'}$UGAAUCCU) in the equivalent region of the higher plant rRNAs. If in higher plants, mitochondrial translation proceeds in a prokaryotic manner, then it seems possible that this sequence could be involved in SD-type interactions between the ribosome and the mRNA. This would offer a plausible mechanism for the seemingly specific initiation of translation at a single AUG codon. Regions 5' to the AUG codons initiating sequenced plant mt genes were examined for sequences capable of base-pairing with the proposed SD sequence. In particular, sequences

Fig 4.25 Comparison of nucleotide sequences at the 3' termini of E.coli 16S and maize and wheat 18S rRNAs.

E.coli: AGUCGUAACA AGGUAACCGU AGGGGAACCU GCGGUUGGAU CACCUCCUUA 3'
****

Maize: AGUCGUAACA AGGUAGCCGU AGGGGAACCU GUGGCUGGAU UGAAUCCUUC 3'

Wheat: AGUCGUAACA AGGUAGCCGU AGGGGAACCU GUGGCUGGAU UGAAUCCUUC 3'

Asterisks (*) indicate the E.coli Shine-Dalgarno sequence.

Underlined residues indicate differences in the plant sequences from the E.coli sequence. Note the 100% identify of the wheat and maize sequences.

See text for references.

complementary to $^{5'}$UGAA were sought because:

1) This sequence is specific to the SD region of higher plant 18S rRNAs, and

2) Spencer et al (1983) have shown by RNA sequence analysis that part of the 3' end of the wheat 18S rRNA is removed by post-transcriptional processing. The sequence at the 3' terminus is then UGAAUCC$^{3'}$ or UGAAUC$^{3'}$.

The results of this analysis are shown in Fig. 4.26. Regions capable of basepairing with the rRNA sequence are shown. In all the cases examined a sequence occurs approximately 15 to 20 nucleotides 5' from the AUG codon which has either a 3/4 or 4/4 complementarity with the proposed mt rRNA SD sequence $^{5'}$UGAA. In most of these cases, additional homology to the remainder of the $^{5'}$UGAAUCCU sequence occurs in this region. The AUG codon proposed to initiate the URF-1 gene in maize mt DNA may also be preceded by a similar sequence (T. Brown, pers. commun.).

The point at which translation of the maize COII gene is initiated is not known with certainty, so the analyses of this gene are rather preliminary. The initiation point of COII in Oenothera is more definite since a single AUG codon is found at the predicted 5' end of the gene and is preceded by in-frame termination codons (Hiesel and Brennicke, 1983). The region 5' to this AUG codon contains a sequence complementary to the proposed wheat and maize SD sequence. It seems likely that the same SD sequence will also be found in the Oenothera 18S rRNA, in view of the 100% nucleotide sequence conservation at the 3' termini of the wheat and maize rRNAs.

It is possible that these sequences upstream from the initiation codons could act as binding sites for mitochondrial ribosomes. The functional significance, if any, of these ribosome binding sites (RBSs) remains to be evaluated, possibly by ribosome binding protection

Fig. 4.26 Comparison of sequences 5' to the initiation codons of higher plant mitochondrial genes.

Maize 18S rRNA (3' terminus)	3'	CU UCCUAAGU UAGGUCGGUGUCCAAGGGGAUG
Maize COB	5'	GAGG AGUUGUCA CGAUAGAAAAGAGAAAUGA
Maize COI	5'	AUAA GGUUUUCA AAACGAAAAAAAAAAUGA
Maize COII[1]	5'	GCUC CUACUUCU GGUGCUGCCA AUGA
Oenothera COII	5'	AGCG GAGAGUCA AAAAAGAAACCAAAGCAAAUGA
Maize S2 ORF[2]	5'	AAAA GGCAUUCA AACUAAAAUAGAACGUAUGA

The proposed ribosome binding octanucleotide is boxed.
Nucleotides capable of basepairing with 18SrRNA in this region are indicated with asterisks.
The initiation codons are underlined.

Notes:
1. The exact initiation point of the maize COII gene is not known. See Chapter 5 for discussion.
2. This is the second AUG codon in the long ORF which is preceded by a putative ribosome binding site. However, the first AUG codon is not followed by an A residue. This first AUG codon is 186BP upstream from one shown in fig. 4.26.

See text for references.

experiments (see for example, Ravetch et al, 1977) to identify sequences involved in the formation of a translation initiation complex.

4.4.3 Binding of $tRNA_f^{met}$ to plant mitochondrial mRNAs

The work of Guillemaut and Weil (1975) had shown that bean mitochondria contained two $tRNA^{met}$ species which could be formylated by a mitochondrial transformylase. The existence of a $tRNA_f^{met}$ gene in wheat mt DNA has been demonstrated by DNA sequence analysis (Gray and Spencer, 1983). It therefore seems that translation of higher plant mitochondrial mRNAs, in common with those of prokaryotes, chloroplasts, fungal and mammalian mitochondria, is initiated by binding of a charged $tRNA_f^{met}$ to the initiator AUG codon.

Taniguchi and Weissmann (1978) have shown that, at least in the case of the bacteriophage Qβ coat cistron, interaction of $tRNA_f^{met}$ with the initiation AUG codon was important in the establishment of a 70S translation complex. They suggest that this interaction is brought about by base pairing of the CAU anticodon to the AUG initiator. They were able to show furthermore, by in vitro mutagenesis, that the $^{5'}$AUGA quadruplet formed 70S complexes up to five times more efficiently than the quadruplet $^{5'}$AUGG. In addition, they found that while $^{5'}$AUAG would only permit initiation at very low levels, the quadruplet $^{5'}$AUAA was almost as efficient as $^{5'}$AUGG. Hence, the identity of the nucleotide immediately following the initiation codon was important in ensuring efficient translation. The reason is probably that this nucleotide has the potential to base pair with the U residue on the 5' side of the CAU anticodon in the $tRNA_f^{met}$. It has now been shown that A residues frequently follow AUG initiation codons in a variety of bacterial and bacteriophage (e.g. M13; van Wezenbeck et al, 1980) genes, so the importance of this interaction is probably widespread in the prokaryotes.

The sequence analysis of the wheat mt $tRNA_f^{met}$ gene shows that the anticodon loop contains the sequence $^{5'}$UCAU, in common with almost all other $tRNA_f^{met}$ species which have been analysed (Gray and Spencer, 1983). It is therefore highly likely that the maize tRNA will contain the same sequence. Examination of the sequenced plant mt genes shows that in every case, the proposed initiation codon is followed by an A residue (Fig. 4.26). Thus, an interaction of the mitochondrial $tRNA_f^{met}$ with the mRNA, of the type proposed by Taniguchi and Weissmann, could be important in the establishment of a complex for the translation of plant mitochondrial mRNAs. This A residue could also confer an additional level of specificity on the correct AUG codon for translation initiation.

4.4.4 Discussion

Comparison of sequences 5' from the initiation codons of a variety of plant mt genes with the 'Shine-Dalgarno' (SD) region of plant mt 18S rRNAs, has allowed a sequence ($^{5'}$UUCA) to be identified which may form part of the ribosome binding site (RBS) for the mRNAs transcribed from these genes. This sequence is normally 15-20 nucleotides from the AUG codon, which is somewhat more distant than the SD sequence in bacterial mRNAs (normally 3-11 nucleotides from the initiation codon; Shine and Dalgarno, 1974). The difference in spacing need not imply that the sequence is not involved in ribosome binding since the fine details of translation initiation are almost certain to differ between E.coli and higher plant mitochondria. Furthermore, the proposed yeast ribosome binding site can be found between 8 and 116 nucleotides from the AUG codon. The universality of this sequence preceding all the protein coding plant mt genes studied, and its homology to a higher plant-specific sequence in the small (18S) rRNA, argues strongly in favour of its involvement in ribosome binding.

A second prokaryotic feature of translation initiation in higher plant mitochondria is the preferential occurrence of the nucleotide 'A' immediately after the AUG initiation codon. Table 4.4 shows the initiation codon usage in a variety of plant, fungal and animal mitochondrial genes. Whilst there is a very strong bias in favour of AUGA in higher plants, this trend is not so marked in any other species. It is interesting that in Drosophila, the initiation quadruplet AUAA precedes some mitochondrial genes (de Bruijn, 1983). The AUA triplet is not in phase with the reading frame specifying the remainder of the gene, so this is rather different from the AUGA "quadruplet" favoured in higher plants, where the AUG triplet is in frame. This does however indicate the potential existence and importance of four base pair codon:anticodon interactions.

The proposed SD interaction at the ribosome binding site and the preference for AUGA initiation sequences both suggest that translation of plant mitochondrial mRNAs is more prokaryotic in nature than translation in either fungal or animal mitochondria. Whilst this may have evolutionary significance, a more practical implication could be in the construction of cell free systems for the in vitro translation of plant mt mRNAs. If the proposed SD sequence is actually involved in mitoribosome binding, then it is clear that E.coli ribosomes could not select the correct binding site in plant mt mRNAs. Even if these ribosomes could bind at this point, the longer distance from the AUG codon may preclude translation initiation. However, since translation seems otherwise to proceed in a rather bacterial fashion, it is possible that co-operation of plant mitoribosomes with E.coli initiation, elongation and termination factors could occur. Thus, an E.coli S-30 cell free system supplemented with plant mitochondrial ribosomes might be used to bring about in vitro translation of plant mitochondrial mRNAs. Such a system would probably not work for other mRNAs from the

Table 4.4 Initiation codons and following nucleotide in various mitochondrial genes.

ORGANISM	COB	COI	COII
Z mays	ATGA	ATGA	ATGA
S. cerevisiae	ATGG	ATGG	ATGT
K. lactis	ATGT	na	na
B. taurus	ATGA	ATGT	ATGG
H. sapiens	ATGA	ATGT	ATGG
R. rattus	ATGA	na	na
M. musculus	ATGA	ATGT	ATGG

na = not available

mitochondria of other eukaryotes since:

1) the existence of ribosome binding sites in these mRNAs has not been convincingly demonstrated; and

2) translation appears decidedly non-prokaryotic in both fungi and mammalian mitochondria (section 4.4.1).

A problem which has been encountered when describing possible ribosome binding sequences is the occurrence of similar sequences within gene coding regions. The majority of these sequences can be discounted because they do not precede AUG codons. However, a few do precede AUG codons (there are three examples in the maize COB gene) at approximately the predicted distance (for example, one such sequence is 5'UAGCUUUCAACAGCGUAGAACAAAUUAUGA between nucleotides 1750 and 1779 in Fig. 4.17; the SD sequence and AUG codon are underlined). Any model of mRNA translation must also explain why such sequences do not act as internal translation initiation points. At least three possible explanations exist:

1) Internal 'RBSs' could be obscured by secondary structures in the mRNA. At least one such site is found in the stem section of a potential hairpin loop in the COI mRNA (P. Isaac, pers. commun.) and is therefore inaccessible to 18S rRNA basepairing.

2) When the mRNA becomes saturated with ribosomes in the polysomal configuration, further binding of ribosomes to internal sites may be precluded on steric grounds.

3) If the 5' nontranslated regions of the mature mRNAs are relatively short and do not contain putative RBSs translation initiation may simply occur at the first region capable of binding to the 18S rRNA and which precedes an AUG codon. Since neither the lengths of 5' non-translated regions of the mRNA, nor all the factors involved in the

formation of a translation initiation complex, are known, the likelihood of this possibility cannot be assessed.

Similar 'internal' ribosome binding sites occur in prokaryotic mRNAs and do not appear to interfere with their translation. For example, the M13 gene IV ribosome binding site is $^{5'}$AAGGTAATTCAAATGA yet the sequence $^{5'}$CAGGTAATTGAAATGA occurs within this gene [in-frame met codons are underlined] and does not appear to act as a translation initiation point (van Wezenbeck et al, 1980). Similarly, the semi-synthetic human growth hormone gene cloned in a bacterial expression vector was found to be efficiently translated from an RBS with the sequence $^{5'}$AGGAAACAGCUAUGU. This gene also contains a sequence $^{5'}$AGGAAGGACAUGG, which does not act as an RBS, approximately 500 nucleotides from the start of the gene. If these sequences can be masked in the bacteria, there is no obvious reason why they should not be similarly blocked in plant mitochondria.

Two sequences (the RBS and AUGA qualruplet) appear important in the initiation of translation of plant mRNAs and in the conference of ribosome binding specificity on the correct AUG codon. The presence of these sequences may prove useful in the identification of ORFs in plant mt DNA which have protein coding potential. It is perhaps significant in this respect that the ORF lying upstream from COB, between nucleotides 292 and 465 (section 4.2.3.5) does not possess either of these features.

The recently published DNA sequence of the maize S2 episomal DNA (Levings and Sederoff, 1983) was analysed for the presence of potential protein-coding genes using the RBS and 'AUGA' criteria. The sequence analysis revealed the presence of two long, non-overlapping ORFs/ of 3294 and 1017 BP respectively on opposite DNA strands. The

longer ORF contains six AUG codons preceded by putative RBSs (identified by the presence of a sequence including 5'UCA approximately 10-20 nucleotides from the AUG codon on the 5' side). Three of these AUG codons are followed by A residues. The first of these could encode a polypeptide of 958 amino acids (M_r ca. 100,000). The occurrence of the four nucleotides in the third position of the coding triplets is: A = 28.7%, G = 20.5%, C = 14.9% and T = 35.8%. The preference for T in the third position has been recognised as a characteristic of all plant mitochondrial protein coding genes, (section 4.2.3.2) and confirms that this ORF could have a protein-coding function. The 1017 BP ORF contains only one AUG codon preceded by a putative ribosome binding site, and this codon is followed by a C residue rather than A. This reading frame could encode a polypeptide of 161 amino acids (M_r ca. 17,000) and shows a nucleotide distribution in the third triplet position (A = 26.7%, G = 15.9%, C = 20.5%, T = 37.1%) similar to that of protein-coding genes.

It is not known whether these putative ribosome binding sites are genuine points of at which translation of S2-encoded genes is initiated. It has been shown that mitochondria containing the S2 (and S1) episomes synthesise a spectrum of high molecular weight polypeptides (Forde and Leaver, 1980). It would be interesting to speculate that at least some of these could arise from overlapping translation products from the various putative initiation sites in the large S2 ORF.

4.5 LOCATION OF THE COB GENE ON THE MAIZE MITOCHONDRIAL DNA RESTRICTION MAP

4.5.1 Origin of the map

A cosmid library has been prepared by D. Lonsdale and his colleagues (at the Plant Breeding Institute, Cambridge) from mt DNA of the male fertile (Normal cytoplasm) maize line WF9. Restriction enzyme mapping of these clones has allowed this group to construct a preliminary map of the mitochondrial genome.

The cosmid library used in these mapping experiments was a gift of D. Lonsdale and contained 384 cosmids constructed by insertion of sized (ca. 35 KB), partially digested Sau3A fragments of mt DNA, into the single BamH1 site of the vector pHC79 (Hohn and Collins, 1980). Assuming a genome size of 700 KB (D. Lonsdale, pers. commun.), the probability of any point on the genome being excluded from this library is, on statistical grounds, $[(700-35)/700]^{384}$, which is $<10^{-8}$. The library was supplied as a grid of cosmid-containing E.coli colonies on Whatman 541 paper (prepared as described in Chapter 3), and was hybridised with COB specific DNA probes in order to identify the clones containing this gene. DNA was isolated from these cosmid clones and subjected to fine detail restriction mapping in order to establish the precise position of the gene. No differences between the mitochondrial restriction maps of the male-fertile lines B37N and WF9N have been detected (D. Lonsdale, pers. commun.)

4.5.2 Identification of cosmid clones containing the COB gene

The M13 mp8 clone number 640, which contains an internal portion of the maize COB gene (section 4.2.2.4), was ^{32}P labelled by second strand synthesis and used to probe the cosmid library. The filter containing the library was pre-hybridised for 60 minutes in 3 x SSC, 10 x Denhardt's solution, 0.1% w/v SDS and 100 µg/ml sonicated denatured herring sperm DNA at 65°C. Approximately 10^6 dpm of denatured probe DNA was added and allowed to hybridise for a further 16 h. The filter was then washed twice in 3 x SSC at 65°C for 30 minutes each and then in 0.3 x SSC for 30 minutes at 65°C. Following autoradiography of the filter, four strongly hybridising clones designated 9-1D10, 9-2F1, 9-2F6 and 9-3F5 were detected along with several fainter ones, such as 9-1D2 (Fig. 4.27). These cosmids were shown, after consultation with D. Lonsdale, to contain mt DNA inserts which overlapped on the mt DNA restriction map.

Fig 4.27 Identification of cosmid clones containing the maize COB gene.

The cosmid library, immobilised on Whatman 541 paper, was probed with ^{32}P-labelled M13 clone 640. After fluorography, strongly hybridising clones 9-1D10, 9-2F1, 9-2F6 and 9-3F5; and weakly hybridising clones such as 9-1D2 (arrowed) were identified.

9-1 D2
9-1 D10
9-2F1
9-2F6
9-3 F5

Cosmid DNA was prepared from 100 ml cultures of the clones 9-1D10 and 9-1D2 as described in Chapter 2 then digested with SmaI. After electrophoresis on a 0.7% agarose gel, DNA fragments containing the COB gene were detected by hybridisation of the labelled probe (clone number 640) to a Southern blot of this gel (Fig. 4.28). The autoradiograph shows that the COB gene probably resides on the 11.9 KB SmaI fragment; the larger fragment of 9-1D2 which hybridises to the probe is a compound molecule containing mt DNA and vector DNA.

The EcoR1 clone pbE4 contains most of the maize COB gene (Fig. 4.12). Linearisation of this clone with SmaI indicates a single recognition site ($^{5'}$CCCGGG) for this enzyme is found in the plasmid. Sequence analysis of pBR328 (Prentki et al, 1981) has shown the vector DNA does not contain this sequence. Furthermore, sequence analysis of a 2.7 KB segment of the insert including the COB gene shows no SmaI sites are found in this region (Fig. 4.29). By double digestion of pbE4 (Fig. 29), the SmaI site was shown to divide the 5.1 KB EcoR1 fragment into 3.5 and 1.6 KB segments. The sequence data confirmed that the 3.5 KB SmaI-EcoR1 fragment must contain the maize COB gene. The gene could therefore be mapped within the 11.9 KB SmaI fragment. The orientation of this fragment in the mt genome was deduced from the following cosmid hybridisation data:

1) The cosmid 8-3H4, which contains the right-hand 4 KB of the 11.9 KB SmaI fragment (D. Lonsdale, pers. commun.; see Fig. 4.30) does not hybridise to the M13 clone 640. The mt DNA insert in this clone is derived from the small EcoR1-HindIII sub-fragment of pbE4. Fig. 4.29 shows that this sequence is between 2.8 and 3.5 KB from the mapped SmaI site. This clone should therefore identify any cosmid containing DNA extending more than 2.8 KB from the SmaI site nearest the COB gene.

2) The cosmid 9-1D2 which contains the left hand 4-5 KB of the 11.9 KB SmaI fragment (Fig. 4.30), but not the right hand 7-8 KB

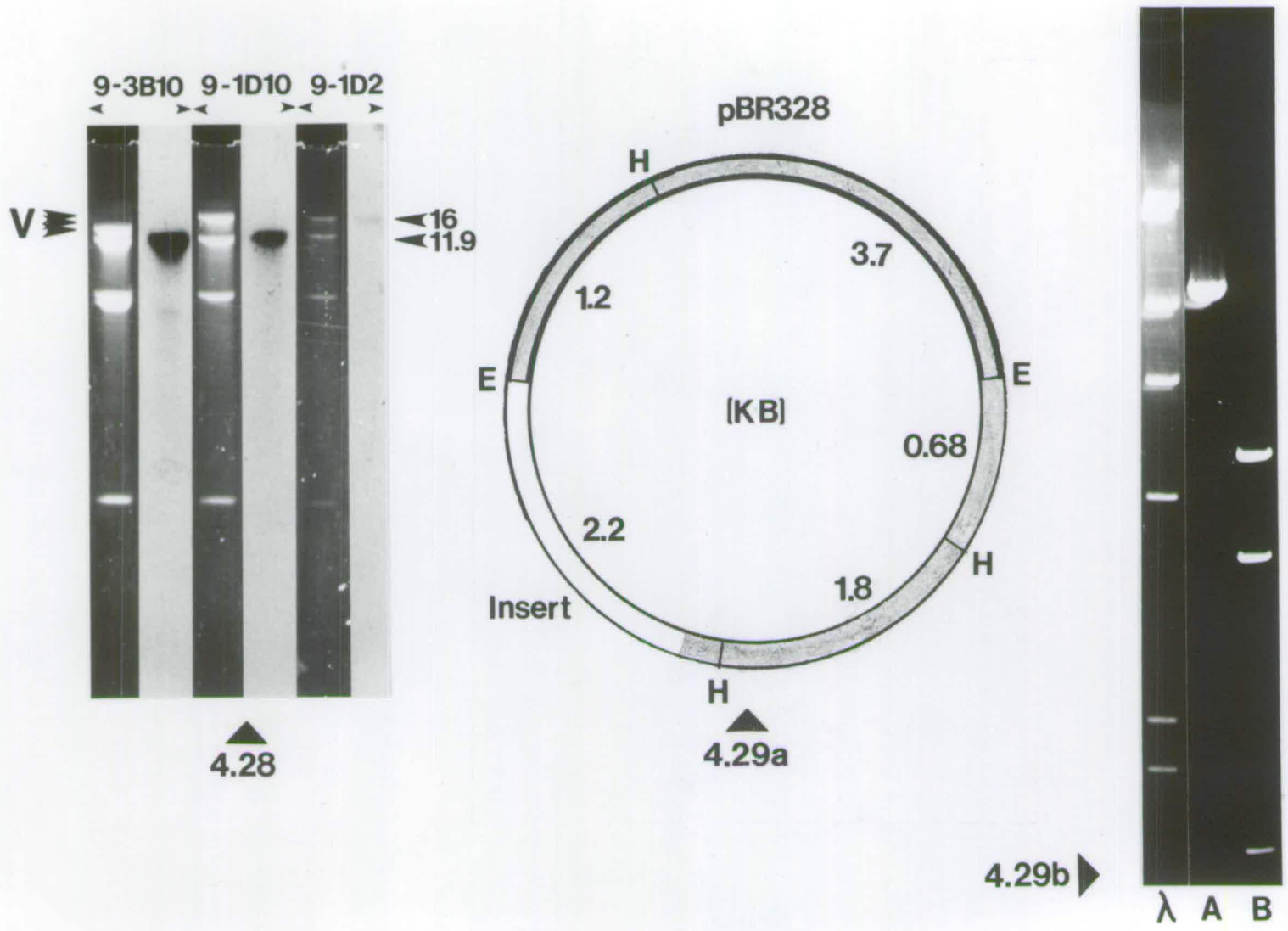
Fig 4.28 Hybridisation of M13 clone 640 to SmaI - digested COB homologous cosmid DNA.

Vector - containing fragments are marked (V).
Note hybridisation to the 11.9KB SmaI fragment of mt DNA.

Fig 4.29 Location of SmaI sites in pbE4.

4.29a Restriction map of pbE4 showing sequenced regions (shaded) where SmaI sites cannot lie.

4.29b Single (lane A) and double (lane B) digestion of pbE4 with SmaI or SmaI plus EcoRI. See text for details.


9-3B10
9-1D10
9-1D2
V
16
11.9
4.28
pBR328
H
3.7
1.2
E
E
(KB)
0.68
2.2
H
1.8
Insert
H
4.29a
4.29b
λ
A
B

hybridises to the M13 clone 640. The portion of the gene contained in clone 640 cannot therefore lie within 7 KB of the rightmost SmaI site. Since this portion of the gene has been shown to fall between 2.8 and 3.5 KB of a SmaI site, this data unequivocally places the maize COB gene towards the left hand end of the 11.9 KB SmaI fragment and in the orientation shown in Figure 4.30.

4.5.3 Map position of COB with respect to other maize mitochondrial genes

The position of the maize COI gene has been mapped precisely on the mt DNA restriction map (P. Isaac, pers. commun.) using a similar approach to that described for COB. The restriction map between these two points is known (D. Lonsdale, pers. commun.), showing that the two genes are encoded on the same DNA strand and their 5' termini are separated by 43.1 KB (Fig. 4.30). Brown et al (1983) have shown that maize mt DNA contains a sequence homologous to the human URF-1 gene. This locus has been mapped to a 7.0 KB SmaI fragment adjacent to the 11.9 KB fragment containing COB (Fig. 4.30). The precise position and orientation of this gene are not yet known.

The region of the maize mt genome between the COB and COI loci also contains a sequence homologous to the chloroplast RUBP carboxylase (large subunit) gene (Lonsdale et al, 1983). The maize mt DNA seems to contain a complete copy of this gene, which is probably not expressed in vivo, on the opposite DNA strand from COB and COI. Its 3' terminus is approximately 7 KB from the 5' terminus of COI (Fig. 4.30). This region also contains a 12 KB sequence which is homologous to part of the inverted repeat from chloroplast DNA (Stern and Lonsdale, 1982). In chloroplasts this sequence contains a 16S rRNA gene and the coding sequences of $tRNA^{Ile}$ and $tRNA^{Val}$. These mapping data are summarised in Fig. 4.30, and a representation of the map of the whole maize mitochondrial genome is shown in Fig. 4.31.

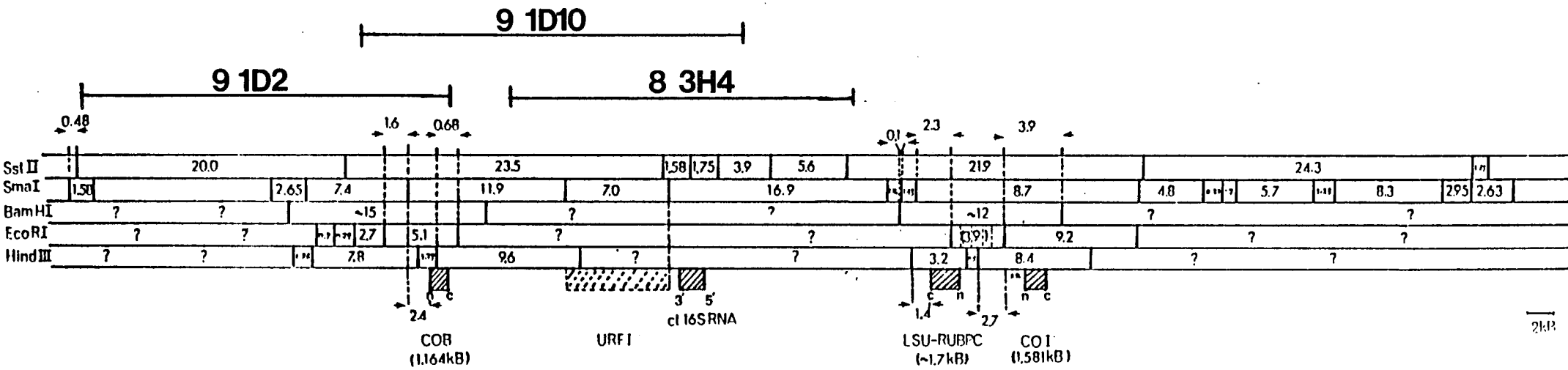


Fig 4.30 Restriction map of maize mt DNA sequences linking COB and COI.

