NUCLEOTIDE SEQUENCE FROM A RIBOSOMAL RNA TRANSCRIPTION UNIT OF *Xenopus laevis*

Lucinda M. C. Hall

Thesis submitted to the University of Glasgow

for the degree of Doctor of Philosophy

Department of Biochemistry

September 1981

NAME	ADDRESS AND AFFILIATION	DATE

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

J. S. HERON LTD
5 QUEENS CRESCENT
ST. GEORGES CROSS
GLASGOW 041 332 1883

Page

CONTENTS

FIGURES AND TABLES

ACKNOWLEDGEMENTS

I would like to express my gratitude to Dr. Ted Maden for his encouragement and good advice throughout this project.

I would like to thank Mary Robertson for preparing plasmid DNA on several occasions.

My thanks to Professors R.M.S. Smellie and A.R. Williamson for providing facilities in this department. Also to Mr. P. Ferry and the staff of the Wellcome Cell Culture Unit for maintaining and providing cells.

I am very grateful to Mrs. Strachan for miraculously quick and efficient typing and to Mr. Ian Ramsden for photography.

Finally thanks to all my friends in C13 and elsewhere, and to Michael, for reminding me that there is more to life than ribosomal DNA.

ABBREVIATIONS

Abbreviations used are as recommended in the Biochemical Journal Instructions to Authors, 1981, with the following additions:-

rRNA	ribosomal RNA
rDNA	genes for ribosomal RNA and associated spacer DNA
ITS	internal transcribed spacer
ETS	external transcribed spacer
NTS	non-transcribed spacer
bp	base pairs
kb	kilobases (1000 bases)
DBM paper	diazobenzyloxymethyl paper

SUMMARY

I have determined the sequence of the central part of a ribosomal transcription unit from Xenopus laevis, using the plasmid pX1r101. The sequence comprises over 200 bp at the 3' end of the 18S gene, the first internal transcribed spacer, the 5.8S gene, the second internal transcribed spacer, and over 100 bp at the 5' end of the 28S gene. The two transcribed spacers have G + C contents of over 80% and include long homopolymeric tracts of G or C (10-15 residues). ITS1 also has long tracts of purines containing several A residues.

The Xenopus sequence is compared to other organisms where data is available. The gene sequences show a high level of homology with sequences from other eukaryotes and also some homology with the prokaryote E. coli. No sequence homology is found between the internal transcribed spacers of Xenopus and yeast (Saccharomyces).

Tentative secondary structure models are proposed for the Xenopus sequence and again compared to possible models from other organisms. Secondary structure may be highly conserved within the mature rRNAs, even in regions where the primary sequence is variable between species. In the transcribed spacers one hairpin may be held in common by Xenopus and yeast but other secondary structures are not obviously conserved.

I have attempted to characterise some ribosomal RNA precursors in Xenopus tissue culture cells by both 'Northern' transfers and S1 nuclease protection mapping. Various artifacts limit the usefulness of these techniques in this system. However it is proposed that a putative '30S' precursor exists containing the RNA of 5.8S, ITS2 and 28S, and having the same 5' end as 5.8S rRNA.

Evidence from the sequence supports the proposition that 5.8S rRNA in eukaryotes is structurally equivalent to the 5' end of 23S rRNA in E. coli The results lead to speculation of the relationship between rRNA processing in eukaryotes and E. coli.

CHAPTER 1. GENERAL INTRODUCTION

Ribosomal RNA is a major component in the protein synthetic machinery of the cell. All organisms (except viruses) have rRNA genes, usually in multiple copies. In most eukaryotes these genes are arranged in tandem arrays at one or a few chromosomal locations.

A considerable amount of work on ribosomal RNA and its genes (rDNA) has been reported and many experiments pioneering new techniques and ideas have arisen from this system. For example post-transcriptional processing of RNA was shown first for ribosomal RNA (see Perry, 1967), the first eukaryotic genes to be cloned in *E. coli* were those coding for ribosomal RNA from *Xenopus laevis* (Morrow *et al*, 1974) and the first reported intervening sequence was in the ribosomal genes of *Drosophila melanogaster* (Glover & Hogness, 1977).