The gene encoding COII is separated from all the other protein-coding genes by at least 100 KB of DNA. A continuous map has been constructed by the group of Lonsdale, linking these genes in one direction. It has not yet been possible to link the genes in the other direction and complete the circular map because problems have been encountered linking the two ends of the map (see Fig. 4.31).

Maize mitochondrial genes were mapped by hybridisation of specific DNA probes to cosmids containing restriction fragments of mt DNA whose location on a physical map was known. This has allowed, for the first time, a genetic linkage map of higher plant mitochondrial genome to be constructed. This had not been previously possible because:

1) Non-lethal mitochondrial mutations for use as marker loci in genetic crosses were not available (cf. S.cerevisiae).

2) The large size of the higher plant mitochondrial genome precludes complete sequence analysis (cf. the mammals).

3) No cell-free system for mt DNA exists to allow the translation products of different segments of mt DNA to be identified.

The most striking feature to emerge from this analysis is the large intergene distance which separates all the mapped loci. Even the two closest (COB and URF-1) are nearly 10 KB apart, whereas COII is over 300 KB - four times the size of the entire S.cerevisiae mitochondrial genome - from the nearest mapped protein coding gene. It seems unlikely that the 'spacer' DNA is filled with active genes, since isolated maize mitochondria only synthesised 20-30 polypeptides in vitro (Leaver and Forde, 1980), the genes for which would only occupy ca. 30 KB. Furthermore, no large ORFs have been detected near to any of the sequenced protein-coding genes in maize mt DNA. It would be interesting to probe mt RNA with a selection of clones covering different portions of the maize mt genome, in order to assess which

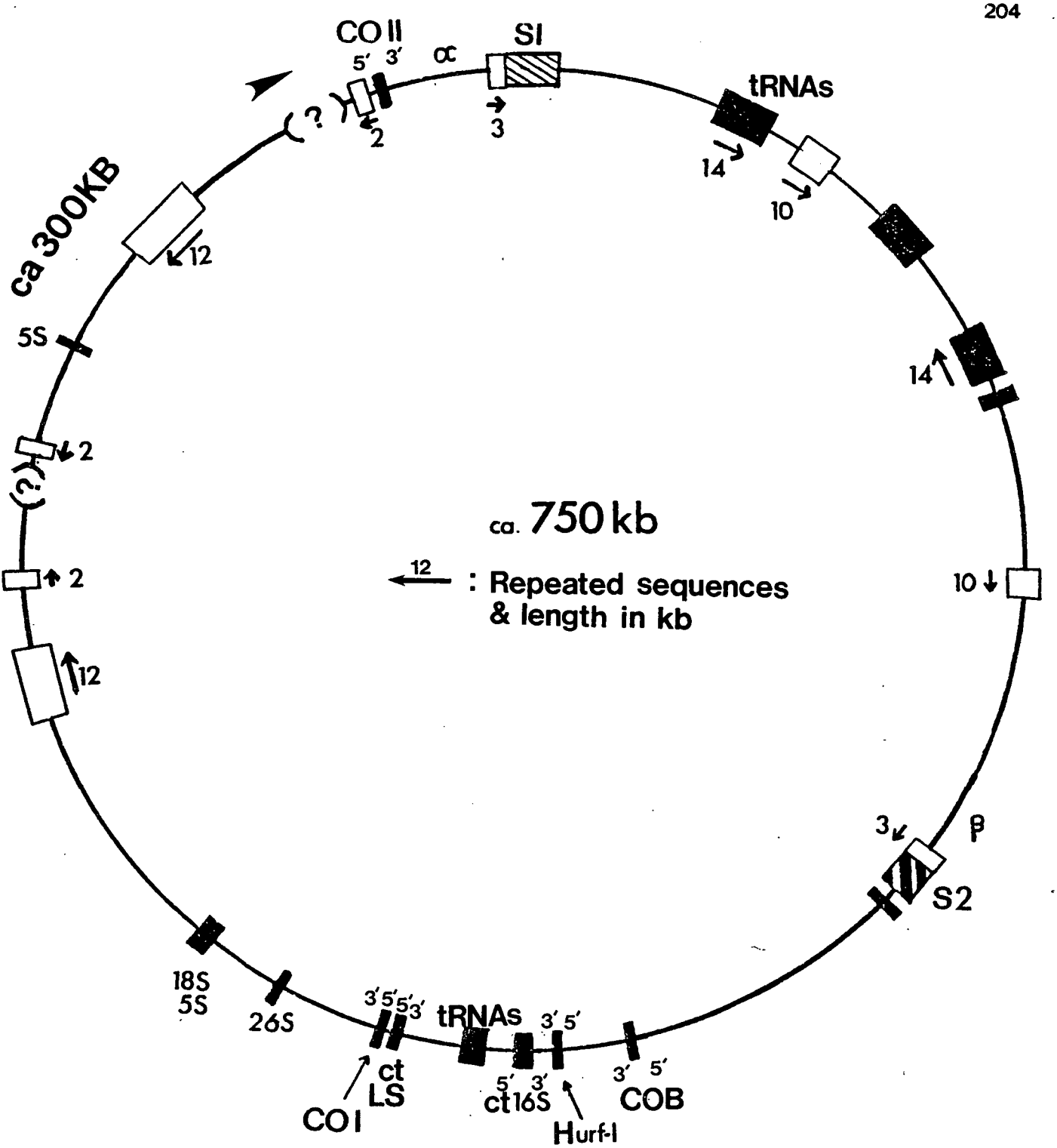


Fig 4.31 Sketch map of the maize mt genome showing known coding regions. The major circle of the tripartite structure (see Chapter 1) is shown.

From D. Lonsdale (pers. commun.).
The area indicated between brackets has not been mapped accurately but could be shorter than the ca 300KB indicated here.

regions were being actively transcribed and which could therefore contain active genes. The mapping data (Fig. 4.31) suggest a possible 'clustering' of the protein coding genes, although the spacing of these genes is still wide and the COII lies outside this region. The order of the genes in maize mt DNA (COII → COB → URF1 → COI (→ COII)) is topologically equivalent to that in Aspergillus (Brown et al, 1982), man and S.cerevisiae (Borst and Grivell, 1981) although no homologue to URF-1 has yet been described in yeast mt DNA. It is too early to decide whether this conserved grouping of genes is of any major significance because too few genes have been mapped in maize mt DNA. Major rearrangements in the positions of other genes in the mt DNA seem to occur between yeast and the mammals (Borst and Grivell, 1981).

Now that genes have been mapped in the mitochondrial DNA of fertile maize (normal cytoplasm), it will be possible to compare arrangements in other maize lines and investigate their possible phenotypic implications. It will also be of interest to investigate the arrangement in other plant species and therefore assess the structural constraints placed on gene reassortment during mitochondrial evolution.

4.6 SUMMARY AND OUTLOOK

4.6.1 Summary: general aspects of the COB gene and apocytochrome b in maize

The isolation and sequence analysis of the maize COB gene has made a significant contribution to our overall understanding of the higher plant mitochondrion. By comparison of this sequence with that of the COB gene in other eukaryotes, and also by comparison with the sequences of other maize mitochondrial genes, a number of predictions about the structure of higher plant mitochondrial genes and their expression could be made. Furthermore, by comparison of the predicted

amino acid sequence of the apocytochrome b polypeptide encoded by the COB gene with that of other apocytochromes b, several deductions about the structure and function of this polypeptide in maize could be made. The findings of these analyses are summarised below:

1) The maize COB genes does not appear to contain introns, unlike the maize COII gene (Fox and Leaver, 1981) and unlike the COB gene in many fungi (Nobrega and Tzagoloff, 1980; Waring et al, 1981; Citterich et al, 1983). The maize COI gene does not contain an intron (P. Isaac, pers. commun.) so it appears that mitochondrial genomes cannot be assigned exclusively to two categories, i.e. a mammalian type (no split genes) or an archetypal fungal class (COB and COI have introns, COII is not split). Furthermore, the lack of introns in both COB and COI precludes the possibility of a cob-box type phenomenon in maize mitochondria (Jacq et al, 1982). The finding (Hiesel and Brennicke, 1983) that the COII gene in Oenothera lacks an intron suggests that introns in plant mitochondrial genes may be the exception rather than the rule.

2) Codon usage in the maize COB gene resembles that in COII which employs a genetic code very similar to the 'universal' codon assignment. Specifically, in COB, UGG is the preferred tryptophan codon and UGA triplets, which encode tryptophan in all other mitochondria, are absent, suggesting they could act as termination codons. A single CGG codon occurs in the coding sequence and this has been translated as tryptophan in view of the findings of Fox and Leaver (1981) and of Heisel and Brennicke (1983). No other nonstandard codon assignments were predicted from sequence analysis of this gene. Examination of the nucleotide usage distribution in the third codon position of COB and other known higher plant mitochondrial genes has revealed a bias in favour of T. This is different from the fungi (Waring et al, 1981) and from the mammals (Anderson et al, 1981). It could be a

useful criterion in assessments of the protein-coding potentials of ORFs in higher plant mt DNA, detected by sequence analysis outside identified genes.

3) Transcription of the COB gene in maize, as revealed by 'Northern' blot analysis, is complex and several high molecular weight RNA species hybridise to COB specific DNA probes. The most abundant transcript, possibly the mature mRNA, is about 2.2 Kbases long. This complex pattern of transcription and the presence of an mRNA much longer than the gene which encodes it is also a characteristic of the transcription of yeast genes, whether or not they contain introns (Tzagoloff, 1982). The split COII gene from maize displays a similarly complex pattern, but the COI gene, which contains no introns, apparently only hybridises to a single transcript. This transcript is, again, considerably longer than the gene. A model has been proposed to explain these findings, whereby a single promoter and terminator sequence specify transcription of a single RNA species. This is extensively processed via smaller intermediates, possibly including splicing out of mRNAs for co-transcribed genes, to the mature COB and COII messages. The COI promoter and terminator sequences would, in this model, be found at the same positions as the termini of the mature mRNA so additional processing would be unnecessary. The transcript pattern of COII would probably be additionally complicated by the presence of an intron.

4) Comparison of 'upstream' sequences, 5' to the initiation codon, of all the sequenced plant mitochondrial genes has revealed the presence of a short sequence which could basepair with a complementary sequence near the 3' terminus of the maize 18S rRNA. The sequence shows between three and five matches to the complementary sequence 5'UGAAUCCU in the rRNA, and finishes 10-20 nucleotides from the AUG

codon. The sequence in the rRNA terminates two nucleotides from the end of the molecule (Chao et al, 1983) and is in the same relative position as the 'Shine and Dalgarno' sequence (Shine and Dalgarno, 1974) $^{5'}$CACCUCCU in the E.coli 16S rRNA. The latter sequence has been shown to assist in binding of ribosomes to the mRNA (Steitz and Jakes, 1975). In view of the strong primary and secondary sequence homology between the higher plant 18S and the E.coli 16S rRNAs (Spencer et al, 1983) it is postulated that these subunits could fulfil similar roles in the respective ribosomes. Hence the 3' terminal sequence of the maize 18S rRNA could be involved in interactions with the conserved complementary sequence upstream from the mitochondrial gene initiation codons. This hypothesis is supported by the observation that the conserved upstream sequence always contains at least three continuous nucleotides complementary to the portion ($^{5'}$UGAA) of proposed Shine-Dalgarno sequence which is specific to higher plant mt 18S rRNA.

5) In all the higher plant mt genes studied, the initiation codon (AUG) is followed by the nucleotide residue A. This seemingly prokaryotic feature could be involved in four-base pair interactions with the anticodon loop of the initiator $tRNA_f^{met}$ (Taniguchi and Weissmann, 1978). Both this nucleotide, and the proposed Shine-Dalgarno sequence, could lend specificity for ribosome binding and translation initiation to the AUG codon occurring at the start of the gene. This could be important in view of the long 5' leader sequences which are likely to exist in the mRNAs, predicted on the basis of transcript analysis. Identification of genuine protein-coding ORFs could also be facilitated by searches for these ribosome-binding and translation initiation signals.

6) The position of the COB gene on a preliminary restriction map of the maize mitochondrial genome has been ascertained by hybridisation of specific probes to a cosmid clone library (prepared by D. Lonsdale). By similar approaches, the COI, COII and URF-1 loci have been mapped. The gross gene order is the same as found in fungi and the mammals but the intergenic distances are much larger (10-300 KB). Some of these loci appear loosely grouped but the functional significance of this is unknown. These are the first mapping data obtained for any higher plant mitochondrial genome, and comparison with maps from other organisms should yield interesting information on higher plant mitochondrial evolution.

7) Analysis of the predicted polypeptide from the COB gene indicates that apocytochrome b in maize is 388 amino acids in length (assuming no processing) and will have an M_r of 42,900. This is very similar (± 6 residues) to apocytochrome b in all the other eukaryotes which have been studied. Hydrophobicity plots (Kyte and Doolittle, 1982) show that the maize polypeptide displays a similar profile to all other analysed apocytochromes b. This suggests firstly that the amino and carboxy termini are not encoded by separate exons and secondly that the maize protein could fold and function in the inner mitochondrial membrane in the manner proposed for these other apocytochromes b (Saraste, 1983).

8) 'Invariant' histidine, arginine and lysine residues have been identified which are present in the same relative positions of the maize and all other predicted apocytochrome b sequences. This finding supports the model for protohaem binding proposed by Saraste (1983) and Widger et al (1983) and indicates that a single cytochrome b polypeptide probably contains two protohaem moieties, each in a rather different hydrophobic environment.

9) The maize, and all other apocytochromes *b*, display sequence homology to two chloroplast DNA-encoded proteins, cytochrome b_6 and the '17K' protein (Widger *et al*, 1983). These two chloroplast genes are adjacent on the ct DNA map and the polypeptides can be aligned end to end with homologous sequences in the mitochondrial apocytochromes *b*. A gap of six amino acids separates the carboxy terminus of the aligned b_6 polypeptide and the amino terminus of the aligned '17K' protein. The maize apocytochrome *b* sequence, uniquely, contains a methionine codon corresponding in position to the initiation methionine of the aligned 17K protein. This could lead to speculation on the origin of the mitochondrial gene, particularly since DNA transfer from the chloroplast to mitochondria appears to occur with regularity in the higher plants (Stern *et al*, 1983). It might be proposed that the chloroplast genes were transferred to the mitochondrion of a primitive progenitor aerobe where they subsequently become fused and expressed as a single mitochondrial gene.

4.6.2 Future prospects

The sequence analysis of *COB* and a small number of other genes from higher plant mitochondria has opened the door to a large and interesting area of molecular biology. It is certain that the sequence analyses of all the 'standard' mitochondrial genes (including *COI II*, ATPase subunits 6 and 9, and possibly some URF sequences) will be completed in the near future, both in maize and in other higher plants. As the gene probes become available, these loci will be positioned on the appropriate mitochondrial genome maps. These analyses will however merely repeat work carried out many years previously in the mammals and fungi. They may not explain any of the unique and puzzling features of higher plant mt DNA - for example, its disproportionately large and variable (Ward *et al*, 1981) size, and the

presence of chloroplast DNA in the mitochondrial genome. DNA sequence analyses alone will not answer these questions and a variety of biochemical and biophysical analyses will be required before the functioning of the higher plant mitochondrion is more fully understood. Two such approaches which could be used in the near future are described below.

4.6.2.1 Identification of novel higher plant mitochondrial genes.

The higher plant mitochondrion synthesises in vitro more polypeptides than do mitochondria isolated from either fungi or from mammals (Leaver and Forde, 1980). Although the extra genes required to encode these proteins could not possibly saturate the large mitochondrial genome, their identification may give some insight into the particular protein requirements of this organelle in higher plants. However, the identification of these genes is limited by the availability of hybridisation probes for known mitochondrial genes from other organisms. Clearly, because a large number of genes exist in higher plant mitochondria which are probably not found in mitochondria from any other eukaryotes, alternative means for their identification must be sought.

Hack and Leaver (1983) have used such an alternative approach to demonstrate that the α subunit of the mitochondrial F_1 ATPase in maize is the product of mitochondrial protein synthesis. This unexpected finding has been confirmed by 'heterologous' hybridisation of α subunit DNA probes from E.coli or the yeast nucleus to maize mt DNA. Specific DNA restriction fragments containing this gene in maize have now been identified and cloned (S. Dunbar, pers. commun.)

With the availability of complete cosmid libraries of maize mt DNA, it should be possible to identify transcriptionally active regions of the mt genome simply by hybridisation of these mapped cosmids

(D. Lonsdale, pers. commun.) to 'Northern' blots of mt RNA. Once identified, more specific RNA-hybridising mt DNA fragments could be cloned and sequenced, thereby revealing the presence of genes which could not be identified by any other means. This approach, unlike that used by Hack and Leaver will not, however, necessarily allow identification of the polypeptide encoded by this gene. Possible means to achieve this are discussed in the next section.

4.6.2.2 *In vitro* expression of higher plant mitochondrial genes.

No reliable cell free system for the transcription and translation of mitochondrial genes has yet been devised (Leaver and Gray, 1982). The discovery of prokaryotic-like putative ribosome binding sequences uniquely in higher plant mt mRNAs suggests that an *E.coli* cell free system, supplemented by mitoribosomes, may be capable of at least directing protein synthesis from added higher plant mt mRNAs. Experiments are in progress in this laboratory to test this hypothesis, using hybrid-selected RNAs specific to *COB*, *COI* and *COII*, to programme protein synthesis in *in vitro* systems.

An alternative approach for both the transcription and translation of mitochondrial DNA could make use of bacterial plasmid vectors (Maniatis *et al*, 1982) to express cloned mt DNA sequences using *E.coli* promoters and ribosome binding sites. Such an approach requires previous knowledge about the gene encoded by the mt DNA fragment (for example, its reading frame) but could be useful if direct expression in cell free systems fails. The greatest value of either approach would be the synthesis of mitochondrial proteins in large amounts, from which homologous antibodies could be raised. This would then make possible, for the first time, the unequivocal identification of these proteins in the spectrum of polypeptides synthesised *in vitro* by isolated higher plant mitochondria.

It should be remembered that the identification of mitochondrial genes *per se* will make only a relatively small contribution to a complete understanding of the function of this organelle. The vast majority of mitochondrial proteins are encoded in nuclear genes and imported into the mitochondrion (Gasser *et al*, 1982). The investigation of these genes, and identification of the control factors which co-ordinate expression of the two genomes, is a major area for study in the future and has, as yet, barely been touched upon.

CHAPTER 5 REORGANISATION OF MITOCHONDRIAL DNA IN MALE STERILE MAIZE LINES

5.1 INTRODUCTION

The mitochondrial genome of maize has a higher sequence complexity (600-700 KB; Ward et al, 1981; Lonsdale et al, 1983) than its counterpart in either the fungi (19-108 KB; Borst and Grivell, 1978; Grivell, 1983) or the mammals (16-17 KB; Anderson et al, 1981 and 1982). This complexity does not appear to be made up of large tracts of A + T rich DNA, as judged by the high buoyant density (1.706 g/ml, equivalent to an A + T content of 47%) of the DNA (Shah and Levings, 1974). Hence, as discussed in Chapter 1, the maize mitochondrial genome could theoretically encode many more proteins than any mammalian or fungal mt DNA. Analysis of polypeptides synthesised by isolated maize mitochondria by one- and two-dimensional polyacrylamide gel electrophoresis has shown that up to 50 translation products can be detected (Hack and Leaver, 1983) and are by implication encoded by mitochondrial genes. This is more than the number synthesised by isolated human (26; Attardi, 1981) or yeast (about 8-14; reviewed by Grivell, 1983) mitochondria, but still far fewer than required to saturate the coding capacity of the mt DNA.

The identity of one one of the additional maize polypeptides (the α subunit of the F_1-ATPase; Hack and Leaver, 1983) is known. The remainder are poorly characterised and their identification is hampered by the lack of non-lethal mutations affecting mitochondrial function. The only class of useful marker mutations in maize mt DNA are those leading to cytoplasmically inherited male sterility (cms). The evidence that the determinants specifying this trait reside in the mitochondrion are reviewed by Leaver and Gray (1982), by Levings (1983a) and in Chapter 1 of this thesis.

The mt DNA isolated from male sterile maize lines has been shown to differ significantly in its restriction endonuclease cleavage pattern from that of male-fertile (normal, N) maize lines (see for example, Levings and Pring, 1976). Furthermore, the restriction fragment profiles of the mt DNA isolated from the N and from the C, S and T male sterile cytoplasms are easily distinguished from each other (Pring and Levings, 1978). These restriction enzyme fragment differences probably reflect changes in the organisation of the mt DNAs in these different lines, rather than the effects of partial base methylation (Bonen et al, 1980) or point mutation (Sederoff et al, 1981; Spruill et al, 1981). Somewhat surprisingly, these major DNA rearrangements do not seem to affect mitochondrial function in general, other than during pollen production and anthesis (Warmke and Lee, 1977). In fact, variation within N mt DNA lines has been reported (Levings and Pring, 1977) and also in male fertile lines regenerated from tissue cultured cells of both male sterile (T cytoplasm, Gengenbach et al, 1981; Brettell et al, 1982) and male fertile (R.J. Kemble and R.B. Flavell, pers. commun. in Brettell et al, 1982) maize. It would appear, therefore, that much of the maize mt genome has a non-coding function since it can suffer major rearrangement, frequently without any observable phenotypic effect.

In addition to their specific mt DNA restriction endonuclease digestion patterns, the male-sterile maize lines can be characterised by the synthesis, in isolated mitochondria, of a variety of novel polypeptides not produced by mitochondria from fertile cytoplasms (Forde et al, 1978). Mitochondria from cms T and C lines synthesise polypeptides of M_r 13,000 and 17,500 respectively, which appear to replace polypeptides of M_r 21,000 and 15,500 respectively, in N mitochondria. Cms-S mitochondria synthesise eight novel polypeptides,

of M_r 42,000 to 88,000 (Forde and Leaver, 1980). Nuclear restorer (Rf) genes have been shown specifically to repress synthesis of the 13,000 M_r polypeptide (Forde and Leaver, 1980), but in the tissues studied, do not affect synthesis of C- and S-specific proteins. The M_r 13,000 and 21,000 proteins implicated in T-cms are of particular interest because an involvement of these polypeptides in mitochondrial susceptibility to the T-toxin produced by Dreschlera (= Helminthosporium) maydis race T has been suggested (Forde and Leaver, 1979; Walton, 1983). A causal relationship between these cms-specific proteins and male sterility or toxin susceptibility has not yet been established, however. Neither is it known whether the altered polypeptides result directly from changes in the DNA sequences encoding them, or from more complex differences in post-transcriptional processing, possibly involving factors specified by nuclear genes. It has been shown, for example, that certain nuclear-mitochondrial combinations in Sorghum can lead to male sterility. In one such combination, cms correlates with the synthesis of an altered form of cytochrome oxidase subunit I (Dixon and Leaver, 1982). In this case, the polypeptide difference is almost certainly the result of a nucleo-mitochondrial interaction superimposed on a change in the mt DNA, since the same mitochondria can produce a normal subunit I in different nuclear backgrounds.

It appears, therefore, that higher plant mt DNA in general, and maize mt DNA in particular, is rather unstable and may frequently undergo quite considerable organisational changes. These changes happen to be manifested in male sterile lines, but this may simply be because variation in male fertility is the only phenotypic marker at present available for the classification of higher plant mitochondria. It is by no means proven that these reorganisations are actually involved in male sterility. The aim of the work described in this chapter was to investigate in detail the nature of some of these reorganisations

in the mt DNA from male sterile maize lines. Initially, rearrangements near well-characterised protein coding genes were studied. It was anticipated that these rearrangements might model the types of reorganisation which occur throughout the mitochondrial genomes of male sterile maize, and possibly also give some insight into the molecular basis of cytoplasmic male sterility. Most of the experimental work focusses on the cytochrome oxidase subunit II (COII) gene (formerly mox-1; Fox and Leaver, 1981) because when the work was initiated this was the only maize mitochondrial gene for which complete sequence data was available. Reorganisation of the mitochondrial genome in the vicinity of COB and COI is discussed briefly later in this chapter.

5.2 MITOCHONDRIAL GENOME ORGANISATION 5' TO THE COII GENE IN MALE-FERTILE AND MALE-STERILE MAIZE

5.2.1 Comparison of COII 5' sequences in Normal and cms maize

The entire coding portion of the COII gene from male fertile (N) maize is contained on a 2.4 KB EcoRI fragment of mt DNA. This DNA has been cloned in the vector pBR322 to produce a hybrid plasmid, pZmE1 (Fox and Leaver, 1981). In addition to the COII coding sequences, this clone also contains approximately 100 BP of 5'-, and 700 BP of 3'-flanking sequences (Fig. 5.5). This plasmid, a gift of C.J. Leaver, was nick-translated and used as a hybridisation probe for COII sequences in mt DNA isolated from male fertile (N-type) and male-sterile (C-, S- or T-type) maize.

Maize mt DNA, from N, C, S and T-type cytoplasms was prepared from etiolated coleoptiles, digested with EcoRI and with HindIII and fragments were separated by electrophoresis through a 0.8% agarose gel (see Chapter 2). DNA was blotted to nitrocellulose and hybridised to pZmE1, under stringent (Tm-26°C) conditions. The filter was

fluorographed at -80°C in order to detect mt DNA fragments with complementarity to the probe (Fig. 5.1a, b). Under these conditions, pBR322 (vector) DNA did not hybridise to mt DNA digested with EcoRI (Fig. 5.1c).

Fig. 5.1a shows that within the resolving power of the analytical system used, an identical 2.4 KB EcoRI fragment hybridises to pZmE1 in N, C, S and T-type mt DNA. This indicates that no mt DNA rearrangement occurs within 100 BP of the 5' end of the COII gene, nor within 700 BP of the 3' end. Furthermore, no additional strongly hybridising fragments were detected, suggesting that major segments of the COII and flanking regions are not duplicated elsewhere in the mitochondrial genome (but see section 5.2.2.3).

Hybridisation of pZmE1 to HindIII digested B37 N, C, S and T-type mt DNA (Fig. 5.1b) shows that two major fragments are identified in each case. In both N and S-type mt DNA, these fragments are 5.1 and 1.45 KB. In C and T-type, the 1.45 KB fragment is apparently conserved, but the larger fragment is replaced by smaller ones of 3.7 and 2.8 KB in C and T respectively. The orientation of the large and small HindIII fragments with respect to the COII gene could be determined because the 2.4 KB EcoRI fragment was shown to contain a single HindIII site, 1.9 KB from the EcoRI site nearest the 5' end of the gene (Fox and Leaver, 1981; see also Fig. 5.5). The 1.45 KB HindIII fragment must therefore lie downstream from COII, whereas the larger fragment contains the 5' terminus of the gene. Identical results (not shown) were obtained when mt DNA from the WF9N line was used in place of B37N. Similarly, mt DNA from the nuclear restored (Rf) cms lines B37CRf and B37SRf exhibited seemingly identical restriction fragment profiles and pZmE1 hybridisation patterns to the respective non-restored (B37C and B37S) lines. These experiments

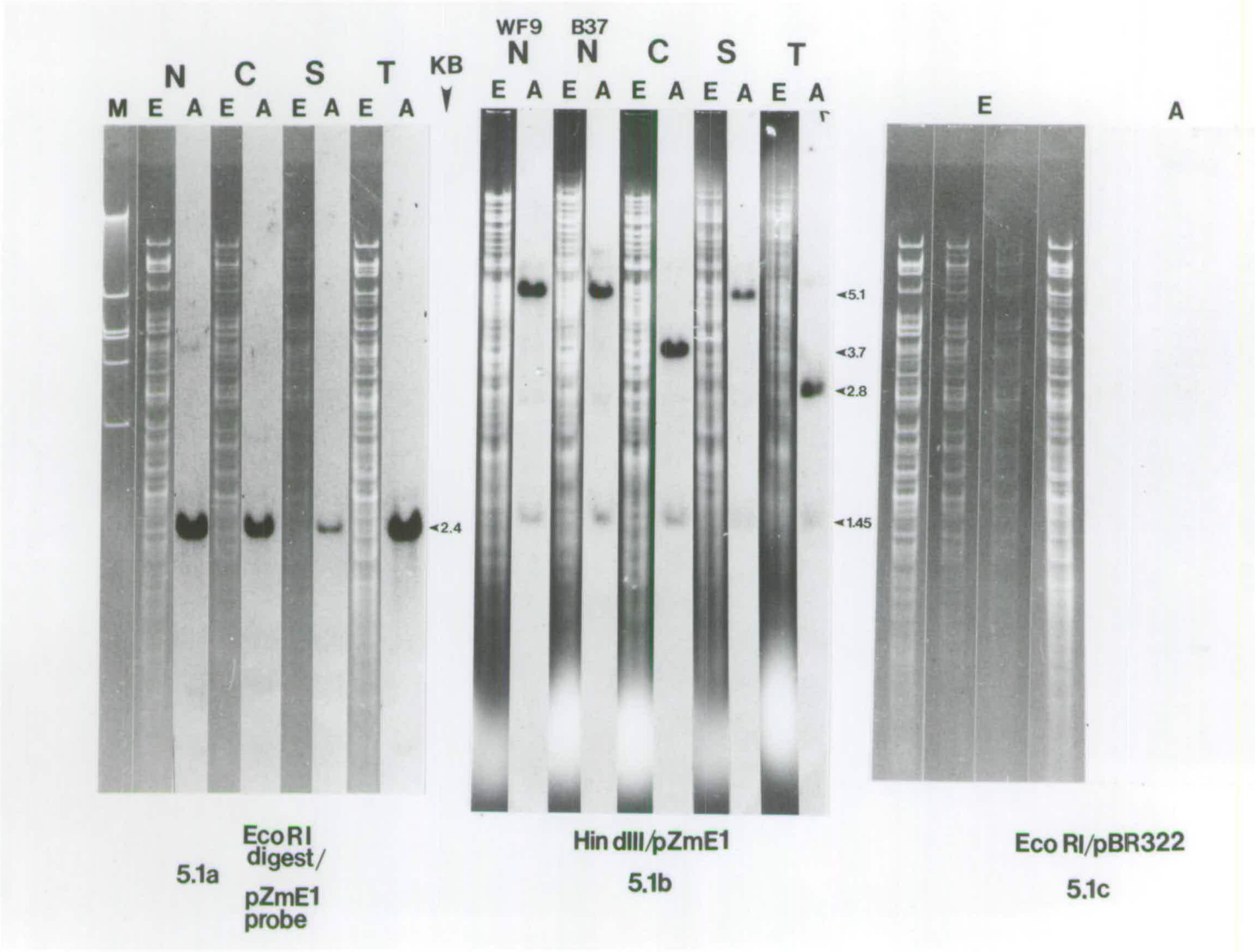
Fig 5.1 Hybridisation of nick-translated pZmE1 and pBR322 to B37N, C, S and T mt DNA.

5.1a EcoR1 digested DNA probed with pZmE1

5.1b HindIII digested DNA probed with pZmE1

5.1c EcoR1 digested DNA probed with pBR322

Abbreviations: M = marker DNA fragments (λ DNA + EcoR1)
E = Ethidium bromide stained electrophoretogram
A = As 'E', after transfer to nitrocellulose, hybridisation and fluorography.


N C S T KB
M E A E A E A E A
2.4
WF9 B37
N N C S T
E A E A E A E A E A
5.1
3.7
2.8
1.45
E A
5.1a EcoRI digest/ pZmE1 probe
Hin dIII/pZmE1
5.1b
Eco RI/pBR322
5.1c

could not, unfortunately, be repeated with B37T Rf because seed of this maize line was not obtainable.

These hybridisation experiments unequivocally shown that a DNA rearrangement has occurred 5' to the COII gene in C and T-type mt DNA. They do not, however, indicate the type of rearrangement occurring in each case. In order to characterise these reorganisations in greater depth, the cloning and mapping experiments described in section 5.2.2 were undertaken. In subsequent sections, only the DNA rearrangement in T-type mt DNA is considered.

5.2.2 Mapping and cloning DNA sequences 5' to COII in N- and T-type mt DNA

5.2.2.1 Isolation of a clone containing the 5.1 KB COII-homologous HindIII fragment from N-type mt DNA.

A clone containing extended sequences upstream from the COII gene was required to enable comparative mapping studies of this region of the N and T mitochondrial genomes to be carried out. To this end, the 2.4 KB EcoRI fragment of mt DNA cloned in pZmE1 was purified from vector sequences by gel electrophoresis and electroelution, then ^{32}P-labelled by nick-translation. This mt DNA fragment was then used to probe for homologous sequences in the pBR328 library of cloned HindIII fragments from B37N mt DNA (section 3.2.3). Following hybridisation, the Whatman 541 paper filters containing a replica of the library were fluorographed, allowing the detection of seven colonies containing probe-specific sequences. Small-scale preparations of plasmid DNA were made from these colonies, treated with HindIII and insert fragment sizes estimated after electrophoresis on a 1% agarose gel. The recombinant plasmids were found to include some containing the 1.45 KB HindIII fragment and others with the 5.1 KB insert. One of the latter was selected, designated pZmH5 (see Fig. 5.5) and used to initiate a larger scale plasmid DNA preparation.

5.2.2.2 Construction of M13 subclones of pZmH5.

A simple restriction map (Fig. 5.2) of pZmH5 was constructed after digestion with HindIII and EcoRI, both individually and in combination. This map was used to guide subsequent cloning experiments, allowing the construction of a series of EcoRI-HindIII and EcoRI-EcoRI subclones containing specific portions of the sequence upstream from COII, including the region which is seemingly reorganised in T. M13 was used as a cloning vector because:

1) the restriction map showed that none of the DNA fragments were >3KB in length; and

2) subsequent clones could be sequenced directly by the chain termination reaction (Sanger et al, 1977).

Prior to M13 subcloning, a plasmid derivative (pZmEH88) of pZmH5 was constructed, containing the whole of the 5.1 KB HindIII fragment except the 1.9 KB EcoRI-HindIII fragment homologus to pZmE1 (see Figs. 5.2) and 5.5). This clone was constructed by digestion of pZmH5 with EcoRI and HindIII, followed by self ligation then selection of clones containing the appropriate inserts (a 2.4 KB EcoRI-EcoRI fragment [not to be confused with the non-identical fragment in pZmE1], and a 0.8 KB EcoRI-HindIII fragment, see Fig. 5.2) from Ap^r Tc^s Cm^s transformants of E.coli HB101.

pZmEH88 DNA was double digested with EcoRI and HindIII and aliquots ligated to M13 mp8 RF DNA digested with EcoRI or, to M13 mp8 and mp9 double digested with EcoRI and HindIII. E.coli JM101 was transformed with a portion of each ligation and five white plaques resulting from each transformation were picked at random and used to initiate simultaneous small scale RF and single stranded DNA preparations (Chapter 2). The RF DNA was double digested with EcoRI and HindIII and fragments were separated by electrophoresis in a 1% agarose gel.

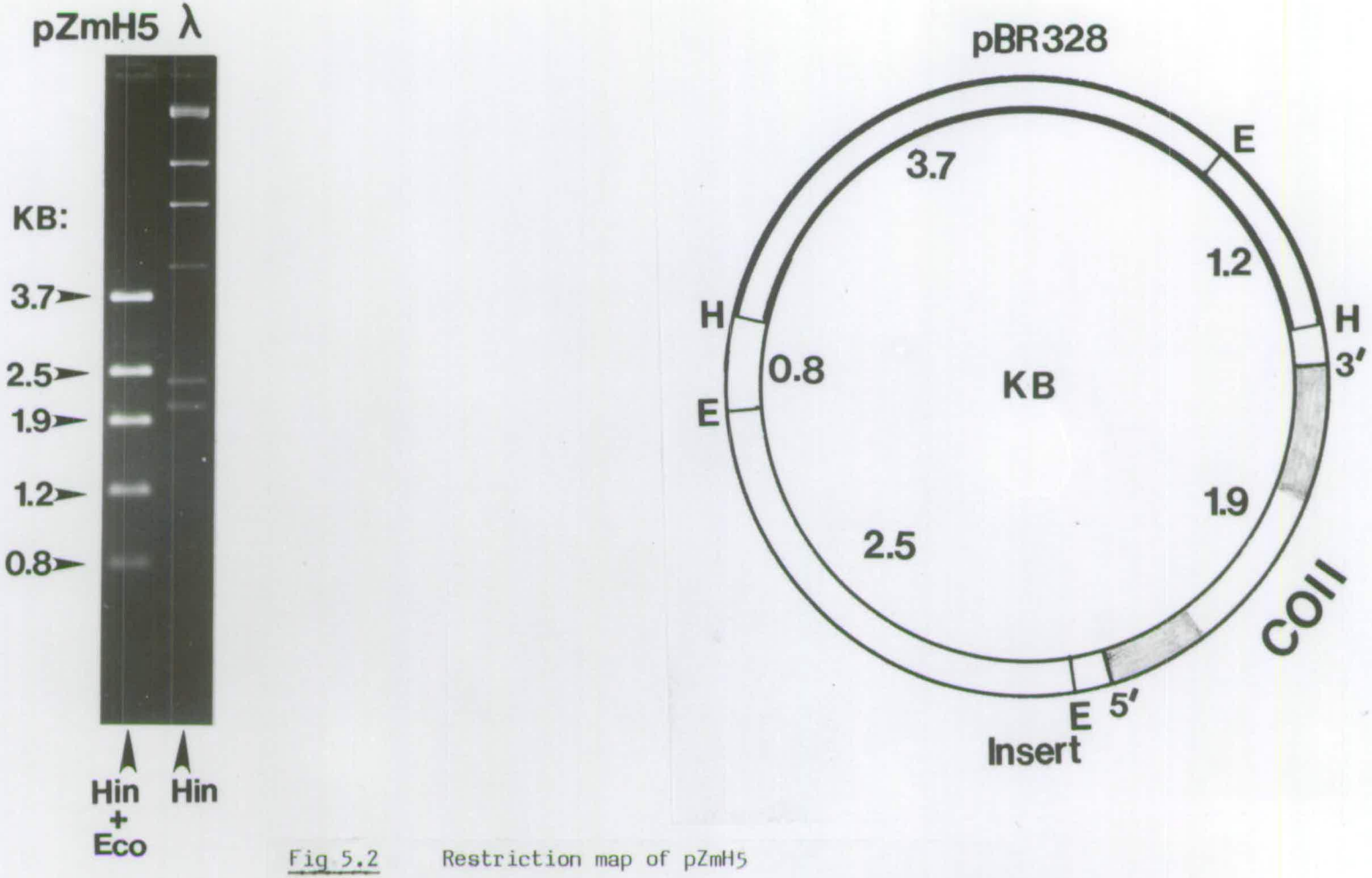


Fig 5.2 Restriction map of pZmH5

The map was constructed after digestion of pZmH5 with HindIII, with EcoR1 and with both enzymes in combination followed by agarose gel electrophoresis to separate and size fragments (see stained gel).
H = HindIII ; E = EcoR1 ; exons shaded.

M13 mp8 clones 1530-1534 were found to contain the 2.4 KB EcoRI fragment, and the 0.8 KB HindIII-EcoRI fragment was contained in mp8 and mp9 clones 1542 and 1544 respectively.

5.2.2.3 Construction of an expanded EcoRI-HindIII restriction map about COII in N- and T-type mt DNA.

B37N and T mt DNAs were each digested with EcoRI and with HindIII then fractionated by electrophoresis on a 0.8% agarose gel (Fig. 5.3a). DNA fragments were blotted to nitrocellulose then hybridised with ^{32}P-labelled pZmE1, pZmH5 and the M13 subclones 1532 and 1542 of the latter (Fig. 5.3b). In this way, a comprehensive map of EcoRI and HindIII sites about COII in N- and T-type mt DNA could be constructed. The main conclusions from this mapping data are summarised below:

1) The sequence reorganisation upstream from COII in T-type mt DNA is not a simple restriction site mutation, since pZmH5 hybridises only to a 2.8 KB HindIII fragment in T. If a novel HindIII site had been introduced into the 5.1 KB HindIII fragment from N mt DNA, 2.8 KB and 2.3 KB HindIII fragments from T mt DNA should hybridise to this probe. The reorganisation must therefore occur less than 2.8 KB from the mitochondrial HindIII site in pZmE1 (see Fig. 5.5), i.e. less than (2.8-1.9) = 0.9 KB from the EcoRI site near the 5' end of COII (see Fig. 5.5).

2) The M13 subclone (1542) of pZmH5 does not hybridise to T mt DNA, strongly suggesting that the reorganisation event involves a deletion of N-specific sequences in T mt DNA. The minimum length of this deletion is therefore 0.8 KB, the length of the EcoRI-HindIII fragment contained in clone 1542 (Fig. 5.5).

3) A number of faintly hybridising DNA fragments are apparent after a longer fluorographic exposure of the filters (arrowed in Fig. 5.3c). These fragments are found within both N and T mt DNA and could result

Fig 5.3 Mapping HindIII and EcoR1 sites 5' and 3' to COII in maize.

5.3a B37N(N) or B37T(T) mt DNA digested with HindIII(H) or EcoR1(E), separated on a 1% agarose gel and stained with ethidium bromide λ DNA digested with HindIII (M) provides molecular weight markers.

5.3b DNA fragments shown in 5.3a blotted to nitrocellulose, hybridised with pZmH5(H5), pZmE1(E1), M13 clone 1532 (32) or clone 1542 (42) and autoradiographed.

5.3c Longer fluorographic exposure of pZmH5 hybridisations to show faintly hybridising submolar DNA fragments.

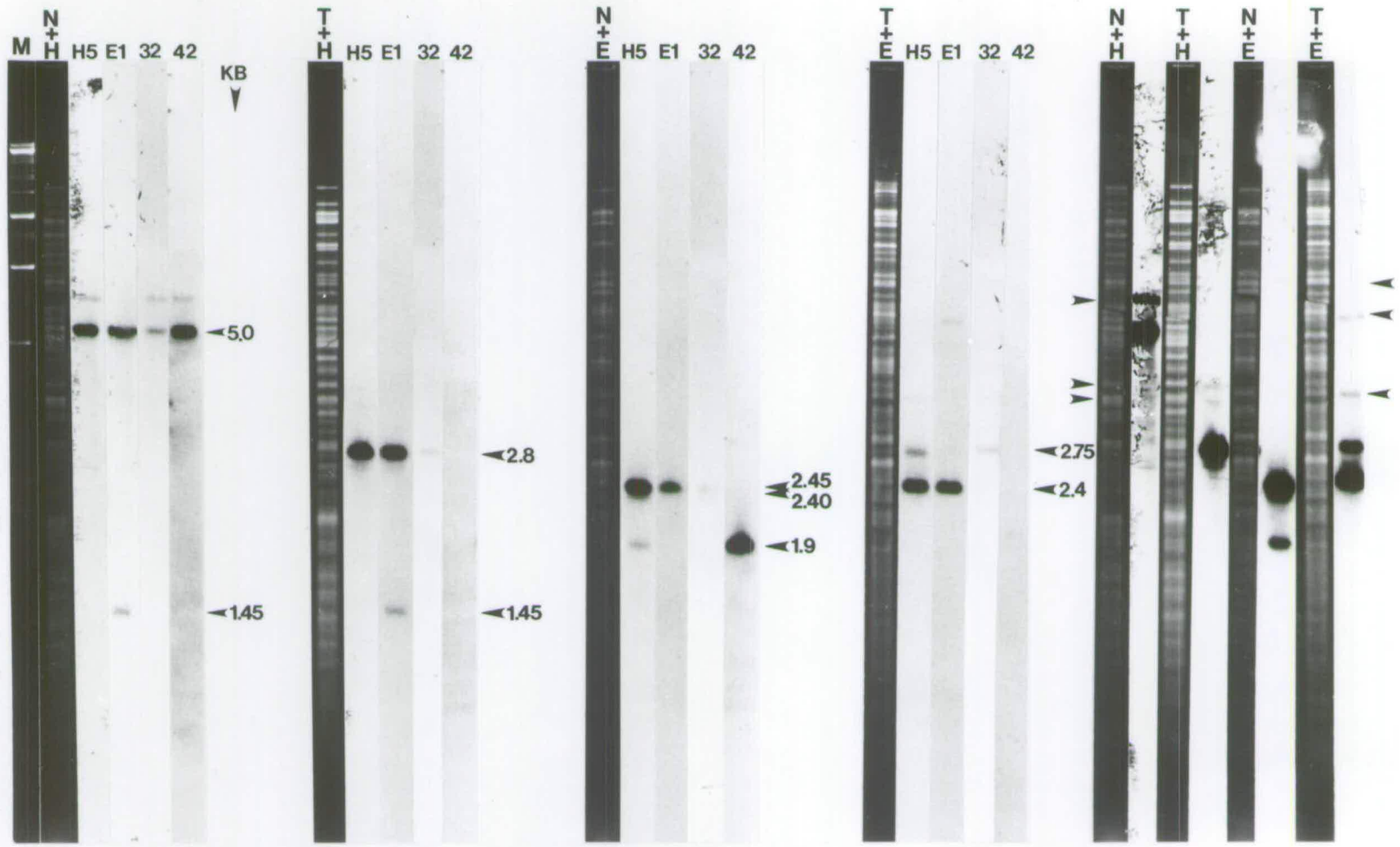

M
N+H
H5 E1 32 42
KB
5.0
1.45
T+H
H5 E1 32 42
2.8
1.45
N+E
H5 E1 32 42
2.45
2.40
1.9
T+E
H5 E1 32 42
2.75
2.4
N+H
T+H
N+E
T+E
a
b
a
b
a
b
a
b
c
(All+H5)

Fig 5.3

from: a) partial duplication of short portions of completely homologous sequence elsewhere in the mt genome; b) duplication of longer sequences having only limited homology to the probe; or c) duplication of segments of the sequence on low abundance, 'submolar' mt DNA species. These alternative hypotheses are evaluated in section 5.4.4. It does appear, however, that this 'microheterogeneity' is not a phenomenon unique to COII, because similar faint bands are also detected with COB probes (section 5.3).

5.2.2.4 Mapping BamHI and XhoI sites near COII.

A map of BamHI and XhoI restriction sites near COII was constructed firstly to extend mapping data 3' to the gene and secondly to locate more precisely the point at which the reorganisation upstream from COII occurs in T-type mt DNA.

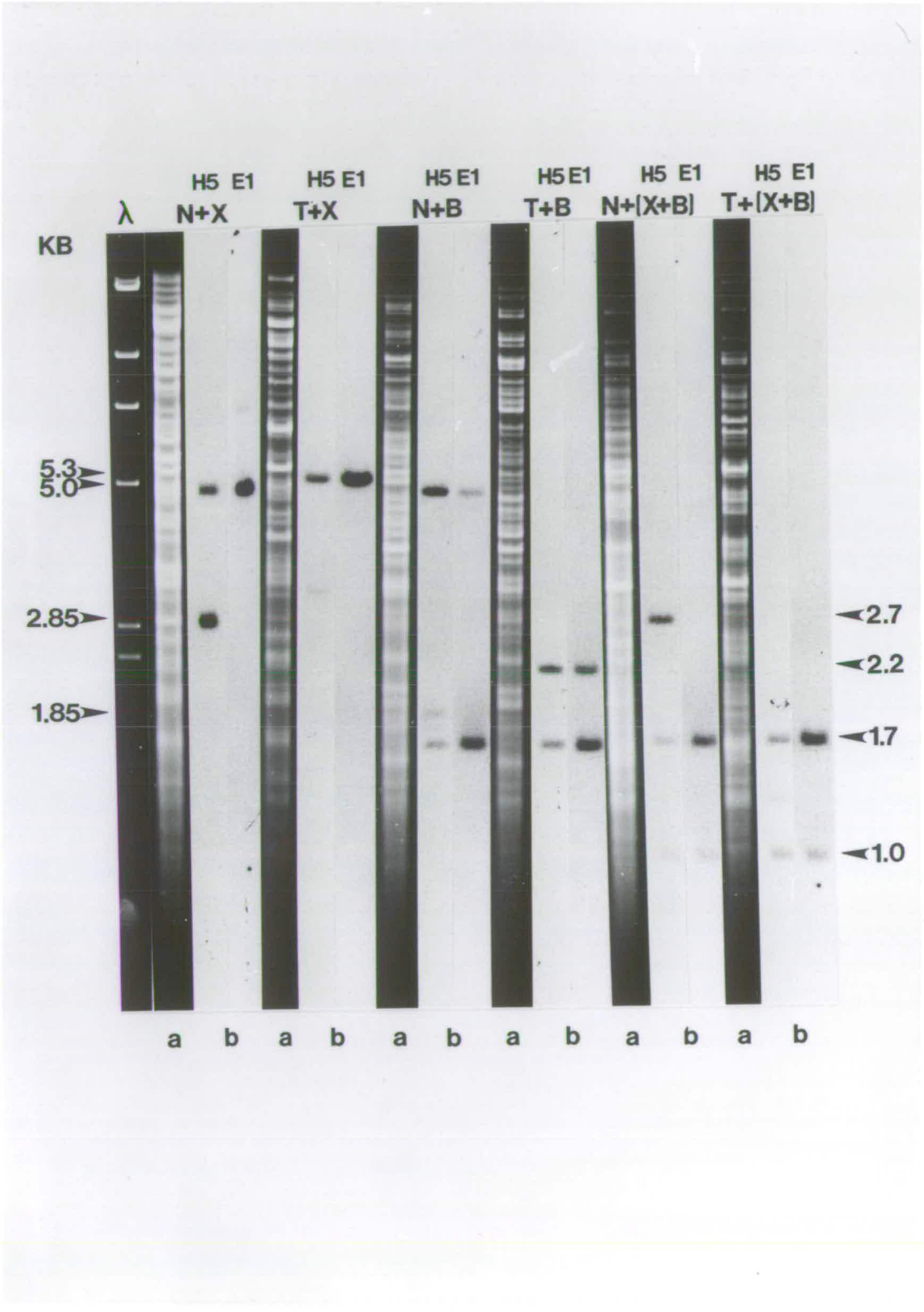
B37N mt DNA was treated with BamHI, with XhoI, and with both enzymes in combination (Fig. 5.4a). DNA fragments were separated by electrophoresis in duplicate tracks on a 0.8% agarose gel then transferred to nitrocellulose. The immobilised DNA was probed with nick-translated pZmE1 and with pZmH5 (Fig. 5.4b) allowing BamHI and XhoI sites to be mapped. Since the 2.4 KB EcoRI fragment of mt DNA cloned in pZmE1 contains no Xho sites and a single BamHI site 0.8 KB from the EcoRI site nearest the 5' end of the gene (Fox and Leaver, 1981; see also Fig. 5.5), these sites could be uniquely orientated with respect to the gene. The combined mapping data for N and T-type mt DNA near COII displayed in Fig. 5.5.

The BamHI and XhoI mapping data confirmed the presence of a rearrangement in T-type mt DNA, and showed that an XhoI site ca. 0.15 KB upstream from COII was common to both N and T mt DNA. The rearrangement therefore occurs between this site and the T-specific HindIII site ca. 0.8 KB further upstream. The restriction map also indicates that

Fig 5.4 Mapping BamH1 and XhoI sites 5' and 3' to the maize COII gene.

5.4a Ethidium bromide stained agarose gel showing mtDNA from B37N(N) or B37T(T) maize after digestion with XhoI(X), BamHI(B) or both enzymes in combination (X + B).

5.4b DNA fragments shown in 5.4a after transfer to nitrocellulose, hybridisation with nick-translated pZmH5 (H5) or pZmE1(E1) and fluorography.


λ
H5 E1
N+X
H5 E1
T+X
H5 E1
N+B
H5 E1
T+B
H5 E1
N+[X+B]
H5 E1
T+[X+B]
KB
5.3
5.0
2.85
1.85
2.7
2.2
1.7
1.0
a b a b a b a b a b a b

a rearrangement may have occurred downstream from COII, beyond the common 1.45 KB HindIII fragment. Both pZmE1 and pZmH5 hybridise to a 5.0 KB XhoI fragment from N mt DNA, but in T the equivalent fragment has an apparent size of 5.3 KB (Fig. 5.4b). Since the XhoI site 5' to COII is conserved in N and T (confirmed by sequence analysis, section 5.2.3.1), the rearrangement must occur beyond the 3' end of COII. This rearrangement has not been investigated in further detail.

The XhoI site immediately upstream from COII provided a useful means to determine the orientation of the EcoRI fragment inserted in M13 mp8 clones 1530-1534 (section 5.2.2.2). RF DNA was prepared from these clones and double digested with HindIII and XhoI. The size of the smaller Xho-HindIII fragment, containing mt DNA and part of the polynucleotide linker from M13 mp8, will therefore be either 150 + 29 or (2400 - 150) + 29 BP, depending on the orientation of the insert. Clones 1530 and 1534 yielded the 2.2 KB fragment, whereas clones 1531, 1532 and 1533 contained the smaller (179 BP) fragment (data not shown; see Fig. 5.6 for restriction map of clone 1532). The latter clones are of greater interest because sequence analysis, which starts near the HindIII site in M13 mp8, will continue directly from the COII upstream sequence determined by Fox and Leaver (1981) from the EcoRI clone pZmE1.