The aim of this project was to characterise the region containing the 'internal transcribed spacers' in the ribosomal DNA of *X. laevis* (see figure 1.1 and below). To place this aim in context I shall start by summarising some general features of rRNA and rDNA in eukaryotes. In Chapter 2 I shall describe the specific objectives in more detail and discuss the experimental approach. For general reviews of rRNA and rDNA see for example Long and Dawid (1980a), Cox (1977), Hadjiolov and Nikolaev (1976) and Perry (1976).

1.1. Structure of ribosomal DNA

In eukaryotes 18S, 5.8S and 28S rRNAs are transcribed as part of a large precursor molecule (40S RNA in *Xenopus*, 45S RNA in mammals). The precursor also contains an external and two internal transcribed spacers, regions of RNA which are excised in an ordered series of events during the maturation of rRNA. The unit of DNA coding for the

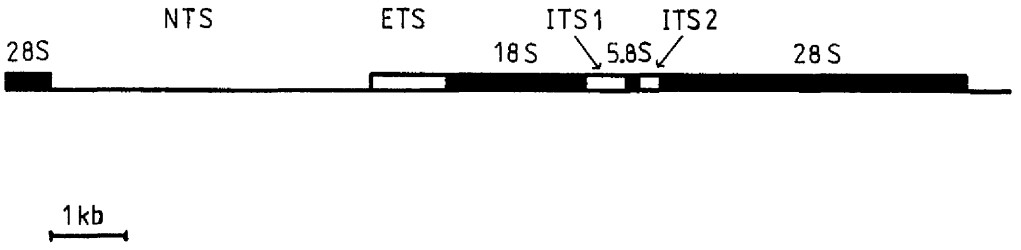


Figure 1.1. A unit of ribosomal DNA from Xenopus laevis. The transcription unit is represented with a thick line and the non-transcribed DNA with a thin line.

NTS - non-transcribed spacer

ETS - external transcribed spacer

ITS - internal transcribed spacer

rRNA precursor is tandemly repeated, and between each transcription unit lies non-transcribed spacer DNA. Xenopus laevis has about 600 rDNA repeats per haploid genome and other eukaryotes have numbers ranging from less than a hundred to several thousand repeats (see Long & Dawid, 1980a). The structure of a ribosomal DNA repeat is summarised in Fig 1.1.

Ribosomal DNA occurs at a single chromosomal location in Xenopus laevis, termed the nucleolar organiser (Pardue, 1973). The genes for 5S rRNA are separate and are found at the telomeres of several chromosomes (Pardue et al, 1973).

The general organisation of rDNA is similar for most eukaryotes, but there are differences in detail in both the transcribed and non-transcribed regions. The lengths of elements within the repeat unit are variable. For instance in Xenopus and plants the region which is transcribed is approximately 7.5 kb, whereas it is about 12 kb in mammals, and intermediate in birds. The difference reflects mainly an increase in the length of the transcribed spacers, but the 28S gene is also slightly longer in mammals and birds than in Xenopus (Loening et al, 1969; Schibler et al, 1975). Even greater differences are found in the non-transcribed spacers. Mouse, man and calf all have total rDNA repeat lengths of 30-40 kb (Arnheim & Southern, 1977; Meunier-Rotival et al, 1979). In contrast the repeat length in various plants is 8.6-9 kb (Goldsbrough & Cullis, 1981; Friedrich et al, 1979; Gerlach & Bedbrook, 1979). In Xenopus it is 11-15 kb (see 1.2). Taking the values for the length of the transcription unit given above, this represents a length variation in the non-transcribed spacer of more than an order of magnitude between diverse eukaryotes.

The ribosomal transcription unit of Drosophila has two special characteristics which should be mentioned, and which are also found in

some other invertebrates. The genes for 5.8S and 28S rRNA each contain a short 'spacer', which is transcribed and then cleaved during rRNA processing. This results in each RNA occurring in two non-covalently linked parts in the mature ribosome (Jordan, 1975; Jordan et al, 1976; Pavlakis et al, 1979). The second special feature is the presence of a large intervening sequence in some copies of the 28S gene (Glover & Hogness, 1977).