5.2.2.5 Isolation of a clone containing the 2.8 KB EcoRI fragment upstream from COII in T mt DNA.

The mapping data presented in the preceding section shows that a DNA rearrangement, including a sequence deletion, occurs upstream from the COII gene in the mitochondrial genome of maize plants carrying the T-type male sterile cytoplasm. To map precisely the site of this deletion with respect to N mt DNA, it was necessary to isolate and clone a fragment of T mt DNA which spanned the juncture of the

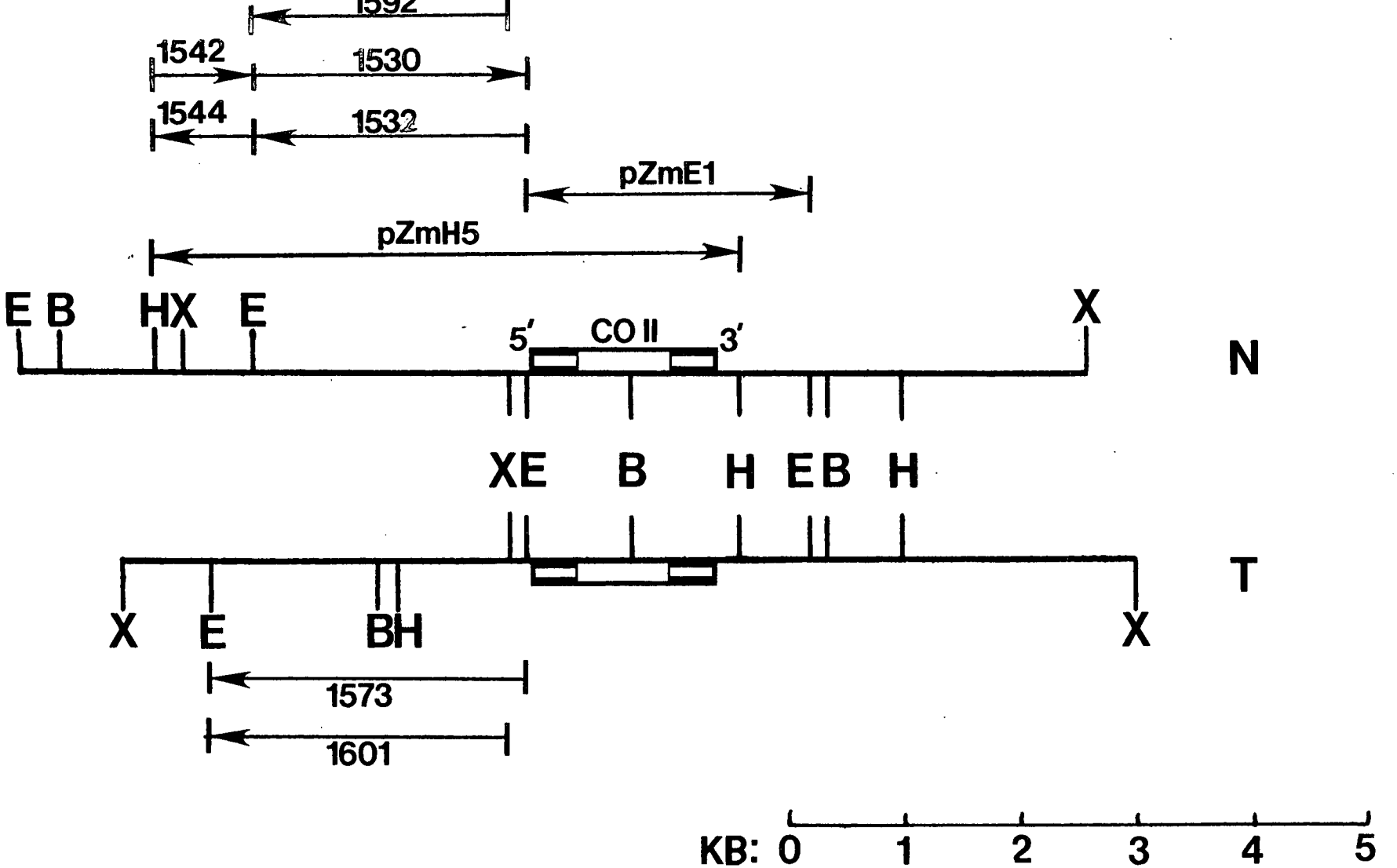


Fig 5.5 Restriction map near COII in B37N(N) and B37T(T) mt DNA
Compiled from data in figs. 5.3 and 5.4

deleted/rearranged sequences. The restriction map (Fig. 5.5) shows that the 2.8 KB EcoRI fragment would contain this sequence. Isolation of a clone containing this fragment would allow a continuous DNA sequence from COII to the point of the reorganisation to be determined.

The construction of a 'library' of cloned EcoRI fragments from T-type mt DNA, in the M13 vector mp8, has been described in Chapter 3. This library was probed, under 'stringent' conditions, with pZmH5. Approximately 20 M13 plaques hybridised to the total library of 1500 recombinants (Fig. 3.3). A selection of these were used to inoculate 1 ml cultures from which single stranded and RF DNA was prepared. Aliquots of the RF DNA were digested with EcoRI to identify clones with inserts of about 2.8 KB. The orientation of the DNA insert could be determined since the 2.8 KB EcoRI fragment contains a single HindIII site 1 KB from the EcoRI site nearest the 5' end of COII (Fig. 5.5). Digestion of the RF DNA with HindIII would therefore give rise to fragments of either 1.0 and 9.0 KB or 1.8 and 8.2 KB; assuming M13 mp8 = 7.2 KB and the EcoRI and HindIII sites in this vector are 29 BP apart (Messing and Vieira, 1982). One clone, designated number 1573, appeared to contain a 2.8 KB EcoRI insert in the desired orientation (i.e. the HindIII site near to the primer binding site, Fig. 5.6).

In order to verify that the EcoRI fragments contained in M13 clone 1573 and its equivalent from N mt DNA (clone 1532) had not become artefactually rearranged during the cloning process, B37N and T mt DNAs were digested with EcoRI and electrophoresed on a 0.8% agarose gel in parallel with 1532 and 1573 RF DNA, also digested with EcoRI (Fig. 5.7a). DNA fragments were transferred to nitrocellulose and probed with nick-translated pZmH5. The autoradiograph (Fig. 5.7b) of the nitrocellulose filter after hybridisation shows:

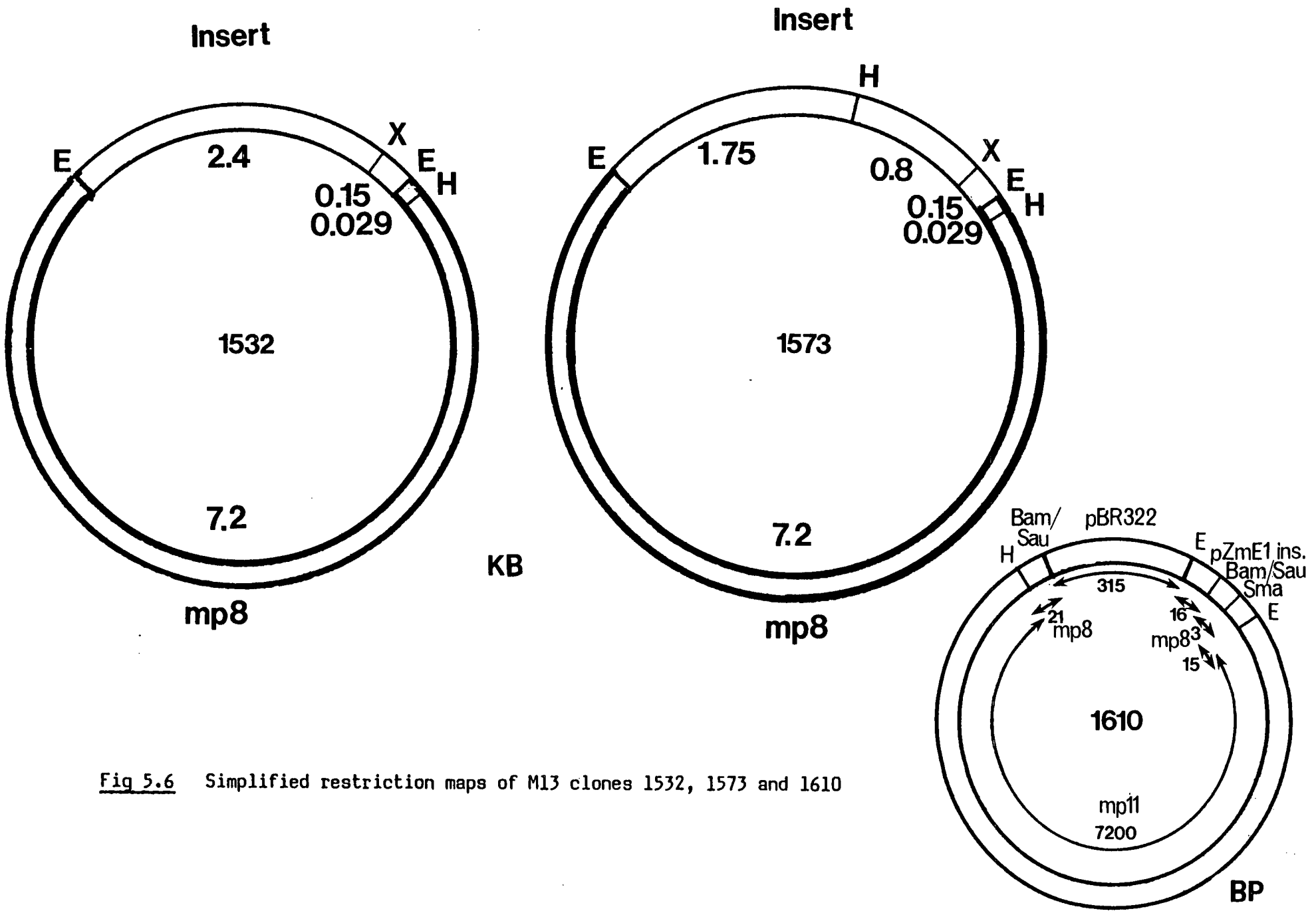


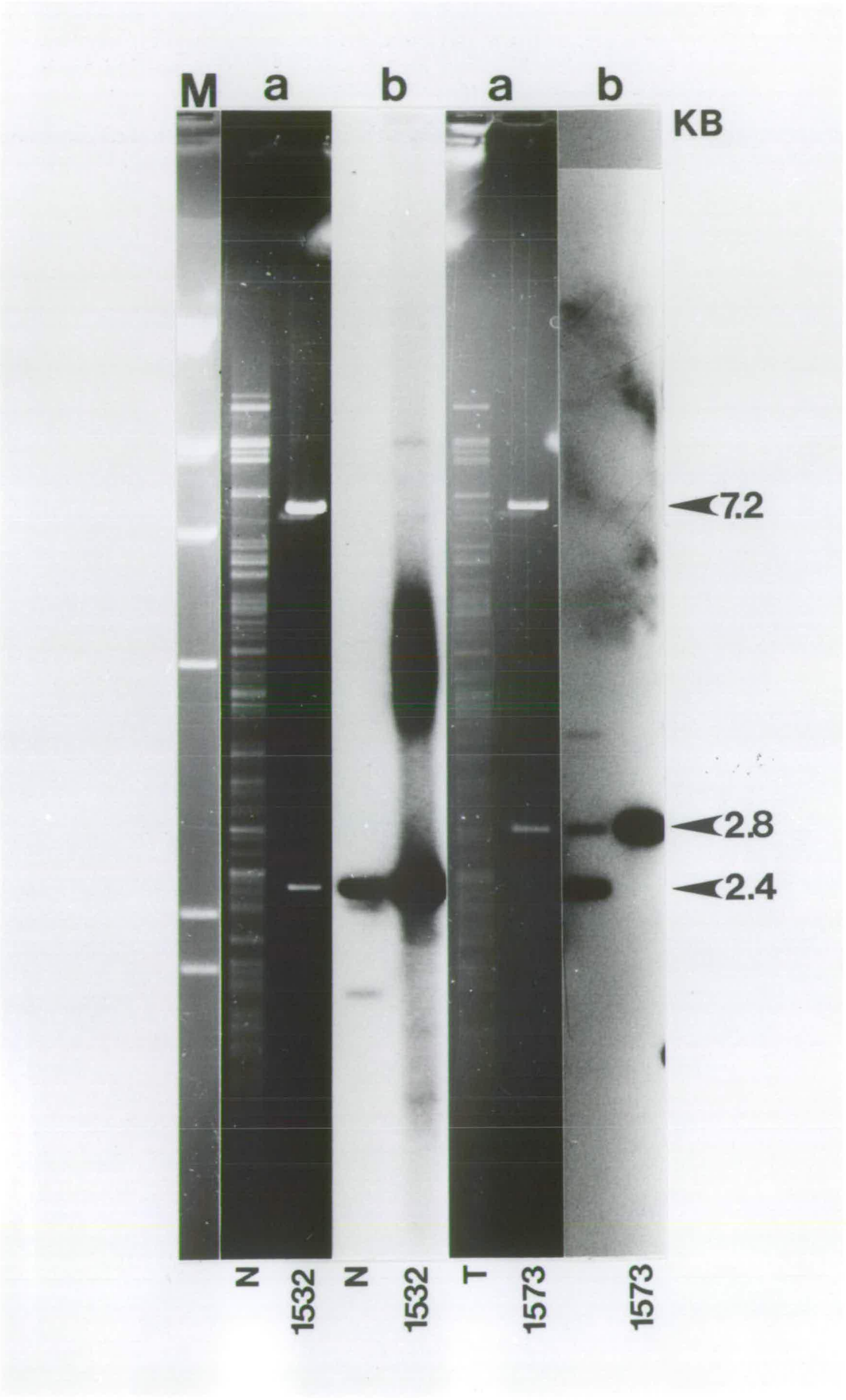
Fig 5.6 Simplified restriction maps of M13 clones 1532, 1573 and 1610

Fig 5.7 Verification of M13 clones 1532 and 1573.

5.7a Ethidium bromide stained gels showing EcoR1 digested B37N(N), 1532RF, B37T(T) and 1573RF DNA. λ DNA digested with HindIII provides molecular weight markers (M).

5.7b DNA shown in fig. 5.7a after blotting and hybridisation with pZmH5.

Note: N and 1532 (16 hr exposure)
T and 1573 (80 hr exposure)


M
a
b
a
b
KB
7.2
2.8
2.4
N
1532
N
1532
T
1573
1573

1) The inserts in clones 1532 and 1573 co-migrate with hybridising fragments from N and T mt DNA of 2.4 and 2.8 KB respectively. The hybridisation signal of pZmH5 to N mt DNA is particularly intense because the probe is complementary to two EcoRI fragments which cannot be distinguished in this electrophoretogram. The common EcoRI fragment hybridising in N and T is the 2.4 KB species containing COII (see Fig. 5.5).

2) Hybridisation of pZmH5 to the 2.8 KB EcoRI fragment in clone 1573 is much weaker than the equivalent hybridisation to clone 1532, even allowing for differences in the DNA loadings of the gel (Fig. 5.7a). This indicates that the sequence homologous to the probe is much shorter in the 2.8 KB DNA fragment than in the 2.4 KB fragment from N. This is consistent with the prediction that the mt DNA rearrangement in T occurs within 0.9 KB of the EcoRI site near the 5' terminus of COII (section 5.2.2.3).

5.2.3 Determination of DNA sequences 5' to COII in N- and T-type mitochondrial DNA.

5.2.3.1 Sequence analysis of clones 1532 and 1573.

The clones 1532 and 1573, from N and T mt DNA respectively, should be similar in sequence at least as far as the common XhoI site ca. 150 BP from the EcoRI site (section 5.2.2.4 and Fig. 5.6). The partial nucleotide sequences of single stranded DNA isolated from these M13 clones was determined by the dideoxy chain termination method (Sanger et al, 1977; see also Chapter 2). A short (2 hr) electrophoresis yielded ca. 200 BP of sequence and showed that N and T mt DNAs are identical in this region: The presence of an XhoI site ($^{5'}$CTCGAG) 174 BP from the EcoRI site was also demonstrated in both N and T. A longer (5½ hr) electrophoresis yielded a further 150 BP of identical sequence. A very long (11¼ hr) electrophoresis showed that the N and T

mt DNA sequences diverged at a point approximately 500 BP from the EcoRI site. However, the sequencing autoradiograph was too blurred to be read in this region so subclones (section 5.2.3.2) were constructed to enable this sequence to be determined with greater clarity.

The sequence data gathered from this experiment is included in Fig. 5.8. The upstream N clones 1530, 1582 and 1541 (sections 5.2.2.2 and 5.2.2.4) were also partially sequenced, but the data is not presented here as the sequence has not been confirmed from analysis of overlapping clones.

5.2.3.2 Construction and sequence analysis of XhoI-SalI derivatives of clones 1532 and 1573.

RF DNA from M13 clones 1532 and 1573 was prepared and double digested with XhoI + SalI, excising a 190 BP fragment containing 16 BP from the M13 polyclonal linker and 174 BP of the mt DNA insert. It was known that the mt DNA in both of these clones contained only a single XhoI site, and assumed that neither insert contained a SalI site. This assumption appears to be justified by the success of the cloning strategy. Since XhoI and SalI generate compatible 5' single stranded protrusions ($AGCT^{5'}$), the digested RF DNA, when self ligated, will yield a high proportion of recombinant molecules lacking the 190 BP XhoI-SalI fragment, by the favoured intramolecular ligation reaction. E.coli JM101 was transformed with a small aliquot of each ligation mixture and eight of the resulting recombinant phage plaques were selected at random and sequenced. Several clones (e.g. 1592 derived from N mt DNA and 1601 from T) yielded sequence continuing from the XhoI site detected by sequence analysis of the EcoRI bounded clones. These data revealed that the N and T mt DNAs were identical up to a point 518 nucleotides from the EcoRI site. Beyond this point, the two sequences were completely dissimilar (see Figs. 5.8 and 5.9). The significance of these sequences is discussed in section 5.4.4.

Fig 5.8 DNA sequences 5' to, and including, COII in N and T maize mt DNA.

The figure displays the determined sequences 5' to the AUG codon presumed to initiate COII (boxed) in N and T mt DNA. The point at which N and T diverge is indicated by a vertical line. The sloping lines indicate the extent of the displaced NmtDNA sequence appearing in T.

The position of the intron in COII is shown by a vertical arrow. The termination codon of COII is indicated by an asterisk (*). The EcoR1 (GAATTC), XhoI (CTCGAG) and Sau3A (GATC) sites near the 5' end of COII which were used in cloning experiments are underlined. Direct repeats near the EcoR1 site (R1 - R4) are indicated by horizontal arrows.

The lower portion of the figure shows a sequencing strategy map used in the compilation of the complete sequence. The sequence shown by the thick bar is from Fox and Leaver (1981).

Note The first 80BP of this sequence must be regarded as preliminary and has not been re-verified.

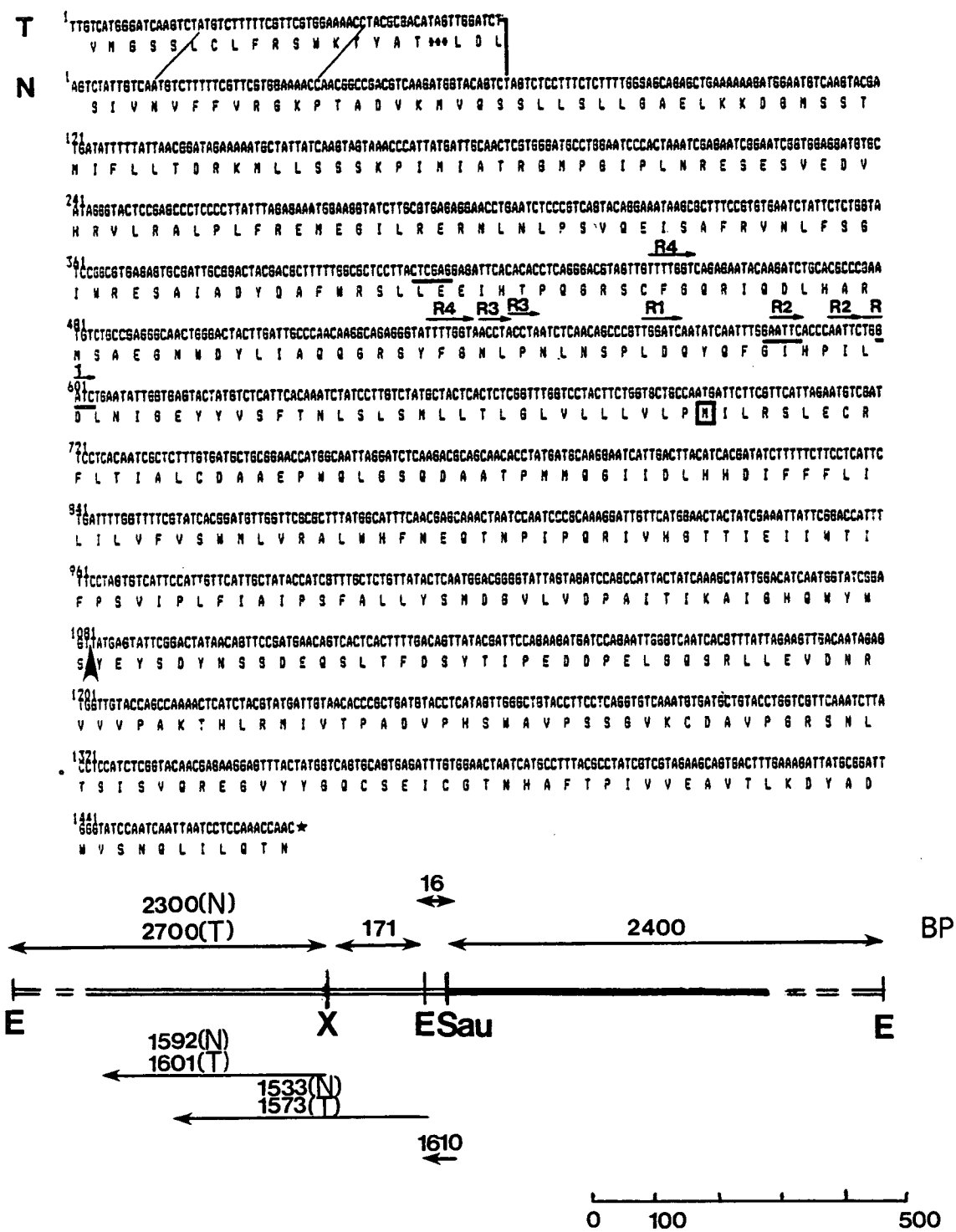
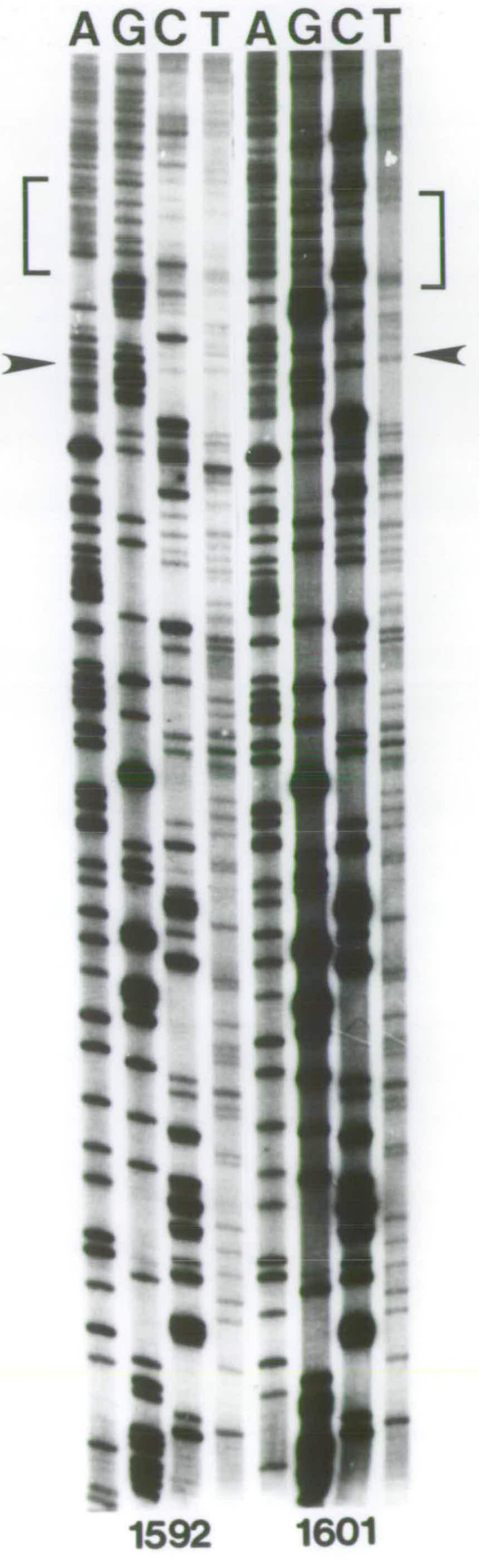

T
N
R4
R4
R3
R3
R1
R2
R2
R
16
2300(N)
2700(T)
171
2400
BP
E
X
ESau
E
1592(N)
1601(T)
1533(N)
1573(T)
1610
0
100
500

Fig 5.9 Sequencing gel of M13 clones 1592 (from NmtDNA) and 1601 (from T) showing the divergence of sequence (arrowed). The extent of the displaced sequence (see fig 5.8) is indicated by brackets.


A G C T A G C T
1592
1601

5.2.3.3 Determination of DNA sequences upstream from COII within pZmE1.

The COII upstream sequences determined in the experiments described above could not immediately be linked to the published COII sequence (Fox and Leaver, 1981) because the position at which the published sequence terminated with respect to the EcoRI site was not known. In an attempt to determine this linking sequence, pZmE1 DNA was digested with EcoRI and ligated to M13 mp8 RF DNA cut with the same enzyme. After transformation of E.coli JM101, ten recombinant M13 plaques were selected and RF DNA was prepared from each. The DNA was digested with HindIII in order to ascertain the orientation of the insert with respect to the single asymmetrically located HindIII site in the 2.4 KB EcoRI fragment (Fox and Leaver, 1981). All the clones contained the 2.4 KB fragment as expected since the vector DNA fragment (4.36 KB, Sutcliffe, 1978) is too large to be stably integrated in M13. However, every insert was found to be in the wrong orientation, i.e. the EcoRI site closest to the 3' end of the gene was nearest to the M13 sequencing primer binding point.

The desired sequence could not be deduced from these EcoRI clones, so use was made of the M13 mp8 subclone VJ73 (constructed by V. Jones for use as an RNA probe). This subclone contains a Sau3A bounded DNA fragment spanning the EcoRI site nearest the 5' end of COII. The fragment contains 315 BP of pBR322 sequence (Sutcliffe, 1978) and approximately 10-20 BP of mt DNA (T.D. Fox, pers. commun.). The Sau3A site in the pZmE1 insert is not shown in the restriction map described by Fox and Leaver (1981), but appears near the 5' end of their sequence.

In order to isolate the DNA fragment used for the construction of VJ73, pZmE1 DNA was digested with Sau3A. Fragments near 300-350 BP were excised and DNA was eluted from the gel as described by Maxam and Gilbert (1977). Following ethanol precipitation, the fragments were ligated to M13 mp8 digested with BamHI and used to transform

E.coli JM101. Single stranded DNA was isolated from ten recombinant white plaques and sequenced. No clones were identified having sequence reading from the mitochondrial Sau3A site, but several (e.g. VJ73) contained the desired insert in the opposite orientation (i.e. reading from the pBR322 Sau3A site).

RF DNA was prepared from this clone and approximately 100 ng was double digested with SmaI and HindIII. This procedure liberates an M13 mp8- pBR322-mt DNA hybrid fragment of about 350 BP. This restricted DNA was ligated to a five-fold excess of mp11 RF DNA also digested with SmaI and HindIII. mp11 was a gift of C. Midgley (Department of Molecular Biology, University of Edinburgh) and was used in preference to mp9 in view of the problems associated with the latter vector (see section 3.2.4). An excess of mp11 DNA ensured that the 350 BP fragments cloned preferentially in this vector and was therefore reversed in orientation with respect to the starting clone (see Fig. 5.6).

A T-screening reaction (section 2.2.12.2.4) of single stranded DNA isolated from eight recombinant phage plaques revealed that one (number 1610) yielded the predicted pattern of T residues and was subjected to the full sequencing reaction. The 16 BP Sau3A-EcoRI sequence derived from this clone is included in Fig. 5.8. Note that the sequence published by Fox and Leaver (1981) (5'CACCCAATCCTCGATC) appears to reach the EcoRI site but that this sequence differs slightly from that contained in clone 1610 ($^{5'}$CACCCAATTCTGGATC). Determination of this sequence allowed the complete nucleotide sequence for some 650 nucleotides upstream from COII to be compiled (Fig. 5.8). An open reading frame found in this sequence is discussed in section 5.3.

5.3 GENOME REARRANGEMENTS NEAR COB AND COI

The sequence analysis of the maize apocytochrome b (COB, see Chapter 4) and cytochrome oxidase subunit I (COI, P. Isaac, pers. commun.) genes has recently been completed. Restriction maps of the mt DNA adjacent to these loci have been constructed from N-type mt DNA (see Fig. 4.18 for the COB restriction map). The experiments described below were carried out to investigate the restriction map about COB in the male sterile maize lines. Results of similar mapping experiments for COI are described briefly later in this Chapter.

Mt DNA prepared from male fertile (N) or male sterile (C, S and T) B37 maize was digested with EcoRI, with HindIII and with BamHI (Fig. 5.10a), fractionated by electrophoresis through a 1% agarose gel, then blotted to nitrocellulose. This DNA was probed with the ^{32}P-labelled clone pbE4 (section 4.2.2.3). This plasmid contains the entire maize COB gene except the last 22 nucleotides, and approximately 3.9 KB of sequence 5' to the initiation codon. The hybridising DNA fragments are shown in the fluorograph (Fig. 5.10b) of the filter. The results can be summarised as follows:

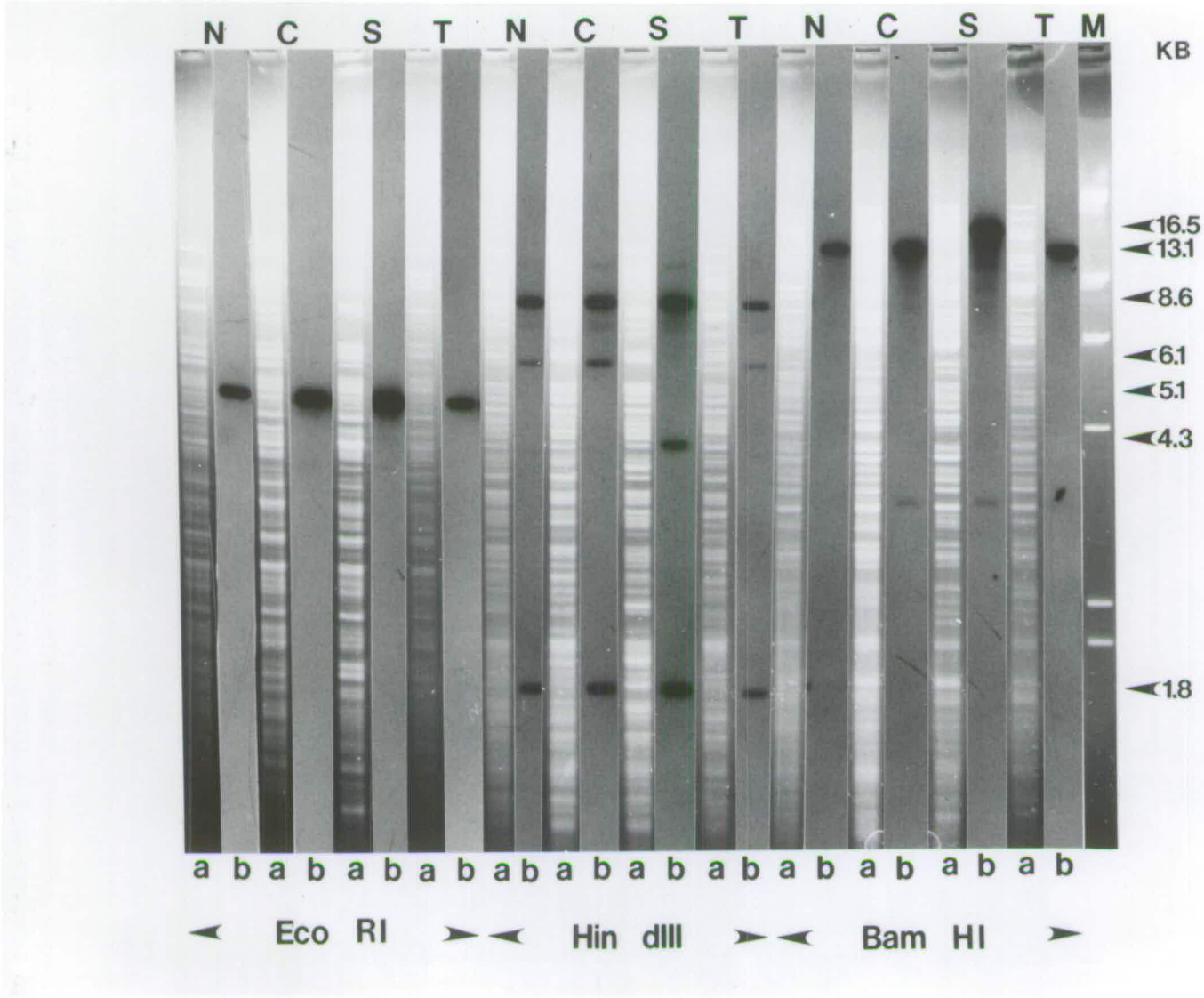
1) A single major EcoRI fragment of 5.1 KB hybridises in all the maize lines, indicating that no major rearrangements have occured upstream from COB in the mt DNA from male sterile maize.

2) The size, and relative intensity of hybridisation, of the three major HindIII fragments from N mt DNA homologous to pbE4 confirms the restriction map predicted by analysis of the BamHI clone pbB2 (Fig. 4.18), i.e. the 3' end of the COB gene lies on the 6.1 KB HindIII fragment, the 5' end on the 1.8 KB fragment [cloned in pZmH1790], and upstream from this is an 8.6 KB fragment.

Fig. 5.10 Genome organisation near the maize COB gene.

5.10a Ethidium bromide stained electrophoretograms of B37N, C, S and T mt DNA digested with EcoR1, HindIII and BamH1. λ DNA digested with HindIII provides molecular weight markers (M).

5.10b DNA fragments, shown in fig 5.10a, after transfer to nitrocellulose, hybridisation with nick-translated pbE4 and autoradiography.


N C S T N C S T N C S T M
KB
16.5
13.1
8.6
6.1
5.1
4.3
1.8
a b a b a b a b a b a b a b a b a b a b a b a b a b
Eco RI
Hin dIII
Bam HI

3) A genome reorganisation occurs downstream from COB in S mt DNA, because the 6.1 KB HindIII fragment is seemingly replaced by an equivalent band of 4.3 KB. Since the 3' end of the COB gene is some 702 BP from the HindIII site (Fig. 4.17), the rearrangement must therefore occur less than 3.6 KB from the 3' end of the gene. T and C mt DNAs apparently resemble N in this region.

4) Hybridisation to BamHI digested mt DNA confirms that a reorganisation occurs near COB in S but not C or T mt DNA. The 13.1 KB BamHI fragment in N, C and T is replaced by a larger species of 16.5 KB in S. Since the BamHI site upstream from COB is closer to the 5' end of the gene than is the HindIII site at the 5' end of the 8.6 KB HindIII fragment (Fig. 4.18), this hybridisation data reaffirms that the genome alteration is downstream from COB. Furthermore, the BamHI site 3' to COB in N mt DNA is approximately 1.9 KB from the terminus of the gene. The rearrangement must therefore lie less than this distance from the end of the gene. Note that the BamHI fragment from N mt DNA (13.1 KB) identified by pbE4 is seemingly smaller than the fragment hybridising to heterologous COB probes (ca. 21 KB, Fig. 4.2). This apparent difference probably arises because of the difficulties in sizing very large restriction fragments on agarose gels. The figure of 13.1 KB is probably more accurate because the gel from which this value is derived was electrophoresed for longer than that used in the heterologous hybridisation experiment. In addition, this value correlates well with estimates of the size (ca. 15 KB) of the larger BamHI insert in the clone pbB2 (section 4.2.2.3).

5) The autoradiograph shows that a number of mt DNA fragments hybridise weakly to pbE4 (bands arrowed in Fig. 5.10b). These cannot be explained in terms of partial restriction digestion products nor in terms of the restriction map. This weak hybridisation parallels that

detected with COII-specific probes (section 5.2.2.3) and indicates that sequence 'microheterogeneity' may be a general phenomenon occurring throughout the N, C, S and T mitochondrial genomes. Possible explanations are discussed in Section 5.4.4.

In order to investigate mt DNA organisation near the COI gene, B37N and T mt DNA, digested with EcoRI, HindIII and with BamHI was probed, under nonstringent conditions, with the yeast COI (oxi3) clone pKL-D (Hensgens et al, 1983a) (P. Isaac, pers. commun.). The hybridisation data (not shown) indicates that whereas a 9.2 KB EcoRI fragment of N mt DNA hybridises to the probe, in T a 7.2 KB fragment is identified (Fig. 5.11). More extensive mapping data suggests that this alteration occurs between 3.7 and 4.7 KB of the 3' terminus of COI and is downstream from the gene. The organisation of sequences in this region of S and C mt DNA has not been investigated.

The nature and extent of the genome rearrangements occurring downstream from the COB and COI genes in various male sterile maize lines has not yet been ascertained. However, the existence of such rearrangements in the vicinity of all the three mitochondria genes which have been studied shows that their occurrence is widespread, even if their significance is unknown. Further study of mt DNA organisation in normal (fertile) and male sterile maize lines will therefore be of great relevance in the elucidation of the molecular basis of maize mt genome evolution, and possibly also of the male sterile phenotype.

5.4 DISCUSSION

5.4.1 Definition of aims

Comparison of the restriction enzyme fragment profiles of even

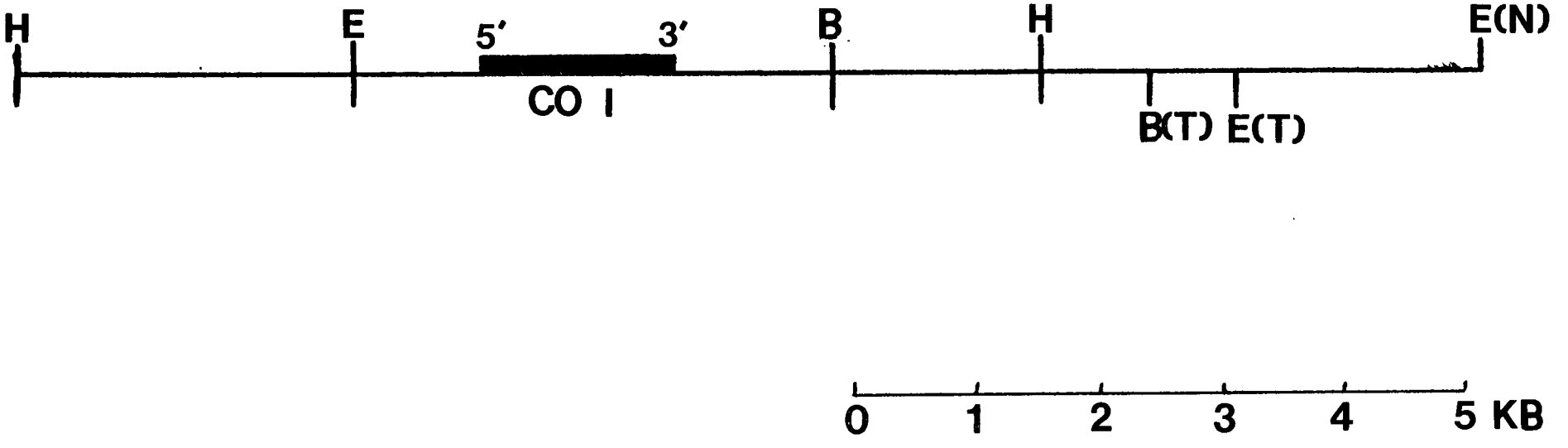


Fig 5.11 Preliminary restriction map of mt DNA flanking COI in N and T maize (from P. Isaac, pers. commun.)

closely related higher plant species has shown considerable organisational differences distinguish their mt DNAs (Sederoff et al, 1981; Weissinger et al, 1983; Stern et al, 1983). In Zea mays, four major groups (N, C, S and T) of mt DNA can be distinguished by their restriction endonuclease cleavage patterns (Pring and Levings, 1978; Levings et al, 1979). These correlate with the cytoplasmic male sterile phenotype, plants carrying the N cytoplasm being fertile, whereas those carrying C, S and T are male sterile. In addition, minor heterogeneity within each group has been demonstrated (Levings and Pring, 1977; Levings et al, 1979; Pring et al, 1980; McNay et al, 1983) and even in a single mt DNA isolate, 'microheterogeneity' and submolar sequences can be detected in DNA hybridisation experiments (Spruill et al, 1980; Dale, 1981).

The aim of the work described in this Chapter was to investigate in greater detail some of these organisational differences, using the altered mt DNAs found in cms maize lines as examples of the type of variation which are apparently widespread throughout most higher plant mitochondrial genomes. The results of these investigations, presented in sections 5.2 and 5.3,are discussed below.

5.4.2 Mitochondrial DNA rearrangements near structural genes.

5.4.2.1 The extent of mt DNA rearrangement.

Hybridisation of DNA clones containing the mitochondrial COII, COB and COI genes to blots of restriction enzyme digested N, C, S and T mt DNA has shown (Figs. 5.1, 5.10, 5.11) that near all three genes, mt DNA rearrangements (as compared to N mt DNA), occur in at least one of the male sterile (C, S or T) lines. These arrangements are summarised in Table 5.1, and show that sequence rearrangements occur abundantly throughout the different maize mitochondrial genomes. On this basis, it seems probable that if mapping studies were extended

Table 5.1 Rearrangements near maize mitochondrial genes.

	GENE			
	COB	COII	COII	COI
Cytoplasm in which rearrangement occurs	S	C	T	T
Nature of rearrangement	?	?	Deletion	?
Direction of rearrangement with respect to gene	3'	5'	5'	3'
Distance from nearest gene terminus	< 3.6 KB	< 3.2 KB	672 BP	3-5 KB

sufficiently around these loci, differences could be found in all the male sterile lines.

Since cms is known to be a mitochondrially encoded trait (see Leaver and Gray, 1982; Levings, 1983b and Chapter 1), and because mt DNA reorganisation correlates with the cms phenotype, it is likely that these arrangements could be causally related to male sterility. It was not initially anticipated, however, that rearrangements in the vicinity of COI, COII and COB would be directly related to the male sterile phenotype, and indeed those rearrangements near COI and COB are probably too distant from the genes to affect either their coding or regulatory regions. It was hoped, rather, that the results obtained from these studies would allow a model for more general mitochondrial rearrangements, including those leading to cms, to be constructed. Somewhat unexpectedly, however, it appears that the rearrangement 5' to COII discussed in section 5.4.3 could be of significance in cms.

5.4.2.2 The propagation of mt DNA rearrangement.

Although it appears that maize mt DNA can rearrange with rapidity, the mechanism by which these rearrangements take place is not clear. For example, it is unknown how the mt DNAs of the male sterile maize lines evolved from that of the presumed ancestral male fertile progenitor (Sederoff et al, 1981; Weissinger et al, 1983; Kemble et al, 1983). The absence of naturally occurring mt DNAs displaying restriction profiles intermediate between those of male fertile and male sterile lines suggests that the numerous rearrangements could have occurred simultaneously. Novel restriction digest profiles resembling those of neither male fertile nor male sterile lines, can be identified in the mt DNAs of maize plants regenerated from cells grown in tissue culture (Gengenbach et al, 1981). It would therefore appear that

massive genome reassortments can indeed occur in only a small number of mitochondrial generations. The means by which a mutation, presumably originating in a single mt DNA molecule, is transmitted to every mitochondrion in a cell, in the absence of any obvious selection advantage, is unknown. Hauswirth and Laipis (1982) postulate (in mammals) that specific zones of the oocyte cytoplasm (which is the mitochondrial donor) may be assigned to different regions of the future embryo. Thus, a single mitochondrion may provide the ancestral genotype from which all the mitochondria in, say, the germ line, are derived. The descendants of a single mitochondrion could thus spread to the next generation. It is not known if this mechanism might also operate in higher plants, but in any case, the model does not explain how an alteration in a single copy of the mitochondrial genome, amongst the several present in each organelle (Bogenhagen and Clayton, 1974; Borst and Kroon, 1969) could be duplicated and transmitted to each of the multiple copies.

5.4.2.3 Implications of mitochondrial genome reorganisation to gene order.

It seems unlikely that the rearrangements detected near COII, COB and COI in various male sterile maize lines will be sufficiently extensive to alter the gross order of the genes. However, until complete restriction maps of N, C, S and T mt DNAs are available, this premise will remain unproven. Although rearrangements in the mitochondrial genomes of Tetrahymena (Goldbach et al, 1978), Paramoecium (Cummings et al, 1980), Aspergillus (Earl et al, 1981) and Neurospora (Collins and Lambowitz, 1983) species have been documented, none of these changes alters the gene order. It would be interesting to investigate whether similar constraints are placed on mitochondrial genome arrangement in Zea, both within the N, C, S and T lines of

Z.mays and between closely related Zea species. This would be relatively easy to test for 'closely' linked (i.e. <50 KB apart) genes such as COB and COI because restriction maps linking the two genes in the various mt DNAs could be constructed, allowing any alterations to be detected. It would be more difficult to construct a similar map for COII, however, as this locus, in N, is several hundred KB from the other mapped genes (D. Lonsdale, pers. commun., see also section 4.5.3).

5.4.3 The deletion near COII in T-type maize mt DNA

5.4.3.1 Extent and nature of the deletion.

Restriction mapping experiments showed that a DNA rearrangement occurs less than 1 KB from the presumed 5' terminus of the COII gene in T mt DNA. Sequence analysis of N and T mt DNA (Fig. 5.8) confirmed that this rearrangement occurs precisely 627 nucleotides upstream from the AUG codon suggested by Fox and Leaver (1981) to represent the start of the maize COII gene. Hybridisation experiments indicated that the rearrangement in T mt DNA involves the deletion of DNA sequences as compared to N. It is tempting to speculate that the lesion in T is a simple deletion, occurring at the point from which the N and T DNA sequences diverge. This is not, however, proven and the alteration could involve a more complex rearrangement of sequences. It is therefore important that the restriction map of N and T mt DNA presented in Fig. 5.5 be extended further upstream in order that the extent and nature of the rearrangement can be assessed.

The preliminary sequence data showed that as far as the point of rearrangement, the 518 BP of sequence which was examined was absolutely identical in N and T (Fig. 5.8). This suggests either that extremely strong selection pressures act on this sequence, requiring its preservation, or that the N and T cytoplasms diverged very recently

on an evolutionary timescale. The T type of male sterility was first reported in Texas as recently as 1944 (Rogers and Edwardson, 1952), so the second explanation is perhaps more plausible. It is nevertheless interesting that certain mt DNA sequences are exactly preserved in the male sterile lines, whereas adjacent sequences appear to be completely dispensible.

The sequence analysis near the point of rearrangement also showed that the N and T sequences do not diverge immediately, rather the deletion is 'staggered' (Fig. 5.8). The sequences are identical up to a point 627 nucleotides from the 5' end of COII, and thereafter appear different. Approximately 20 BP 5' to the rearrangement point in T, is a 25 nucleotide sequence with seemingly exact homology to a similar stretch of N mt DNA terminating 27 BP upstream from the rearrangement. It must be emphasised that the sequences in this region could not be determined with complete accuracy because they lie more than 350 BP from the point at which the sequencing reaction was initiated (see section 5.2.3.2). The presence and extent of the 'stagger' can however be deduced simply by comparison of the patterns in the relevant N and T sequencing gels (Fig. 5.9). A search for similar sequences and 'staggers' near rearrangements elsewhere in the genome could give a valuable insight into the mechanisms by which genome reorganisation occurs in maize mt DNA. It would be of particular interest to determine the point at which the rearrangement upstream from COII occurs in C mt DNA. By this means, a recombination 'hot spot', near which rearrangements occur in both C and T mt DNA, might be identified.

5.4.3.2 Effects of the deletion on an Open Reading Frame upstream from COII

The preliminary sequence of N and T mt DNA as far as the

rearrangement, contains an open reading frame continuous with the COII gene. If the published COII upstream sequence is correct (Fox and Leaver, 1981), then no termination codons separate this ORF from that encoding cytochrome oxidase subunit II. This ORF, between the N-T divergence point and the AUG codon thought to represent the start of the COII gene (Fox and Leaver, 1981; Heisel and Brennicke, 1983), could encode a polypeptide of 209 amino acid residues. In purely random DNA, the chance of a sequence of this length containing no termination (TAA or TAG) codons is $[(64 - 2)/64]^{209}$, i.e. about 0.13%. If TGA is included as a termination codon, the probability is reduced to 0.004%. Clearly, therefore, this ORF must have some significance in maize. The finding that the ORF appears to extend into the region altered in T mt DNA also raises questions about the role of the putative polypeptide it encodes in cytoplasmic male sterility.

If the ORF is transcribed and cotranslated with COII, then a 'hybrid' polypeptide of high molecular weight would be produced. It seems extremely unlikely that the mature cytochrome oxidase subunit II in higher plants contains a long amino-terminal extension, especially since antibodies raised against yeast subunit II immunoprecipitate a maize mitochondrial translation product of M_r similar to that of COII in yeast (C.J. Leaver, pers. commun.). Four possible roles could be ascribed to this upstream COII ORF in N maize mitochondria:

1) It is not expressed (i.e. as a pseudogene) and therefore has no role at all. This seems rather unlikely, particularly since some mitochondrial transcripts hybridising to COII-specific DNA probes are large enough (6.4 Kbases; V. Jones, pers. commun.) to accommodate the entire COII coding sequence plus that of the ORF. Use of ORF-specific DNA probes may help to clarify the transcriptional processing

pattern of this region in both N and T mitochondria. Furthermore, the codon usage in the sequenced region of the ORF is similar to that in other maize mitochondrial genes. The distribution of nucleotides in the third codon position (T = 29.5%, A = 26.2%, G = 23.8%, C = 20.5%) is similar to that determined in COB, COI and COII, with a rather less marked preference for T in the third position.

2) The ORF may be co-translated with COII and form an amino-terminal extension which is subsequently processed from the mature polypeptide. The yeast (Sevarino and Poyton, 1980) and Neurospora crassa (van den Boogart et al, 1982) COII polypeptides are synthesised as precursors with amino terminal extensions of 12-15 amino acids. Similarly, COI in Neurospora crassa (Burger et al, 1983; de Jonge and de Vries, 1983) and Aspergillus nidulans (T. Brown, pers. commun.) is thought to be translated as a precursor with an amino terminal extension of about 50 residues. However, the existence of a mitochondrial protein with a predicted pre-piece containing in excess of 200 amino acids, has never previously been reported. It would be difficult to envisage a function for such a long leader peptide. Sequence analysis of the mature maize COII polypeptide may assist in determination of the processing pattern of this protein. If translation of COII really begins several hundred nucleotides upstream from the previously assumed start of the gene, then the absence of a good ribosome binding sequence (Chapter 4) near this point might be explained.

3) The ORF could represent the gene for an important structural protein which is co-translated with COII and subsequently proteolytically cleaved to yield two separate polypeptides. Using cloned DNA sequences specific to the ORF, it may be possible, with the use of a bacterial expression vector, to generate sufficient ORF-derived protein to allow antibodies to be raised against it. In this way, maize

mitochondrial polypeptides could be probed, either by immunoprecipitation or 'Western' blotting (Gershoni and Palade, 1983) and the protein, if any, encoded by the ORF identified. The predicted amino acid sequence of this polypeptide does not appear to have homology with any of the known human mitochondrial proteins (Anderson et al, 1981) or predicted URF sequences. A hydropathy plot of the 209 residue portion of the ORF determined here is presented in Fig. 5.12.

4) The ORF could specify a maturase enzyme, cotranslated with the first exon of COII, which is involved in splicing of the intron sequence (Fox and Leaver, 1981) from the mature COII mRNA. It is perhaps significant in this context, that the Oenothera COII gene (Heisel and Brennicke, 1983), which does not contain an intron, is preceded by several in-frame termination codons, a putative ribosome binding site, and no sequence with homology to the maize ORF. It is not known if these ORF sequences occur elsewhere in the Oenothera mitochondrial genome. Hybrid exon-intron open reading frames are thought to be important in the splicing of certain yeast mitochondrial mRNAs (Jacq et al, 1982), so this type of phenomenon is not unprecedented. However, in yeast, these maturase ORFs normally reside within the intron whose splicing they catalyse. In addition, they are normally continuous with an exon on the 5', rather than the 3', side.

The existence of a long upstream ORF continuous with COII is in itself interesting. The fact that this ORF presumably starts in a region of mt DNA which is apparently altered in different maize lines is even more surprising. It has not been possible to extend sequence analysis far into the respective N and T sequences, so it is not possible to determine whether the alteration could lead to a truncation of the coding sequence. It is tempting to speculate that the

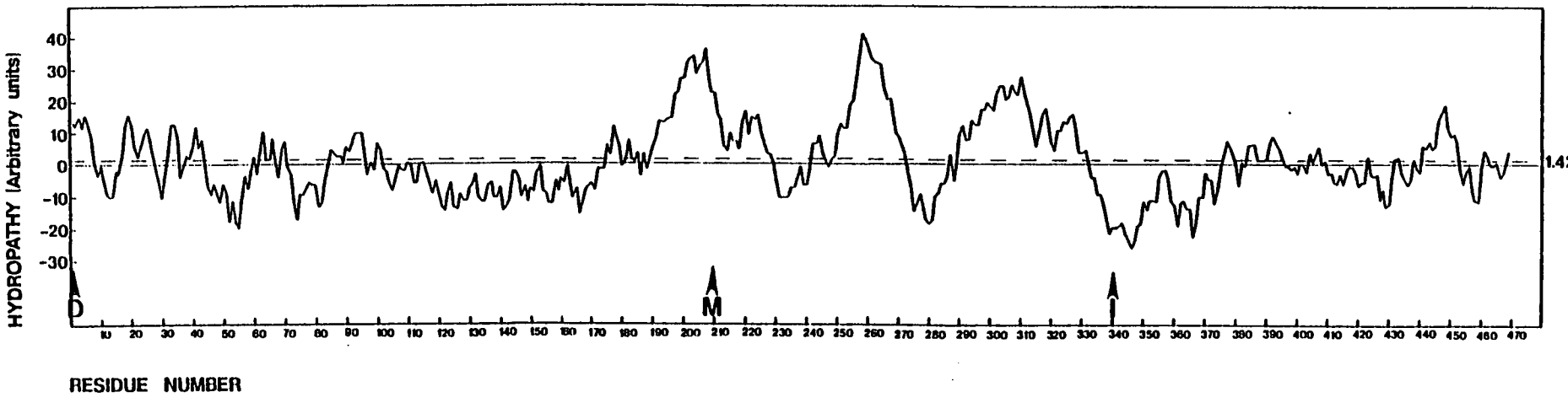


Fig 5.12 Hydropathy profile of Normal (B37N) maize cytochrome oxidase subunit II and predicted ORF-encoded pre-piece (using the indices of Kyte and Doolittle, 1982, and eleven amino acid averages).

Vertical arrows indicate: the points at which the predicted I sequence diverges from N (D); the presumed initiation codon of COII (M) and the position of the intron (I).

rearrangement may lead to the formation of an altered polypeptide, perhaps related to the 13,000 M_r protein specific to T mitochondria. This might be investigated using antibodies r aised against the ORF encoded protein, produced as described above.

It seems possible that if the ORF and COII are co-expressed, then alteration of sequences within the ORF, or its regulatory regions, may also affect expression of COII. This would be particularly true if the ORF-maturase model is correct, since there is no evidence that the COII gene in T mt DNA does not contain an intron. However, preliminary immunoprecipitation data (J. Bailey-Serres, pers. commun.) suggests that COII is apparently not altered in T type mitochondria, hence the role of the ORF in the expression of COII is questionable. Alternatively it is possible that the rearrangment in T mt DNA preserves the reading frame, thereby allowing the synthesis of a functional polypeptide.

If the rearrangement in T is a simple deletion, and if this deleted sequence was shown by sequence analysis to contain an ORF (or ORFs), then a striking parallel with the 'optional' introns in yeast mt DNA might be established (Sanders et al, 1977; Grivell, 1983). These maize mitochondrial genome rearrangements might also mimic the type of reorganisation caused by bacterial insertion sequences or maize nuclear controlling elements (Fincham and Sastry, 1974).

5.4.4 Microheterogeneity in maize mitochondrial DNA

Hybridisation of COB and COII probes to blots of maize mt DNA from either N, C, S or T cytoplasms (Figs. 5.1 and 5.10) always reveals the presence of fragments which hybridise with low intensity and which cannot be rationalised in terms of the predicted restriction map. This sequence 'microheterogeneity' has also been reported in maize by Lonsdale et al(1983) and in a variety of other higher plants, for

example wheat (Falconet et al, 1983) and Brassica campestris (Palmer and Shields, 1983). A number of possible explanations of this phenomenon were presented in section 5.2.2.3. It seems likely that microheterogeneity arises from intra-molecular recombination within the large circular DNA molecule containing the entire mitochondrial genome (see Chapter 1), yielding a variety of lower abundance, smaller circles containing altered sequence organisation in the region of the recombination.

Lonsdale (pers. commun.) has shown by restriction enzyme analysis of cosmid clones containing the maize COII gene and flanking sequences, that this locus may reside near a region of high frequency recombination. These recombinational events would generate novel sequence combinations 5' to COII and may account for some of the faintly hybridising bands. The low intensity of hybridisation to these fragments shows that these events probably only occur, at most, infrequently. The short (5-7 nucleotide) direct repeated sequences occurring just upstream from COII (Fig. 5.8), may in some way be involved in this genome 'instability'.

The detection of faintly hybridising bands with COB probes suggests that microheterogeneity may occur throughout the mitochondrial genome and is not confined to a single 'hot spot' near COII. These changes may be indicative of the generally rather fluid nature of the maize mitochondrial genome as discussed in section 5.1. It is possible that the altered restriction enzyme digestion patterns observed in mt DNAs from cms maize could result from similar recombinational events which have somehow become stabilised and amplified throughout the mitochondrial population. It is also possible that the 'micro-heterogeneity' may simply reflect partial duplication of sequences contained in the cloned probes, in other regions of the genome. For example, if the ORF upstream from COII were involved in splicing,

similar sequences may be found elsewhere, adjacent to other split genes.

5.4.5 Overview and summary

The aim of this work, as outlined in section 5.1, was to investigate genome rearrangements in maize mt DNA, in order to assess the generality of the phenomenon, and its involvement, if any, in cytoplasmic male sterility.

In sections 5.2 and 5.3, DNA rearrangements near a variety of mitochondrial genes in male sterile maize lines were identified. In addition 'microheterogeneity' of sequences near these loci within a single maize line, was demonstrated. These results imply that maize mt DNA may recombine and rearrange at a rapid rate, and that some of these rearrangements may be involved in cms.

DNA hybridisation and sequence analysis has allowed a deletion occurring 5' to the COII gene to be characterised, and has indicated the kinds of feature which might be common to all the frequently-occurring mt DNA rearrangements. In addition, this sequence analysis rather unexpectedly revealed the presence of a long ORF continuous with COII, which is probably altered by the T rearrangement. The significance of this sequence in the expression of COII in N and T mitochondria is unknown at present and represents an interesting problem for future investigation.

The studies described in this Chapter are only preliminary, but they suggest many fruitful lines of research which could be followed up. These investigations should significantly enchance our understanding of the structure and function of the higher plant mitochondrial genome.

CHAPTER 6 DISCUSSION

6.1 SUMMARY AND AIMS OF THIS CHAPTER

The identification, cloning and characterisation of the maize mitochondrial COB gene, and analysis of sequences 5' to the COII gene, have been reported in this thesis. These two lines of research have broadened our understanding of the maize mitochondrial genome in the following ways:

1) Sequence analysis of the COB gene confirmed that maize mitochondria use a genetic code which differs from that employed in fungal and animal mitochondria.

2) Membrane-spanning domains in the COB-encoded polypeptide (apocytochrome b) were predicted, suggesting that it could function as in animal and fungal mitochondria.

3) Analysis of mt RNA showed that the COB gene is transcribed in a complex manner, and a unifying model which could explain all the observed patterns of transcription was proposed (see also section 6.2.1).

4) An octanucleotide sequence, which might act as a ribosome binding site, was identified immediately preceding the initiation codons of several plant mitochondrial genes.

5) The 627 BP of sequence preceding the COII gene contains an open reading frame continuous with the first exon of this gene. At 627 BP 5' to COII, a rearrangement occurs in mt DNA from cms-T plants as compared to that from normal (fertile) plants. This rearrangement is likely to affect expression of the ORF, and possibly also of COII, in these plants.

6) In collaboration with P. Isaac and D. Lonsdale, COB, COI and COII have been located on a maize mt DNA restriction map, allowing a genetic linkage map of higher plant mt DNA to be constructed for the first time.

These findings have been discussed in Chapters 4 and 5. The purpose of this Chapter is not to reiterate these discussions but to consider plant mitochondrial genes in a broader context, and to touch on some more speculative aspects of mitochondrial genome organisation and expression.

6.2 CONTROL OF GENE EXPRESSION IN MAIZE MITOCHONDRIA

6.2.1 Transcription: Processing and splicing

Transcript analysis of maize mitochondrial genes by 'Northern' blotting and DNA:RNA hybridisation has shown that, at least in five day old coleoptiles, several different patterns of transcription can be identified (Chapter 4, also V. Jones, pers. commun.). COI-specific DNA probes identify a single major transcript, and up to two additional minor transcripts. COB, which like COI does not contain introns, is transcribed in a more complex manner and at least five different RNA species hybridise to COB-specific probes. COII also displays a complex transcriptional pattern, probably due in part to the presence of an intron in the gene. The various transcripts of COB, COI and COII in maize are summarised in Table 6.1.

In Chapter 4, a model to account for these patterns, involving transcription between fixed promoter and terminator sequences followed by differential processing, was proposed. Recent evidence from studies of S.cerevisiae mitochondria has shown that RNA may be processed by similar pathways in this organism. For example, the yeast COII gene does not contain introns and is transcribed as a single RNA species of 3.6 Kbases (Thalenfeld et al, 1983). On the other hand, at least ten transcripts derived from the COII gene, which does not contain an intron either, can be identified (Coruzzi et al, 1981). Similarly, transcription of COB is complex, probably due to the presence of two or five introns in the gene (Christianson et al, 1983).

Table 6.1. Transcripts of COB, COI and COII in Z.mays and S.cerevisiae mitochondria.

	Gene		
	COB	COI	COII
Z. mays	> 9, 4.3, 2.2	2.4, 2.3	6, 5.1, 3.9, 3.5, 3.2, 2.45, 1.95
S.cerevisiae	(Up to ca. 10 transcripts) 1.9	(Up to ca. 50 transcripts) 2.1	(Up to ca. 10 transcripts) 0.85

Transcript sizes in k bases

Major transcripts underlined.

It is clear that two processes contribute to the overall complexity of the transcription patterns of these genes. Firstly, the presence of introns automatically means that at least two RNA species will contain the coding information of the gene: the primary transcript containing intron sequences and the spliced mature mRNA. Secondly, processing of 5' and 3' untranslated sequences from the primary transcript will yield a series of RNA intermediates of varying stabilities. Thus, the extent (and rate) of both splicing and processing will govern the ultimate complexity of the transcription pattern which is generated.

In yeast, almost all stable transcripts appear to have undergone some 5' and 3' processing. Even the single transcript of the COIII gene does not represent the primary transcript: this molecule is probably processed so rapidly that it cannot be detected in hybridisation experiments (Thalenfeld et al, 1983). The COII gene is probably transcribed between single promoter and terminators, however in this case the primary transcript is more stable and is slowly processed in a sequential fashion. Processing of COB (Christianson et al, 1983), S and L rRNA (Christianson et al, 1982; Merten et al, 1980) transcripts in a similar manner has also been demonstrated. It is perhaps surprising that such radically different processing pathways can exist for genes contained in the same mitochondrion, all probably transcribed by the same RNA polymerase (Christianson et al, 1983). Co-transcription of two (or more) genes as a polycistronic RNA could account for some of the processing requirements. However, the reason why processing is required at all for non-cotranscribed genes, is not really known.

All the observed patterns of transcription in maize mitochondria have parallels in yeast, and these patterns are probably generated via the single promoter-terminator model. It therefore seems plausible

that a similar model could operate in higher plants. In order to prove that it does, it will be necessary to identify the requisite promoter sequences, analogous to the nonanucleotide $^{5'}$ ATATAAGTA which is believed to act as a promoter in S.cerevisiae (Osinga and Tabak, 1982; Christianson et al, 1983) and K.lactis (Osinga et al, 1982). If such a sequence is found preceding some, but not all, higher plant mt genes, it will be necessary to prove that the genes lacking it form part of a polycistronic primary transcript. Such experiments may not be easy to perform because the primary transcript may only have a transient existence and therefore not be amenable to analysis by standard techniques such as S1 nuclease mapping. Alternative approaches, such as in vitro 'capping' of unprocessed transcripts (Levens et al, 1981) may have to be adopted to identify these initiation sequences.

The reasons why processing (as opposed to splicing) occurs, are not well understood. However, it seems likely that the operation is largely catalysed by enzymes encoded in the nuclear genome. Thus, assuming that unprocessed transcripts do not have messenger activity, it seems probable that nuclear control of mitochondrial mRNA maturation may allow control over the expression of specific genes. Accumulation of precursor mRNAs encoded by these genes may allow for the rapid expression of these genes without requiring de novo RNA synthesis. The rapid synthesis of six polypeptides in germinating V.faba seeds has been demonstrated by Dixon et al (1980) and this seems to be achieved independently of a general increase in mitochondrial protein synthetic capability. Rapid maturation of 'stockpiled' mRNAs specifying these polypeptides could provide the means to achieve this synthesis without the concomitant production of polypeptides not immediately required during germination.

6.2.2 Intron splicing in maize mitochondria

In yeast mitochondria, a substantial proportion (up to ca. 15 KB) of the genome (ca. 75 KB) is composed of intron sequences in the COB, COI and L rRNA genes. Intron sequences thus account for a large portion of the increased mt genome size when compared to those of the animals. Since the mt DNAs of higher plants are even larger than those of the fungi, it would not be unexpected if, by analogy, their genes contained intron sequences. However, in maize, of the sequenced genes, only COII contains an intron. The presence of an intron in this gene correlates with the presence of a long open reading continuous with, and preceding, the first exon. By analogy with the hybrid intron-exon maturases in yeast, it has been suggested (section 5.4.3.2) that the ORF could be involved in splicing the intron sequences from the precursor mRNA. The sequence of events could then be as follows:

1) The ORF and first COII exon are co translated as far as the first termination codon in the intron (33 triplets beyond the proposed splice point).

2) The hybrid protein functions as a 'maturase', joining (possibly in collaboration with nucleus-encoded enzymes) the ORF and COII exons to form a mature mRNA.

3) A second hybrid protein, containing the ORF-encoded polypeptide fused to the amino terminus of cytochrome oxidase subunit 2, is translated from the mature mRNA.

4) The amino-terminal extension, containing the ORF-encoded sequences, is removed proteolytically, yielding a mature subunit 2 protein and a free ORF-encoded protein. The latter may have some independent function in the mitochondrion, or could be degraded rapidly. Proteolytic removal of a (short) amino-terminal extension peptide from subunit 2

occurs in both yeast (Sevarino and Poyton, 1980) and N.crassa (van den Boogaart et al, 1982). A minor modification of a similar system in maize could allow the proteolysis of a longer pre-piece. The COII gene in Oenothera (Heisel and Brennicke, 1983) does not contain an intron and is not preceded by an ORF. Thus, the correlation between the presence of the intron and the ORF is positive, at least in the two species analysed. It remains to be established whether COII contains an intron in any other higher plant, and if so whether the gene is preceded by an ORF. It will also be interesting to investigate whether other split genes in maize mt DNA - if any - could be spliced in a similar manner.

The mt DNA rearrangement 627 BP 5' to COII in cms-T mitochondria might alter the proposed maturase so that the variant protein is less active than its counterpart in fertile plants. This may not be important during normal plant growth, but may prevent the mitochondrion from synthesising sufficient cytochrome oxidase subunit 2 during the period of rapid mitochondrial biogenesis prior to pollen formation. This could lead to the mitochondrial degeneration and pollen abortion observed in cms-T plants (Warmke and Lee, 1977). Clearly, this model does not adequately explain all aspects of the cms-T phenomenon, for example T-toxin sensitivity and the synthesis of variant polypeptides in cms-T mitochondria. However, it does suggest that splicing could be important in the normal expression of mitochondrial genes, and that at least some forms of cms could arise from defects in the splicing (or processing) pathways. In this context, the function of the nuclear fertility-restoring (Rf) genes might be to suppress these hypothetical splicing defects, paralleling the activity of the nuclear nam2 gene, which can suppress intron-encoded maturase deficiencies (Dujardin et al, 1982) in yeast mitochondria.

6.2.3 Translation and translation products

The recent sequence analyses of several plant mitochondrial genes (see Chapter 4) has, for the first time, allowed meaningful comparisons of sequences 5' to these genes to be made. An octanucleotide, having five out of eight, or better, homology to a concensus sequence 5'AGGATTCA, is found upstream from the maize COB and COI genes, starting 23-28 BP from the ATG initiation codon. A similar sequence, in the same relative position, also precedes the Oenothera COII gene (Hiesel and Brennicke, 1983) and probably also the Sorghum COI (D. Hanson, pers. commun.) and maize URF-1 (T. Brown, pers. commun.) genes. Because the true start of the maize COII gene has not been determined with certainty, it is not known whether a similar concensus sequence also precedes this gene.

It has been proposed (section 4.4) that this octanucleotide acts as a ribosome binding site in a manner analogous to the bacterial Shine-Dalgarno sequence (Shine and Dalgarno, 1974) because a complementary octanucleotide (5'UGAAUCCU) occurs near the 3' end of the maize and wheat mitochondrial 18S rRNAs. If this hypothesis is correct, then at least in one aspect, translation in higher plant mitochondria is more similar to that in the prokaryotes than is that in the mitochondria of any other eukaryote. However, the apparent inability to translate maize mt mRNAs in an E.coli cell free system (Leaver and Gray, 1982) shows that translation in E.coli and maize mitochondria may have some fundamental differences.

The lack of an in vitro system capable of transcribing and translating mt DNA has hampered identification of the genes encoding the various polypeptides synthesised by isolated plant mitochondria (Leaver et al, 1983). However, as DNA sequence data becomes available, it may be possible to use this information to identify the translation

products encoded by a particular piece of DNA. The use of bacterial vectors to express mitochondrial genes in E.coli has been discussed in sections 4.6.2.2 and 5.4.3.2. In this way, sufficient protein might be synthesised to allow specific antibodies against mitochondrial polypeptides to be raised. Alternatively, antibodies might be raised against short synthetic oligopeptides constructed with knowledge of the mt DNA gene sequence. These antibodies could then be used to identify the full length translation products, synthesised in isolated mitochondria. Such an approach has been successfully used to identify the translation products encoded by the human URF A6L, ATPase 6, URF 1 and URF 3 genes (Mariottini et al, 1983; Chomyn et al, 1983).

Even if the amino acid sequence of a mt translation product is known or predicted, it is much more difficult to assign an in vivo function to it. For example, of the eight unidentified reading frames located by sequence analysis of the human mt genome, the function of only one of them (URF A6L = ATPase 8) has been determined. This assignation was only made possible by comparison with S.cerevisiae gene sequences (Macreadie et al, 1983) identified by analysis of mit$^-$ mutants. In the absence of analogous mutations in plant mt DNA, analysis of the function of the 20-50 translation products they encode, could be an extremely lengthy procedure.

6.3 UNORTHODOX FEATURES OF THE HIGHER PLANT MITOCHONDRIAL GENOME

6.3.1 Coding capacity

A variety of experiments (e.g. Ward et al, 1981; see also Chapter 1) have shown that higher plant mt DNA is large (200-2400 KB) and that this complexity varies considerably even between closely related species. This raises two questions which are addressed below:

1) Why is the mt genome so large?

The obvious answer to this question is that higher plant mitochondria encode more genes than do their fungal and mammalian counterparts. However, analysis of the proteins synthesised by isolated plant mitochondria shows that this is not so: at most 30 additional polypeptides are made (e.g. the α subunit of the F_1-ATPase; Hack and Leaver, 1983), only requiring a small proportion of the total potential coding capacity. An alternative explanation, that the bulk of the genome is composed of repeated sequences, can be discounted on the basis of kinetic (Ward et al, 1981) and hybridisation (Spruill et al, 1981) analyses which show that <5% of the genome is reiterated.

Yeast mt DNAs show variations in size between 18.9 and 108 KB (Clark-Walker et al, 1981). However, at least in S.cerevisiae, much of the 'extra' sequence compared to mammalian mt DNAs is occupied by intergene A + T-rich 'spacer' DNA and by long intervening sequences within genes. In higher plants, the first of these explanations is not tenable because the A + T content in most species is relatively low (ca. 53% in most species: Wells and Ingle, 1970), and the second is made more unlikely by the finding that the majority of higher plant mt genes sequenced to date do not contain introns.

It appears, therefore, that higher plant mt DNAs present a 'C-value paradox' (c.f. Cavalier-Smith, 1978) since the amount of DNA they contain apparently considerably exceeds the amount required for coding purposes. The function of this additional DNA is not known, nor is it understood what evolutionary pressures have caused animal mt DNAs to become streamlined to a minimum, or yeast mitochondria to maximise their A + T contents, while higher plants apparently exhibit little constraint on the amount of DNA their mitochondria contain.

Cavalier-Smith (1978) and Orgel and Crick (1980) note that an increase in the amount of nuclear DNA often correlates with a lengthened

cell cycle and increased longevity of the species. It is not known whether similar selection forces may act on the mitochondria, although it seems unlikely because some of the longest-lived eukaryotes (e.g. man) actually have the smallest mt genomes. It is possible that the excess DNA is without function, i.e. is 'selfish' and only exists because suitable means to eliminate it have not evolved (Doolittle and Sapienza, 1980; Orgel and Crick, 1980).

2) Why are mt genomes so variable in size?

The mt DNA complexities of two closely related curcurbits, watermelon and muskmelon, are 330 and 2400 KB respectively. It seems higher improbable that the mt DNA coding requirements are nearly ten times as great in one species as in the other. The excess sequence is therefore presumably noncoding and could have two possible origins:

a) Mt DNA sequences could have been duplicated, subsequent degeneration of one copy leading to an increased apparent mt DNA complexity. Although this is in principle possible it seems unlikely that sufficient duplication and degeneration could have occurred in the presumably short timespan since, say muskmelons diverged from watermelons, unless an improbably high muation rate is invoked.

b) It could arise by the wholesale transfer of DNA sequences from another cellular compartment - e.g. the nucleus or chloroplast - into the mitochondrion. However, while transfer of DNA from chloroplast to mitochondrion has been demonstrated (section 6.3.3) the amounts involved are generally rather small (Stern et al. 1983) and not correlated with total mt genome size. Furthermore, the presence of nuclear DNA sequences in plant mitochondria has not been proven.

At present, therefore, the origin of the (apparently) excess DNA in the higher plant mt genome, and the reasons why the amount varies between different species, are not known. It is possible that further sequence analyses may reveal something of the function of this DNA, but is is equally possible that it may have no function at all.

6.3.2 Mitochondrial genes and the endosymbiont hypothesis

The 'endosymbiont' hypothesis for the origin of plastids and mitochondria has been widely promulgated in recent years. Persuasive evidence has accumulated and shows beyond reasonable doubt that plastids evolved from primitive photosynthetic bacteria (see Gray and Doolittle, 1982 for review). A bacterial ancestry of mitochondria also seems probable, although in this case the data is much less convincing (Gray, 1983; Gray and Doolittle, 1982). Plant mitochondria are perhaps the most 'primitive' because their S rRNA sequences most closely resemble those of modern day bacteria (Spencer et al, 1983). However, it should be remembered that in many respects (e.g. the presence of introns, complex patterns of transcript processing and an unusual organisation of rRNA genes) plant mitochondria are decidedly un-bacterial, and any endosymbiotic model must account for these differences.

Since the presumed free-living progenitors of present-day mitochondria must have contained more genes than all present-day mt DNAs, a transfer of coding sequences from endosymbiont to nucleus must have occurred. The question of why the particular set of genes found in almost all mt DNAs has been retained, must therefore be answered. In particular, why is mt DNA retained at all, since its maintenance and expression undeniably represent a substantial drain on the cell's metabolism? Two possible explanations have been offered:

1) The retained genes specify hydrophobic components of the inner

mitochondrial membrane which could not be efficiently imported from the cytoplasm (e.g. Borst and Grivell, 1978). However, the finding that in Neurospora the very hydrophobic F_o-ATPase subunit 9 is encoded in the nucleus (Sebald, 1977), and that efficient mechanisms exist for its import into mitochondria (Viebrock et al, 1982) tends to reduce the validity of this argument. At the other extreme, the hydrophilic α subunit of the F_1-ATPase can be encoded in the mitochondrion in some eukaryotes (Hack and Leaver, 1983).

2) DNA transfer might have been arrested suddenly, leaving no mechanism for the elimination of the few remaining mitochondrial genes. However, the finding that recent DNA transfers to the nucleus have occurred (see below) suggests that this theory might also be incorrect.

There is, therefore, no obvious reason why mitochondria contain DNA. It may simply be that once efficient means have evolved for their expression, no particular selection advantage is conferred by the transfer of a single one of these genes, when the rest remain and have to be expressed anyway.

6.3.3 'Promiscuous' DNA

Until recently, the genomes of the nucleus, plastids and mitochondria were assumed not to share common sequences with each other, and to be confined to separate cellular compartments. However, Stern and Lonsdale (1982) and Lonsdale et al (1983) have shown that the maize mt genome contains stretches of chloroplast DNA, including rRNA and tRNA genes, and the large subunit of RUBP carboxylase. Ellis (1982) has coined the term 'promiscuous' DNA to describe these DNA sequences which have seemingly migrated from one genome to another. Their 'function' is not known, but Stern and Lonsdale (1982) argue that since the arrangement of ct DNA sequences in the maize mt genome differs in fertile and male sterile lines, its presence might be

important for the normal expression of the mt genome. Stern et al (1983) have shown that ct DNA sequences are found in the mt DNAs of a variety of angiosperms, suggesting that its 'importance' could be universal.

Inter-organelle DNA movement has not been restricted to the chloroplast and mitochondrion. Chloroplast DNA sequences have also been found in the nucleus of spinach (Timmis and Scott, 1983), and nuclear copies of mitochondrial sequences have been reported in locusts (Gellissen et al, 1983), sea urchins (Jacobs et al, 1983), Podospora (Wright and Cummings, 1983) and yeast (Farrelly and Butow, 1983); see also Fox (1983) and Lewin (1983) for review.

In most cases, the mt DNA sequences integrated into the nuclear genome only represent partial copies, often highly rearranged, of mitochondrial genes. However, sea urchin nuclear DNA contains a complete copy of the mt COI gene. Sequence analysis has revealed that in the nuclear copy, TGA codons have been specifically converted to TGG triplets. In this case, if transcribed, the mt gene could be expressed in the nucleus, perhaps suggesting that this might represent an early stage in the transfer of a gene from cytoplasm to nucleus. A later stage in the process might be represented by the ATPase subunit 9 gene in Neurospora. Although in this fungus, this mitochondrial protein is encoded by a nuclear gene (Sebald, 1977) and synthesised on cytoplasmic ribosomes, a complete copy of the gene is also found in the mitochondrion (van den Boogaart et al, 1982a; de Vries et al, 1983). This copy, like the sea urchin nuclear COI gene, has the potential to be expressed, although this potential is probably not realised in the mitochondrion. A plausible explanation for gene transfer might therefore require duplication of a mitochondrial gene and transfer of one copy to the nucleus. The mitochondrial copy could be 'inactivated' as the nuclear copy becomes expressed and its

product becomes importable by the mitochondrion. The sea urchin and Neurospora genes might represent intermediates at different stages in the process. Presumably the transfer is slow and rather haphazard because numerous alterations will be required to allow a mitochondrial gene to be expressed in the nucleus, and the selection pressure driving these changes will be very low.

It is not known how a single DNA sequence can be dispersed among two (or more) intracellular genomes. Ellis (1982) suggests three possibilities: either organellar lysis followed by uptake of sequences into the nucleus or mitochondrion; organellar fusion and recombination of their DNAs; or vector-mediated DNA transfer. It seems likely that different mechanisms may have been used by the various 'promiscuous' sequences. However, the vector model is perhaps the most attractive, and the finding that 'transposon-like' DNA elements from maize mitochondria (i.e. the S-plasmids) can be found integrated in nuclear DNA certainly strengthens this view (Kemble et al, 1983).

6.4 CONCLUDING REMARKS

The research described in this thesis has focussed on the analysis of two genes, COB and COII, in maize mitochondria. The results of these analyses have been described in Chapters 4, 5 and 6. Whilst these investigations have answered some questions about genes and their expression in maize mitochondria, many more have been raised. Some of these questions have been discussed in this Chapter, others have been hinted at in earlier sections of the thesis. In my opinion, the major unresolved questions about the maize mitochondrial genome are as follows:

1) Why is the genome so much larger than required to encode the genes it contains, and what is the origin of this excess DNA?

2) How universal is the 'recombinational' model for genome structure, and is the order of genes on the mt DNA in other higher plants similar to that in maize?

3) Why does mt DNA rearrange so fast, and what are the molecular bases of these sequence rearrangements?

4) Which additional genes does the genome contain, not normally found in other mt DNAs?

5) What is the molecular basis of cytoplasmic male sterility and of nuclear restoration of fertility by Rf genes?

6) What is the function of the continuous open reading frame preceding COII exon 1, and does this have any role in cms?

7) What are the promoters for RNA transcription and why are splicing and processing patterns so complex?

8) How is the expression of mt genes controlled, both at a transcriptional and translation level?

9) Why do maize mitochondria contain portions of the chloroplast genome?

10) Is gene transfer from mitochondrion to nucleus still occurring, and if so by what means?

It is likely that complete answers to all these questions will never be realised. It is obvious that DNA sequence analysis alone cannot provide all the answers. However, the experiments suggested in Chapters 4, 5 and 6, perhaps combined with sequence analysis, should help to resolve some of them. In any event, it is clear that analysis of higher plant mitochondria and the genes they contain, will be an interesting and fruitful field for molecular research for several years to come.

BIBLIOGRAPHY

Albring, M., Griffith, J. and Attardi, G. (1977)
Proc. Natl. Acad. Sci. USA. 74, 1348 - 1352

Alexander, N.J., Perlman, P.S., Hanson, D.K. and Mahler, H.R. (1980)
Cell 20, 199 - 206

Ambler, R.P. and Scott, G.K. (1978)
Proc. Natl. Acad. Sci. USA. 75, 3732 - 3763

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G. (1981)
Nature 290, 457 - 465

Anderson, S., de Bruijn, M.H.L., Coulson, A.R., Eperon, I.C., Sanger, F. and Young, I.G. (1982)
J. Mol. Biol. 156, 683 - 717

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G. (1982a)
In: 'Mitochondrial Genes'; Slonimski, P., Borst, P. and Attardi, G. (eds.).
Cold Spring Harbor Laboratory. pp 5 - 43

Anziano, P.Q., Perlman, P.S., Lang, B.F. and Wolf, K. (1983)
Curr. Genet. 7, 273 - 284

Appleyard, R.K. (1954)
Genetics 39, 440 - 452

Attardi, G. (1981)
Trends Biochem. Sci. 6, 86 - 89 and 100 - 103

Attardi, G., Cantatore, P., Chomyn, A., Crews, S., Gelfand, R., Merkel, C., Montoya, J. and Ojala, D. (1982)
In: 'Mitochondrial Genes'; Slonimski, P., Borst, P. and Attardi, G. (eds.).
Cold Spring Harbor Laboratory. pp 51 - 71

Attardi, G., Doersen, C., Gaines, G., King, M., Montoya, J. and Shuey, D. (1983)
In: 'Nucleo-Mitochondrial Interactions'; Wolf, K., Schweyen, R.J. and Kaudewitz, F. (eds.).
De Gruyter, Berlin. (In press)

Bach, S.J., Dixon, M. and Zerfas, L.G. (1946)
Biochem. J. 40, 229 - 239

Bakeeva, L.E., Chentsov, Y.S. and Skulachev, V.P. (1978)
Biochim. Biophys. Acta. 501, 349 - 369

Barrell, B.G., Bankier, A.T. and Drouin, J. (1979)
Nature 282, 189 - 194

Barrell, B.G., Anderson, S., Bankier, A.T., de Bruijn, M.H.L., Chen, E., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G. (1980)
Proc. Natl. Acad. Sci. USA. 77, 3164 - 3166

Beckett, J.B. (1971)
Crop. Sci. 11, 724 - 727

Belliard, G., Vedel, F. and Pelletier, G. (1979)
Nature 281, 401 - 403

Bendall, D.S. (1968)
Biochem. J. 109, Proceedings of the Biochemical Society, 483rd Meeting (1968).
University College of North Wales, Bangor. p 46P

Benton, W.D. and Davis, R.W. (1977)
Science 196, 180 - 182

Berk, A.J. and Sharp, P.A. (1977)
Cell 12, 721 - 732

Bernardi, G. (1982)
Trends Biochem. Sci. 7, 404 - 408

Bernardi, G., Faures, M., Piperno, G. and Slonimski, P.P. (1970)
J. Mol. Biol. 48, 23 - 42

Bibb, M.J., Van Etten, R.A., Wright, C.T., Walberg, M.W. and Clayton, D.A. (1981)
Cell 26, 167 - 180

Birnboim, H.C. and Doly, J. (1979)
Nucl. Acids Res. 7, 1513 - 1523

Blanc, H. and Dujon, B. (1982)
In: 'Mitochondrial Genes'; Slonimski, P., Borst, P. and Attardi, G. (eds.).
Cold Spring Harbor Laboratory. pp 279 - 294

Bogenhagen, D. and Clayton, D.A. (1974)
J. Biol. Chem. 249, 7991 - 7995

Bolden, A., Noy, G.P. and Weissbach, A. (1977)
J. Biol. Chem. 252, 3351 - 3356

Bolivar, F. and Backman, K. (1979)
Methods Enzymol. 68, 245 - 267

Bolivar, F., Rodriguez, R.L., Betlach, M.C. and Boyer, H.W. (1977)
Gene 2, 75 - 93

Bonen, L. and Gray, M.W. (1980)
Nucl. Acids Res. 8, 319 - 335

Bonen, L., Huh, T.Y. and Gray, M.W. (1980)
FEBS Lett. 111, 340 - 346

Bonitz, S.G., Coruzzi, G., Thalenfeld, B.E., Tzagoloff, A. and Macino, G. (1980)
J. Biol. Chem. 255, 11927 - 11941

Bonitz, S.G., Homison, G., Thalenfeld, B.E., Tzagoloff, A. and Nobrega, F.G. (1982)
J. Biol. Chem. 257, 6268 - 6274

Bonner, T.I., Brenner, D.J., Neufeld, B.R. and Britten, R.J. (1973)
J. Mol. Biol. 81, 123 - 135

Borst, P. (1972)
Ann. Rev. Biochem. 41, 333 - 376

Borst, P. (1979)
In: 'Specific Eukaryote Genes, Alfred Benzon Symposium 13';
Munksgard. pp 244 - 258

Borst, P. (1980)
In: 'Proceedings of Mosbach Colloquia';
Bücher, Th., Sebald, W. and Weiss, H. (eds.).
Springer Verlag, Berlin. pp 27 - 41

Borst, P. and Grivell, L.A. (1978)
Cell 15, 705 - 723

Borst, P. and Grivell, L.A. (1981a)
Nature 289, 439 - 440

Borst, P. and Grivell, L.A. (1981b)
Nature 290, 443 - 444

Borst, P. and Hoeijmakers, J.H.J. (1979)
Plasmid 2, 20 - 40

Borst, P. and Kroon, A.M. (1969)
Int. Rev. Cytol. 26, 107 - 190

Bos, J.L., Osinga, K.A., Van der Horst, G., Hecht, N.B., Tabak, H.F., Van Ommen, G-J.B. and Borst, P. (1980)
Cell 20, 207 - 214

Boutry, M., Briquet, M. and Goffeau, A. (1983)
J. Biol. Chem. 258, 8524 - 8526

Brennicke, A. (1980)
Plant Physiol. 65, 1207 - 1210

Brennicke, A. and Blanz, P. (1982)
Mol. Gen. Genet. 187, 461 - 466

Brettell, R.I.S., Thomas, E. and Ingram, D.S. (1980)
Theor. Appl. Genet. 58, 55 - 58

Brettell, R.I.S., Conde, M.F. and Pring, D.R. (1982)
Maize Genet. Co-op. News Letter 56, 13 - 14

Brosius, J., Palmer, M.L., Kennedy, P.J. and Noller, H.F. (1978)
Proc. Natl. Acad. Sci. USA. 75, 4801 - 4805

Brosius, J., Dull, T.J. and Noller, H.F. (1980)
Proc. Natl. Acad. Sci. USA. 77, 201 - 204

Brown , T.A., Davies, R.W., Ray, J.A., Waring, R.B. and Scazzocchio, C. (1983)
EMBO J. 2, 427 - 435

Brown, T.A., Davies, R.W., Waring, R.B., Ray, J.A. and Scazzocchio, C. (1983a)
Nature 302, 721 - 723

Brown, W.M. (1980)
Proc. Natl. Acad. Sci. USA. 77, 3605 - 3609

Brown, W.M., George, M. and Wilson, A.C. (1979)
Proc. Natl. Acad. Sci. USA. 76, 1967 - 1971

Brownlee, G.G., Sanger, F. and Barrell, B.G. (1967)
Nature 215, 735 - 736

Burger, G., Scriven, C., Machleidt, W. and Werner, S. (1982)
EMBO J. 1, 1385 - 1391

Burger, G. and Werner, S. (1983)
In: 'Nucleo-Mitochondrial Interactions'
Wolf, K., Schweyen, R.J. and Kaudewitz, F. (eds.).
De Gruyter, Berlin. (In press)

Burke, J.M. and Raj Bhandary, U.L. (1982)
Cell 31, 509 - 520

Carter, K.R., T'sai, A-L. and Palmer, G. (1981)
FEBS Lett. 132, 243 - 246

Castora, F.J., Arnheim, N. and Simpson, M.V. (1980)
Proc. Natl. Acad. Sci. USA. 77, 6415 - 6419

Cavalier-Smith, T. (1978)
J. Cell Sci. 34, 247 - 278

Chaconas, G. and van de Sande, J.H. (1980)
Methods Enzymol. 65, 75 - 85

Chance, B. (1974)
In: 'Dynamics of Energy - Transducing Membranes';
L. Ernster et al, (eds.).
Elsevier, Amsterdam. pp 553 - 78

Chang, A.C.Y. and Cohen, S.N. (1978)
J. Bacteriol. 134, 1141 - 1156

Chao, S., Sederoff, R.R. and Levings, C.S. III (1983)
Plant Physiol. 71, 190 - 193

Chomyn, A., Mariottini, P., Gonzalez-Cadavid, N., Attardi, G., Strong, D.D., Trovato, D., Riley, M. and Doolittle, R.F. (1983)
Proc. Natl. Acad. Sci. USA. 80, 5535 - 5539

Chopra, I. and Howe, T.G.B. (1978)
Microbiol. Rev. 42, 707 - 724

Chothia, C. (1976)
J. Mol. Biol. 105, 1 - 14

Chourey, P.S. and Kemble, R.J. (1982)
In: "Plant tissue culture 1983"; Fujiwara, A. (ed.).
Japanese Association for Plant Tissue Culture, Tokyo. pp 425 - 426

Christianson, T., Edwards, J., Levens, D., Locker, J. and Rabinowitz, M. (1982)
J. Biol. Chem. 257, 6494 - 6500

Christianson, T., Edwards, J.C., Mueller, D.M. and Rabinowitz, M. (1983)
Proc. Natl. Acad. Sci. USA. 80, 5564 - 5568

Citterich, M.H., Morelli, G. and Macino, G. (1983)
EMBO J. 2, 1235 - 1242

Citterich, M.H., Morelli, G., Nelson, M.A. and Macino, G. (1983a)
In: 'Nucleo-cytoplasmic Interactions';
Wolf, K., Schweyen, R.J. and Kaudewitz, F. (eds.).
De Gruyter, Berlin. (In press)

Clark-Walker, G.D., McArthur, C.R. and Daley, D.J. (1981a)
Curr. Genet. 4, 7 - 12

Clark-Walker, G.D., McArthur, C.R. and Sriprakash, K.S. (1981b)
J. Mol. Biol. 147, 399 - 415

Clark-Walker, G.D. and Sriprakash, K.S. (1981)
J. Mol. Biol. 151, 367 - 387

Clarke, L. and Carbon, J. (1978)
Cell 9, 91 - 99

Clayton, D.A. (1982)
Cell 28, 693 - 705

Clayton, D.A., Davis, R.W. and Vinograd, J. (1970)
J. Mol. Biol. 47, 137 - 153

Clewell, D.B. (1972)
J. Bacteriol. 110, 667 - 676

Clewell, D.B. and Helinski, D.R. (1972)
J. Bacteriol. 110, 1135 - 1146

Cohen, S.N., Chang, A.C.Y. and Hsu, L. (1973)
Proc. Natl. Acad. Sci. USA. 69, 2110 - 2114

Collins, R.A. and Lambowitz, A.M. (1983)
Plasmid 9, 53 - 70

Collins, R.A., Stohl, L.L., Cole, M.D. and Lambowitz, A.M. (1981)
Cell 24, 443 - 452

Coruzzi, G., Bonitz, S.G., Thalenfeld, B.E. and Tzagoloff, A. (1981)
J. Biol. Chem. 256, 12780 - 12787

Crick, F.H.C. (1979)
Science 204, 264 - 271

Cummings, D.J. Maki, R.A., Conlon, P.J. and Laping, J. (1980)
Mol. Gen. Genet. 178, 499 - 510

Cunningham, R.S., Bonen, L., Doolittle, W.F. and Gray, M.W. (1976)
FEBS Lett. 69, 116 - 122

Dagert, M. and Ehrlich, S.D. (1979)
Gene 6, 23 - 28

Dale, R.M.K. (1981)
Proc. Natl. Acad. Sci. USA. 78, 4453 - 4457

Dale, R.M.K., Duesing, J.H. and Keene, D. (1981)
Nucl. Acids Res. 9, 4583 - 4593

Davies, J.W. and Verduin, B.J.M. (1979)
J. Gen. Virol. 44, 545 - 549

Davies, R.W., Waring, R.B., Ray, J.A., Brown, T.A. and Scazzocchio, C.
Nature 300, 719 - 724 (1982)

de Bruijn, M.H.L. (1983)
Nature 304, 234 - 241

de Jonge, J.C. and de Vries, H. (1983)
Curr. Genet. 7, 21 - 28

de Ronde, A., Van Loon, A.P.G.M. and Grivell, L.A. (1980)
Nature 287, 361 - 363

de Vries, H., de Jonge, J.C., Arnberg, A.,
Peijnenburg, A.A.C.M. and Agsteribbe, E. (1983)
In: 'Nucleo Mitochondrial Interactions';
Wolf, K., Schweyen, R.J. and Kaudewitz, F. (eds.).
De Gruyter, Berlin. (In press)

de Zamaroczy, M., Marotta, R., Faugeron-Fonty, Goursot, R.,
Mangin, M., Baldacci, G. and Bernardi, G. (1981)
Nature 292, 75 - 78

Deutsch, J., Dujon, B., Netter, P., Petrochilo, E., Slonimski, P.P.,
Bolotin-Fukuhara, M. and Coen, D. (1974)
Genetics 76, 195 - 219

Denhardt, D.T. (1966)
Biochem. Biophys. Res. Commun. 23, 641 - 646

Dickerson, R.E. and Geis, I. (1969)
'The Structure and Action of Proteins'.
Harper and Row, London. p 28

Dixon, L.K., Forde, B.G., Forde, J. and Leaver, C.J. (1980)
In: 'The Organisation and Expression of the Mitochondrial Genome';
Kroon, A.M. and Saccone, C. (eds.).
Elsevier, Amsterdam. pp 365 - 368

Dixon, L.K. and Leaver, C.J. (1982)
Plant Mol. Biol. 1, 89 - 102

Doolittle, W.F. and Sapienza, C. (1980)
Nature 284, 601 - 603

Dretzen, G., Bellard, M., Sassone-Corsi, P. and Chambon, P. (1981)
Anal. Biochem. 112, 295 - 298

Dujardin, G., Jacq. C. and Slonimski, P.P. (1982)
Nature 298, 628 - 632

Dujardin, G., Labouesse, M., Netter, P. and Slonimski, P. (1983)
In: 'Nucleo-Mitochondrial Interactions';
Wolf, F., Schweyen, R.J. and Kaudewitz, F. (eds.).
De Gruyter, Berlin. (In press)

Dujon, B. (1981)
In: 'The Molecular Biology of the Yeast Saccharomyces.
Life Cycle and Inheritance';
Strathern, J.N., Jones, E.W. and Broach, J.R. (eds.).
Cold Spring Harbor Laboratory. pp 505 - 635

Dujon, B., Colson, A.M. and Slonimski, P.P. (1977)
In: 'Mitochondria 1977';
Bandlow, W., Schewyen, R.J. Rolf, K. and Kaudewitz, F.
De Gruyter, Berlin. pp 579 - 669 (eds.).

Dugiaczyk, A., Boyer, H.W. and Goodman, H.M. (1975)
J. Mol. Biol. 96, 171 - 184

Duvick, D.N. (1965)
Adv. Genet. 13, 1 - 56

Earl, A.J. Turner, G., Croft, J.H., Dales, R.B.G.,
Lazarus, C.M., Luisdorf, H. and Küntzel, H. (1981)
Curr. Genet. 3, 221 - 228

Edwardson, J.R. (1970)
Bot. Rev. 36, 341 - 420

Efstradiatis, A., Vournakis, J.N., Donis-Keller, H., Chaconas, G.,
Dougall, D.K. and Kafatos, F.C. (1977)
Nucl. Acids Res. 4, 4165 - 4174

Ellis, J. (1982)
Nature 299, 678 - 679

Erecinska, M., Wilson, D.F. and Blasie, J.K. (1978)
Biochim. Biophys. Acta 501, 63 - 71

Falconet, D., Lejeune, B., Quetier, F. and Gray, M.W. (1983)
Submitted to EMBO J.

Fangman, W.L. (1978)
Nucl. Acid. Res. 5, 653 - 665

Farrelly, F. and Butow, R.A. (1983)
Nature 301, 296 - 301

Fauron, C.M-R. and Wolstenholme, D.R. (1980)
Nucl. Acids Res. 8, 2439 - 2452

Fiers, W., Contreras R., Haegeman, G., Rogiers, R.,
Van de Voorde, A., Van Heuverswyn, H., Van Herreweghe, J.,
Volckaert, G. and Ysebaert, M. (1978)
Nature 273, 113 - 120

Finchman, J.R.S. and Sastry, G.R.K. (1974)
Ann. Rev. Genetics 8, 15 - 50

Forde, B.G. and Leaver, C.J. (1979)
In: 'The Plant Genome'; Davies, D.R. and Hopwood, D.A. (eds.).
John Innes Charity, Norwich. pp 131 - 146

Forde, B.G. and Leaver, C.J. (1980)
Proc. Natl. Acad. Sci. USA. 77, 418 - 422

Forde, B.G. Oliver, R.J.C. and Leaver, C.J. (1978)
Proc. Natl. Acad. Sci. USA. 75, 3841 - 3845

Forde, B.G. Oliver, R.J.C. and Leaver, C.J. (1979)
Plant Physiol. 63, 67 - 73

Fox, T.D. (1979a)
J. Mol. Biol. 130, 63 - 82

Fox, T.D. (1979b)
Proc. Natl. Acad. Sci. USA. 76, 6534 - 6538

Fox, T.D. (1983)
Nature 301, 371 - 372

Fox, T.D. and Leaver, C.J. (1981)
Cell 26, 315 - 323

Frontali, L., Agostinelli, M., Baldacci, G., Falcone, C. and Zennaro, E. (1982)
In: 'Mitochondrial Genes';
Slonimski, P., Borst, P. and Attardi, G. (eds.).
Cold Spring Harbor Laboratory. pp 327 - 331

Gasser, S.M., Daum, G. and Schatz, G. (1982)
J. Biol. Chem. 257, 13034 - 13041

Gellissen, G., Bradfield, J.Y., White, B.N. and Wyatt, G.R. (1983)
Nature 301, 631 - 634

Gengenbach, B.G., Connelly, J.A., Pring, D.R. and Conde, M.F. (1981)
Theor. Appl. Genet. 59, 161 - 167

Gergen, J.P., Stern, R.H. and Wensink, P.C. (1979)
Nucl. Acids Res. 7, 2115 - 2136

Gershoni, J.M. and Palade, G. (1983)
Anal. Biochem. 131, 1 - 15

Gilbert, W.
Nature 271, 501

Godson, G.N. and Vapneck, D. (1973)
Biochim. Biophys. Acta 299, 516 - 520

Goelet, P., Lomonossoff, G.P. Butler, P.J.G., Akam, M.E., Gait, M.J. and Karn, J. (1982)
Proc. Natl. Acad. Sci. USA. 79, 5818 - 5822

Goldbach, R.W., Bollen-de Boer, J.E., van Bruggen, E.F.J. and Borst, P. (1978)
Biochim. Biophys. Acta 521, 187 - 197

Görtz, G. and Feldmann, H. (1982)
Curr. Genet. 5, 221 - 225

Gray, M.W. (1983)
Bioscience (In press)

Gray, M.W. and Doolittle, W.F. (1982)
Microbiol. Rev. 46, 1 - 42

Gray, M.W. and Spencer, D.F. (1983)
FEBS Lett. 161, 323 - 327

Green, D.E. and Hatefi, Y. (1961)
Science 133, 13 - 19

Grivell, L.A. (1983)
In: 'Nucleo-Mitochondrial Interactions';
Wolf, K., Schweyen, R.J. and Kaudewitz, F., (eds.).
De Gruyter, Berlin. (In press)

Grivell, L.A. (1983a)
Scientific American 248, 78 - 89

Grivell, L.A., Bonen, L. and Borst, P. (1982)
In: "Horizons in Biochemistry and Biophysics Vol. 7";
Palmieri, F. and Quagliariello, E. (eds.).
Wiley, Chichester. In press

Gronenborn, B. and Messing, J. (1978)
Nature 272, 375 - 377

Groot, G.S.P., Mason, T.L. and van Harten-Loosbroek, N. (1979)
Mol. Gen. Genet. 174, 339 - 342

Groudinsky, O., Dujardin, G. and Slonimski, P.P. (1981)
Mol. Gen. Genet. 184, 493 - 503

Grunstein, M. and Hogness, D.S. (1975)
Proc. Natl. Acad. Sci. USA. 72, 3961 - 3965

Guillemaut, P. and Weil, J.H. (1975)
Biochim. Biophys. Acta 407, 240 - 248

Hack, E. and Leaver, C.J. (1983)
EMBO J. 2, 1783 - 1789

Haddock, B.A. and Jones, C.W. (1977)
Bacteriol. Rev. 41, 47 - 99

Haid, A., Schweyen, R.J., Bechmann, H., Kaudewitz, F., Solioz, M. and Schatz, G. (1979)
Eur. J. Biochem 94, 451 - 464

Hatefi, Y., Haavik, A.G. and Griffiths, D.E. (1962)
J. Biol. Chem. 237, 1681 - 1685

Hauska, G., Hurt, E., Gabellini, N. and Lockau, W. (1983)
Biochim. Biophys. Acta 726, 97 - 133

Hauswirth, W.W. and Laipis, P.J. (1982)
In: 'Mitochondrial Genes'; Slonimski, P., Borst, P. and Attardi, G. (eds.).
Cold Spring Harbor. pp 137 - 141

Hensgens, L.A.M., Grivell, L.A., Borst, P. and Bos, J.L. (1979)
Proc. Natl. Acad. Sci. USA. 76, 1663 - 1667

Hensgens, L.A.M., Bonen, L., de Haan, M., van der Horst, G. and Grivell, L.A. (1983a)
Cell 32, 379 - 389

Hensgens, L.A.M. Arnberg, A.C., Roosendaal, E., van der Horst, G., van der Veen, R., van Ommen, G-J.B. and Grivell, L.A. (1983b)
J. Mol. Biol. 164, 35 - 58

Hiesel, R. and Brennicke, A. (1983)
Submitted to EMBO J.

Hines, J.C. and Ray, D.S. (1980)
Gene II, 207 - 218

Hirsch, M. and Penman, S. (1974)
Cell 3, 335 - 339

Hofschneider, P.H. and Preuss, A. (1963)
J. Mol. Biol. 7, 450 - 451

Hohn, B. and Collins, J. (1980)
Gene 11, 291 - 298

Hollenburg, C.P., Borst, P. and van Bruggen, E.F.J. (1970)
Biochim. Biophys. Acta 209, 1 - 15

Holmes, D.S. and Quigley, M. (1981)
Anal. Biochem. 114, 193 - 197

Howley, P.M., Israel, M.A., Law, M-F. and Martin, M.A. (1979)
J. Biol. Chem. 254, 4876 - 4883

Hu, N-t. and Messing, J. (1982)
Gene 17, 271 - 277

Hudspeth, M.E.S., Ainley, M.W., Shumard, D.S., Butow, R.A. and Grossman, L.I. (1982)
Cell 30, 617 - 626

Humphreys, G.O., Weston, A., Brown, M.G.M. and Saunders, J.R. (1979)
In: 'Transformation 1978'; Glover, S.W. and Butler, L.O. (eds.).
Cotswold Press, Oxford. pp 287 - 312

Iams, K.P. and Sinclair, J.H. (1982)
Proc. Natl. Acad. Sci. USA. 79, 5926 - 5929

Itagaki, E. and Hager, L.P. (1968)
Biochem. Biophys. Res. Commun. 32, 1013 - 1019

Jacobs, H.T., Posakony, J.W., Grula, J.W., Roberts, J.W., Xin, J-H, Britten, R.J. and Davidson, E.H. (1983)
J. Mol. Biol. 165, 609 - 632

Jacq, C., Pajot, P., Lazowska, J., Dujardin, G., Claisse, M., Groudinsky, O., de la Salle, H., Grandchamp, C., Labouesse, M., Gargouri, A., Guiard, B., Spyridakis, A., Dreyfus, M. and Slonimski, P.P. (1982)
In: "Mitochondrial Genes"; Slonimski, P., Borst, P. and Attardi, G. (eds.).
Cold Spring Harbor Laboratory. pp 155 - 183

Jovin, T.M. (1971)
Methods Enzymol 21, 179 - 187

Keilin, D. (1925)
Proc. Roy. Soc. B 98, 312 - 339

Keilin, D. (1927)
Compt. Rend. Soc. Biol. Paris 96, Sp39 - Sp68

Kelley, W.S. and Stump, K.H. (1979)
J. Biol. Chem. 254, 3206 - 3210

Kemble, R.J. and Bedbrook, J.R. (1980)
Nature 284, 565 - 566

Kemble, R.J., Gunn, R.E. and Flavell, R.B. (1983)
Theor. Appl. Genet. 65, 129 - 144

Kemble, R.J. Mans, R.J., Gabay-Laughnan, S. and Laughnan, J.R.
Nature 304, 744 - 747

Kemble, R.J. and Thompson, R.D. (1982)
Nucl. Acids Res. 10, 8181 - 8192

Kemble, R.J. Flavell, R.B. and Brettell, R.I.S. (1982)
Theor. Appl. Genet. 62, 213 - 217

Kim, B.D., Lee, K.J. and De Busk, A.G. (1982)
FEBS Lett. 147, 231 - 234

Kim, B.D., Mans, R.J. Conde, M.F., Pring, D.R. and
Levings, C.S. III (1982a)
Plasmid 7, 1 - 14

Kleckner, N., Roth, J. and Botstein, D. (1977)
J. Mol. Biol. 116, 125 - 159

Köchel, H.G. and Küntzel, H. (1981)
Nucl. Acids Res. 9, 5689 - 5696

Koller, B., Delius, H. and Dyer, T.A. (1982)
Eur. J. Biochem 122, 17 - 23

Kolodner, R. and Tewari, K.K. (1972)
Proc. Natl. Acad. Sci. USA. 69, 1830 - 1834

Kovak, L., Lazowska, J. and Slonimski, P.P. (1983)
In: "Nucleo-Mitochondrial Interactions"; Wolf, K.,
Schweyen, R.J. and Kaudewitz, F. (eds.).
De Gruyter, Berlin. (In press)

Kozak, M. (1978)
Cell 15, 1109 - 1123

Kreike, J. (1982)
'The Biosynthesis of Complex III in Yeast'
Ph.D. Thesis, University of Amsterdam.

Kreike, J., Bechmann, H., van Hemert, F.J., Schweyen, R.J., Boer, P.H., Kaudewitz, F. and Groot, G.S.P. (1979)
Eur. J. Biochem. 101, 607 - 617

Küntzel, H. and Köchel, H.G. (1981)
Nature 293, 751 - 753

Küntzel, H. and Noll, H. (1967)
Nature 215, 1340 - 1345

Kyte, J. and Doolittle, R.F. (1982)
J. Mol. Biol. 157, 105 - 132

Labouesse, M. and Slonimski, P.P. (1983)
EMBO J. 2, 269 - 276

Lang, B.F., Ahne, F. Distler, S., Trinkl, H., Kaudewitz, F. and Wolf, K.
In: 'Nucleo-Mitochondrial Interactions'; Wolf, K., Schweyen, R.J. and Kaudewitz, F. (eds.).
De Gruyter, Berlin. (In press)

Laskey, R.A. and Mills, A.D. (1975)
Eur. J. Biochem. 56, 335 - 341

Laughnan, J.R. and Gabay, S.J. (1973)
Theor. Appl. Genet. 43, 109 - 116

Laughnan, J.R. and Gabay, S.J. (1975)
In: 'Maize Genetics and Breeding'; Walden, D.B. (ed.).
Wiley, New York. pp 427 - 446

Laughnan, J.R. and Gabay, S.J. (1975a)
In: "Genetics and Biogenesis of Mitochondria and Chloroplants"; Birky, W.W. Jr., Perlman, P.S. and Byers, T.J. (eds.).
Ohio State University Press, Columbus. pp 330 - 349

Lazarus, C.M., Lünsdorf, H., Hahn , U., Stepien, P.P. and Küntzel, H. (1980)
Mol. Gen. Genet. 177, 389 - 397

Lazowska, J., Jacq, C. and Slonimski, P.P. (1980)
Cell 22, 333 - 348

Leaver, C.J. (1980)
Trends Biochem. Sci. 5, 248 - 252

Leaver, C.J. and Forde, B.G. (1980)
In: 'Genome Organisation and Expression in Higher Plants'; Leaver, C.J. (ed.).
Plenum, New York. pp 407 - 425

Leaver, C.J., Forde, B.G., Dixon, L.K. and Fox, T.D. (1982)
In: 'Mitochondrial Genes'; Slonimski, P., Borst, P. and Attardi, G. (eds.)
Cold Spring Harbor Laboratory. pp 457 - 470

Leaver, C.J. and Gray, M.W. (1982)
Ann. Rev. Plant Physiol. 33, 373 - 402

Leaver, C.J., Hack, E., Dawson, A.J., Isaac, P.G. and Jones, V.P. (1983)
In: 'Nucleo-Mitochondrial Interactions'; Wolf, K., Schweyen, R.J. and Kaudewitz, F. (eds.).
De Gruyter, Berlin. (In press)

Leaver, C.J. and Harmey, M.A. (1973)
Biochem. Soc. Symp. 38, 175 - 193

Leaver, C.J. and Harmey, M.A. (1976)
Biochem. J. 157, 275 - 277

Leaver, C.J. and Pope, P.K. (1976)
In: 'Nucleic Acids and Protein Synthesis in Plants'; Bogorad, L. (ed.)
Plenum Press, New York. pp 213 - 237

Lebacq, P. and Vedel, F. (1981)
Plant Sci. Lett. 23, 1 - 9

Lehrach, H., Diamond, D., Wozney, J.M. and Boedtker, H. (1977)
Biochemistry - USA. 16, 4743 - 4751

Lemberg, R. and Barrett, J. (1973)
'The Cytochromes'
Academic Press, New York

Levens, D., Ticho, B., Ackerman, E. and Rabinowitz, M. (1981)
J. Biol. Chem. 256, 5226 - 5232

Levings, C.S. III (1983a)
Cell 32, 659 - 661

Levings, C.S. III (1983b)
In: 'Genetic Engineering of Plants, An Ecological Perspective'; Kosuge, T., Meredith, C.P. and Hollaender, A. (eds.).
Plenum Press, New York. pp 81 - 92

Levings, C.S. III, Kim, B.D., Pring, D.R., Conde, M.F., Mans, R.J., Laughnan, J.R. and Gabay-Laughman, S.J. (1980)
Science 209, 1021 - 1023

Levings, C.S. III and Pring, D.R. (1976)
Science 193, 158 - 160

Levings, C.S. III and Pring, D.R. (1977)
J. Hered. 68, 350 - 354

Levings, C.S. III and Sederoff, R.R. (1983)
Proc. Natl. Acad. Sci. USA. 80, 4055 - 4059

Levings, C.S. III, Shah, D.M., Hu, W.W.L., Pring, D.R. and Timothy, D.H. (1979)
In: 'Extrachromosomal DNA, ICN-UCLA Symposium on Molecular and Cellular Biology'; Cummings, D.J., Borst, P. Dawid, I.B., Weissman, S.M. and Fox, C.F. (eds.).
Academic Press, New York. pp 63 - 73

Lewin, B. (1977)
'Gene Expression - 3, Plasmids and Phages'.
John Wiley and Sons, New York. p 278

Lewin, R. (1983)
Science 219, 478 - 479

Li, M. and Tzagoloff, A. (1979)
Cell 18, 47 - 53

Li M., Tzagoloff, A., Underbrink-Lyon, K. and Martin, N.C. (1982)
J. Biol. Chem. 257, 5921 - 5928

Linnane, A.W. and Lukins, H.B. (1975)
In: 'Methods in Cell Biology vol. XII'; Prescott, D.M. (ed.).
Academic Press, New York. pp 285 - 309

Lonsdale, D.M., Hodge, T.P., Howe, C.J. and Stern, D.B. (1983)
Cell. (In press)

Lonsdale, D.M., Hodge, T.P., Fauron, C.M-R. and Flavell, R.B. (1983a)
In: 'Plant Molecular Biology' UCLA Symposia on Molecular Biology New Series Vol. 12; Goldberg, R.B. (ed.).
In press

Lonsdale, D.M., Thompson, R.D. and Hodge, T.P. (1981)
Nucl. Acid. Res. 9, 3657 - 3669

Luck, D.J.L. (1963)
J. Cell Biol. 16, 483 - 499

Luck, D.J.L. and Reich, E. (1964)
Proc. Natl.. Acad. Sci. USA. 52, 931 - 938

Luria, S.E. and Delbrück, M. (1943)
Genetics 28, 491 - 511

Machleidt, W. and Werner, S. (1979)
FEBS Lett. 107, 327 - 330

Macino, G. and Tzagoloff, A. (1979)
J. Biol. Chem. 254, 4617 - 4623

Macino, G. and Tzagoloff, A. (1980)
Cell 20, 507 - 517

Macreadie, I.G., Novitski, C.E., Maxwell, R.J., John, U., Ooi, B-G., McMullen, G.L., Lukins, H.B., Linnane, A.W. and Nagley, P. (1983)
Nucl. Acids Res. 11, 4435 - 4451

Mandel, M. and Higa, A. (1970)
J. Mol. Biol. 53, 159 - 162

Maniatis, T., Jeffrey, A. and Kleid, D.G. (1975)
Proc. Natl. Acad. Sci. USA. 72, 1184 - 1188

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982)
'Molecular Cloning, A Laboratory Manual'
Cold Spring Harbor Laboratory. p 468

Mariottini, P., Chomyn, A., Attardi, G., Trovato, D., Strong, D.D. and Doolittle, R.F. (1983)
Cell 32, 1269 - 1277

Marmur, J. and Doty, P. (1962)
J. Mol. Biol. 5, 109 - 118

Marshall, J. (1952)
Phys. Rev. 86, 685 - 693

Matthews, D.E., Gregory, P. and Gracen, V.E. (1979)
Plant Physiol. 63, 1149 - 1153

Maxam, A.M. and Gilbert, W. (1977)
Proc. Natl. Acad. Sci. USA. 74, 560 - 564

McArthur, C.R. and Clark-Walker, G.D. (1983)
Curr. Genet. 7, 29 - 35

McConaughy, B.L., Laird, C.D. and McCarthy, B.J. (1969)
Biochemistry - USA. 8, 3289 - 3295

McDonell, M.W., Simon, M.N. and Studier, F.W. (1977)
J. Mol. Biol. 110, 119 - 146

McLean, J.R., Cohn, G.L., Brandt, I.K. and Simpson, M.V. (1958)
J. Biol. Chem. 233, 657 - 663

McMaster, G.K. and Carmichael, G.G. (1977)
Proc. Natl. Acad. Sci. USA. 74, 5350 - 5354

McNay, J.W., Pring, D.R. and Lonsdale, D.M. (1983)
Submitted to Plant Mol. Biol.

Mendel-Hartvig, I. and Nelson, B.D. (1981)
Biochim. Biophys. Acta 636, 91 - 97

Merten, S., Synenki, R.M., Locker, J., Christianson, T. and Rabinowitz, M. (1980)
Proc. Natl. Acad. Sci. USA. 77, 1417 - 1421

Messing,J. (1979)
Recombinant DNA Technical Bulletin, NIH Publication
No. 79 - 99, 2, No. 2, 43 - 48

Messing, J., Gronenborn, B., Muller-Hill, B. and Hofschneider P.H. (1977)
Proc. Natl. Acad. Sci. USA. 74, 3642 - 3646

Messing,J. and Vieira, J. (1982)
Gene 19, 269 - 276

Mitchell, P.D. (1961)
Nature 191, 144 - 148

Mitchell, P.D. (1976)
J. Theor. Biol. 62, 327 - 367

Montoya, J., Ojala, D. and Attardi, G. (1981)
Nature 290, 465 - 470

Moore, G. (1983)
In: 'Focus' (Bethesda Research Laboratories, Inc.)
Vol. 5, No. 3. p 11

Murphy, W.I., Attardi, B., Tu, C. and Attardi, G. (1975)
J. Mol. Biol. 99, 809 - 814

Murray, N., Brammar, W.J. and Murray, K. (I977)
Mol. Gen. Genet. 150, 53 - 61

Nelson, M.A. and Macino, G. (1983)
In: 'Nucleo-Mitochondrial Interactions'; Wolf, K.,
Schweyen, R.J. and Kaudewitz, F. (eds.).
De Gruyter, Berlin. (In press)

Nelson, N. and Neumann, J. (1972)
J. Biol. Chem. 247, 1917 - 1924

Nelson, N. and Schatz, G. (1979)
Proc. Natl. Acad. Sci. USA. 76, 4365 - 4369

Neupert, W. and Schatz, G. (1981)
Trends Biochem. Sci. 6, 1 - 4

Nobrega, F.G. and Tzagoloff, A. (1980)
J. Biol. Chem. 255, 9828 - 9837

O'brien, T.W. and Matthews, D.E. (1976)
In: 'Handbook of Genetics Vol. 5'; King, R.C. (ed.).
Plenum, New York. pp 535 - 580

Ojala, D., Montoya, J. and Attardi, G. (1981)
Nature 290, 470 - 474

Olszewska, E. and Tait, A. (1980)
Mol. Gen. Genet. 178, 453 - 457

Orgel, L.E. and Crick, F.H.C. (1980)
Nature 284, 604 - 607

Osiewacz, H.D. and Esser, K. (1983)
Curr. Genet. 7, 219 - 223

Osinga, K.A., De Haan, M., Christianson, T. and Tabak, H.F. (1982)
Nucl. Acids Res. 10, 7993 - 8006

Osinga, K.A. and Tabak, H.F. (1982)
Nucl. Acids Res. 10, 3617 - 3626

Palmer, J.D. (1983)
Nature 301, 92 - 93

Palmer, J.D. and Shields, C.R. (1983)
Nature (In press)

Palmer, J.D., Shields, C.R., Cohen, D.B. and Orton, T.J. (1983)
Nature 301, 725 - 728

Peacock, A.C. and Dingman, C.W. (1967)
Biochemistry - USA. 6, 1818 - 1827

Philippsen, P., Kramer, R.A. and Davis, R.W. (1978)
J. Mol. Biol. 123, 371 - 386

Piko, L. and Matsumoto, L. (1976)
Dev. Biol. 49, 1 - 10

Plattner, H., Salpeter, M.M., Saltzgaber, J. and Schatz, G. (1970)
Proc. Natl. Acad. Sci. USA. 66, 1252 - 1259

Powling, A. (1981)
Mol. Gen. Genet. 183, 82 - 84

Pratje, E. Mannhaupt, G., Michaelis, G. and Beyreuther, K. (1983)
EMBO J. 2, 1049 - 1054

Pratt, D., Tzagoloff, H. and Erdahl, W.S. (1966)
Virology 30, 397 - 410

Prentki, P., Karch, F., Iida, S. and Meyer, J. (1981)
Gene 14, 289 - 299

Pribnow, D. (1975)
Proc. Natl. Acad. Sci. USA. 72, 784 - 788

Pring, D.R., Conde, M.F. and Levings, C.S. III (1980)
Crop Sci. 20, 159 - 162

Pring, D.R., Conde, M.F., Schertz, K.F. and Levings, C.S. III (1982)
Mol. Gen. Genet. 186, 180 - 184

Pring, D.R. and Levings, C.S.III (1978)
Genetics 89, 121 - 136

Pring, D.R., Levings, C.S. III, Hu, W.W.L. and Timothy, D.H. (1977)
Proc. Natl. Acad. Sci. USA. 74, 2904 - 2908

Prunell, A. and Bernardi, G. (1974)
J. Mol. Biol. 86, 825 - 841

Quetier, F. and Vedel F. (1977)
Nature 268, 365 - 368

Quetier, F. and Vedel, F. (1980)
In: 'Genome Organisation and Expression in Plants';
Leaver, C.J. (ed.).
Plenum, New York. pp 401 - 406

Radloff, R., Bauer, W. and Vinograd, J. (1967)
Proc. Natl. Acad. Sci. USA. 57, 1514 - 1521

Randerath, K. (1970)
Anal. Biochem. 34, 188 - 205

Rastl, E. and Dawid, I.B. (1979)
Cell 18, 501 - 510

Ravetch, J.V., Model, P. and Robertson, H.D. (1977)
Nature 265, 698 - 702

Rieske, J.S. (1973). See Wikström (1973).

Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Bery, P. (1977)
J. Mol. Biol. 113, 237 - 251

Rogers, J.S. and Edwardson, J.R. (1952)
Agron. J. 44 8 - 13

Rosenberg, M. and Court, D. (1979)
Ann. Rev. Genetics 13, 319 - 353

Sanders, J.P.M., Flavell, R.A., Borst, P. and Mol, J.N.M. (1973)
Biochim. Biophys. Acta. 312, 441 - 457

Sanders, J.P.M., Heyting, C., Verbeet, M. Ph., Meijlink, F.C.P.W. and Borst, P. (1977)
Mol. Gen. Genet. 157, 239 - 261

Sanger, F., Nicklen, S. and Coulson, A.R. (1977)
Proc. Natl. Acad. Sci. USA. 74, 5463 - 5467

Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., Hutchison, C.A. III, Slocombe, P.M. and Smith, M. (1977a)
Nature 265, 687 - 695

Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980)
J. Mol. Biol. 143, 161 - 178

Sanger, F., Coulson, A.R., Hong, G.F., Hill, D.F., and Peterson, G.B. (1982)
J. Mol. Biol. 162, 729 - 773

Saraste, M. (1983)
Submitted to FEBS Lett.

Saraste, M. and Wikström, M. (1983)
In: 'Structure and Function of Membrane Proteins';
Palmieri, F. et al (eds.).
Elsevier, Amsterdam. (In press)

Scazzocchio, C., Brown, T.A., Waring, R.B. and Davies, R.W. (1983)
In: 'Nucleo-Mitochondrial Interactions';
Wolf, K., Schweyen, R.J. and Kaudewitz, F. (eds.).
De Gruyter, Berlin. (In press)

Schatz, G. and Butow, R. (1983)
Cell 32, 316 - 318

Schatz, G. and Mason, T.L. (1974)
Ann. Rev. Biochem. 43, 51 - 87

Schildkraut, C.L., Marmur, J. and Doty, P. (1962)
J. Mol. Biol. 4, 430 - 443

Schreier, P.H., and Cortese, R. (1979)
J. Mol. Biol. 129, 169 - 172

Schweyen, R.J., Francisci, S., Haid, A., Ostermayr, R., Rödel, G., Schmelzer, C., Schroeder, R., Weiss-Brummer, B. and Kaudewitz, F. (1982).
In: 'Mitochondrial Genes'; Slonimski, P., Borst, P. and Attardi, G. (eds.).
Cold Spring Harbor. pp 201 - 212

Sebald, W. (1977)
Biochim. Biophys. Acta. 463, 1 - 27

Sebald, W., Hoppe, J. and Wachter, E. (1979)
In: 'Function and Molecular aspects of Biomembrane Transport';
Quagliariello, E., (ed.)
Elsevier, Amsterdam. pp 63 - 74

Sederoff, R.R., Levings, C.S. III, Timothy, D.H. and Hu, W.W.L. (1981)
Proc. Natl. Acad. Sci. USA. 78, 5953 - 5957

Sevarino, K.A. and Poyton, R.O. (1980)
Proc. Natl. Acad. Sci. USA. 77, 142 - 146

Shah, D.M. and Levings, C.S. III (1974)
Crop. Sci. 14, 852 - 853

Sharp, P.A., Sugden, B. and Sambrook, J. (1973)
Biochemistry - USA 12, 3055 - 3063

Shine, J. and Dalgarno, L. (1974)
Proc. Natl. Acad. Sci. USA. 71, 1342 - 1346

Slater, E.C. (1981)
In: 'Chemiosmotic Protein Circuits in Biological Membranes';
Skulachev, V.P. and Hinkle, P.C. (eds.).
Addison-Wesley, London. pp 69 - 104

Slater, E.C. (1983)
Trends Biochem. Sci. 8, 239 - 242

Slonimski, P.P. and Tzagoloff, A. (1976)
Eur. J. Biochem. 61, 27 - 41

Smith, S.M. and Ellis, R.J. (1981)
J. Mol. Appl. Genet. 1, 127 - 137

Soberon, X., Covarrubias, L. and Bolivar, F. (1980)
Gene 9; 287 - 305

Southern, E.M. (1975)
J. Mol. Biol. 98, 503 - 517

Sparks, R.B. Jr. and Dale, R.M.K. (1980)
Mol. Gen. Genet. 180, 351 - 355

Spencer, D.F., Schnare, M.N. and Gray, M.W. (1983)
Submitted to EMBO J.

Spruill, W.M. Jr., Levings, C.S. III and Sederoff, R.R. (1980)
Dev. Genet. 1, 363 - 378

Spruill, W.M. Jr., Levings, C.S. III and Sederoff, R.R. (1981)
Dev. Genet. 2, 319 - 336

Steitz, J.A. and Jakes, K. (1975)
Proc. Natl. Acad. Sci. USA. 72, 4734 - 4738

Stern, D.B., Dyer T.A. and Lonsdale, D.M. (1982)
Nucl. Acids Res. 10, 3333 - 3340

Stern, D.B. and Lonsdale, D.M. (1982)
Nature 299, 698 - 702

Stern, D.B., Palmer, J.D., Thompson, W.F. and Lonsdale, D.M. (1983)
In: 'Plant Molecular Biology' UCLA Symposia on Molecular Biology New Series Vol. 12; Goldberg, R.B. (ed.).
In press

Stevens, B. (1981)
In: 'The Molecular Biology of the Yeast Saccharomyces. Life Cycle and Inheritance'; Strathern, J.N., Jones, E.W. and Broach, J.R. (eds.).
Cold Spring Harbor Laboratory. pp 471 - 504

Storey, B.T. (1969)
Plant Physiol. 44, 413 - 421

Sutcliffe, J.G. (1978)
Cold Spring Harbor Symp. Quant. Biol. 43, 77 - 90

Suyama, Y. and Miura, K. (1968)
Proc. Natl. Acad. Sci. USA. 60, 235 - 242

Synenki, R.M., Levings, C.S. III and Shah, D.M. (1978)
Plant Physiol. 61, 460 - 464

Szalay, A.A., Mackey, C.J. and Langridge, W.H.R. (1979)
Enzyme Microb. Technol. 1, 154 - 163

Talen, J.L., Sanders, J.P.M. and Flavell, R.A. (1974)
Biochim. Biophys. Acta 374, 129 - 135

Tanaka, T. and Weisblum, B. (1975)
J. Bacteriol. 121, 354 - 362

Tanford, C. (1978)
Science 200, 1012 - 1018

Taniguchi, T. and Weissmann, C. (1978)
J. Mol. Biol. 118, 533 - 565

Taylor, J.M., Illmensee, R. and Summers, J. (1976)
Biochim. Biophys. Acta 442, 324 - 330

Thalenfeld, B.E., Hill, J. and Tzagoloff, A. (1983)
J. Biol. Chem. 258, 610 - 615

Thomas, P.S. (1980)
Proc. Natl. Acad. Sci. USA. 77, 5201 -5205

Thompson, R.D., Kemble, R.J. and Flavell, R.B. (1980)
Nucl. Acids Res. 8, 1999 - 2008

Thuring, R.W.J., Sanders, J.P.M., and Borst, P. (1975)
Anal. Biochem. 66, 213 - 220

Tientze, M., Slaughter, M., Weiss, H. and Neupert, W. (1982)
J. Biol. Chem. 257, 10364 - 10371

Timmis, J.N. and Scott, N.S. (1983)
Nature 305, 65 - 67

T'sai, A-L. and Palmer, G. (1982)
Biochim. Biophys. Acta 681, 484 - 495

Tudzynski, P., Düvell, A. and Esser, K. (1983)
Curr. Genet. 7, 145 - 150

Tzagoloff, A. (1982)
'Mitochondria'.
Plenum Press, New York.

Ullstrup, A.J. (1972)
Ann. Rev. Phytopath. 10, 37 - 50

van den Boogaart, P., van Dijk, S. and Agsteribbe, E. (1982)
FEBS Lett. 147, 97 - 100

van den Boogaart, P., Samallo, J. and Agsteribbe, E. (1982a)
Nature 298, 187 - 189

van Ommen, G-J.B., Groot, G.S.P. and Grivell, L.A. (1979)
Cell 18, 511 - 523

van Wezenbeek, P.M.G.F., Hulsebos, T.J.M. and Schoenmakers, J.G.G. (1980)
Gene 11, 129 - 148

Vedel, F., Mathieu, C., Lebacq, P., Ambard-Bretteville, F., Remy, R., Pelletier, G., Renard, M. and Rousselle, P. (1982)
Theor. Appl. Genet. 62, 255 - 262

Viebrock, A., Perz, A. and Sebald, W. (1982)
EMBO J. 1, 565 - 571

von Heijne, G. and Blomberg, C. (1979)
Eur. J. Biochem. 97, 175 - 181

von Jagow, G., Schägger, H., Engel, W.D., Machleidt, W. and Machleidt, I. (1978)
FEBS Lett. 91, 121 - 125

Wahl, G.M., Stern, M. and Stark, G.R. (1979)
Proc. Natl. Acad. Sci. USA. 76, 3683 - 3687

Wallace, D.C. (1982)
Microbiol. Rev. 46, 208 - 240

Wallace, D.C., Oliver, N.A., Blanc, H. and Adams, C.W. (1982)
In: 'Mitochondrial Genes';
Slonimski, P., Borst, P. and Attardi, G. (eds.).
Cold Spring Harbor Laboratory. pp 105 - 116

Walton, J.D. (1983)
Plant Mol. Biol. Reporter 1, 101 - 103

Ward, B.L., Anderson, R.S. and Bendich, A.J. (1981)
Cell 25, 793 - 803

Waring, R.B., Davies, R.W., Lee, S., Grisi, E., McPhail Berks, M. and Scazzocchio, C. (1981)
Cell 27, 4 - 11

Warmke, H.E. and Lee, S.L.J. (1977)
J. Hered. 68, 213 - 222

Weislander, L. (1979)
Anal. Biochem. 98, 305 - 309

Weiss, H. and Kolb, H.J. (1979)
Eur. J. Biochem. 99, 139 - 149

Weissinger, A.K., Timothy, D.H., Levings, C.S. III, Hu, W.W.L. and Goodman, M.M. (1981)
Maize Genet. Co-op. Newsletter 55, 86

Weissinger, A.K., Timothy, D.K., Levings, C.S. III and Goodman, M.M. (1983)
Genetics 104, 365 - 379

Wells, R. and Birnsteil, M. (1969)
Biochem. J. 112, 777 - 786

Wells, R. and Ingle, J. (1970)
Plant Physiol. 46, 178 - 179

Wesolowski, M. snd Fukuhara, H. (1981)
Mol. Cell. Biol. 1, 387 - 393

Whitfield, P.R. and Bottomley, W. (1983)
Ann. Rev. Plant Physiol. 34, 279 - 310

Widger, W.R., Cramer, W.A., Herrmann, R. and Trebst, A. (1983)
Submitted to Proc. Natl. Acad. Sci. USA.

Wikström, M.K.F. (1973)
Biochim. Biophys. Acta 301, 155 - 193

Wikström, M., Krab, K. and Saraste, M. (1981)
Ann. Rev. Biochem. 50, 623 - 655

Wikström, M. and Saraste, M. (1983)
See Saraste and Wikström (1983)

Wright, R.M. and Cummings, D.J. (1983)
Nature 302, 86 - 88

Wright, R.M. and Cummings, D.J. (1983a)
Curr. Genet. 7, 151 - 157

Wright, R.M., Horrum, M.A. and Cummings, D.J. (1982)
Cell 29, 505 - 515

Wong, F.Y. and Wildman, S.G. (1972)
Biochim. Biophys. Acta 259, 5 - 12

Yamamoto, K., Alberts, B.M., Benzinger, R., Lawhorne, L. and Treiber, G. (1970)
Virology 40, 734 - 744

Yin, S., Heckman, J., Sarnoff, J. and Raj Bhandary, U.L. (1980)
In: 'Organisation and Expression of the Mitochondrial Genome'; Kroon, A.M. and Saccone, C. (eds.).
Elsevier, Amsterdam. pp 307 - 310

Yu, C-A. and Yu, L. (1981)
Biochim. Biophys. Acta 639, 99 - 128

Zwieb, C., Glotz, C. and Brimacombe, R. (1981)
Nucl. Acids Res. 9, 3621 - 3640