In yeast each rDNA repeat also contains a gene for 5S rRNA, but this occurs on the opposite strand and between the major transcription units (Rubin & Sulston, 1973; Aarstad & Oyen, 1975).

In some single-cell eukaryotes the bulk of the rDNA is extra-chromosomal and in several instances, eg, Dictyostelium and Tetrahymena, this DNA takes the form of giant palindromes (Cockburn et al, 1978; Karrer & Gall, 1976).

1.2. Chromosomal and amplified rDNA from X. laevis

Much of the work on Xenopus rDNA has been carried out using amplified DNA from oocytes. During oogenesis extensive amplification of ribosomal DNA takes place, and the amplified rDNA remains extra-chromosomal. The first phase of amplification occurs in both sexes during early germ cell proliferation and results in a 10-40 fold increase in rRNA gene number (Bird, 1977). The second, major, phase of amplification occurs during meiosis in oocytes but not in sperm. This second amplification, apparently involving a rolling-circle mechanism, brings the number of rDNA repeats to a few thousand times the chromosomal level (reviewed by Tobler, 1975). Amplified rDNA differs from chromosomal rDNA in its lack of methylated cytosines; it can thus be purified as a high density satellite after density gradient centrifugation of DNA from oocytes or ovaries (Dawid et al, 1970).

Amplified rDNA has been used in many studies, and most importantly in this context it was used in the construction of recombinant plasmids (see Table 3.1). It is therefore critical to know whether amplified rDNA is representative of chromosomal rDNA, and as part of this question whether any rDNA repeats are different from others.

Initial experiments by Birnstiel et al (1969) showed that the renaturation kinetics of somatic rDNA were consistent with a single repeating unit of a length corresponding to one set of ribosomal genes plus a fixed amount of non-ribosomal DNA. The degree of similarity between repeats was tested by Dawid et al (1970) as follows: amplified rDNA was melted, allowed to reanneal then melted again and the two melting profiles compared. The profiles were very similar, suggesting that when different copies anneal with each other the extent of mismatch is very low. (The DNA was obtained from ovaries from several frogs.) Wensink and Brown (1971) then compared amplified and chromosomal DNA by electron microscopy. Molecules were partially denatured to give a characteristic repeating pattern of double and single stranded regions (depending on local base composition), and repeat lengths were measured. By this criterion all repeat lengths in both types of DNA appeared to be the same, as did the characteristic denaturation pattern.

However more recently the use of restriction enzymes and gel electrophoresis has allowed repeat lengths to be measured and compared with greater accuracy and resolution. Wellauer et al (1974a) digested amplified rDNA with EcoR1 and separated the products by electrophoresis. They found one strong band of DNA with an estimated molecular weight of 3.0×10^6 and several bands of higher molecular weight and varying intensity. The combined molecular weights of all fragments was much too great to correspond to a single repeat unit (estimated from E.M.).

Electron microscopy of single stranded molecules derived from EcoR1 digested rDNA revealed that two types of fragment were present. The small fragment of 3.0×10^6 appeared to contain most of the 28S gene, by comparison to the appearance of 28S rRNA under similar spreading conditions. Each of the higher molecular weight fragments contained most of the 18S gene plus a variable length of non-transcribed spacer and a short section of the 28S gene. It was therefore concluded that each rDNA repeat contains two sites for EcoR1. The smaller fragment generated is homogeneous in size but the larger fragment, containing the non-transcribed spacer, is heterogeneous and several size classes were found (Wellauer et al, 1974a). Wellauer et al (1976a) then used four different cloned copies of the large EcoR1 fragment to analyse the molecular basis of length heterogeneity. The cloned rDNA was denatured and allowed to reanneal, either to its complementary strand as a homoduplex or to rDNA from a second plasmid as a heteroduplex. Examination of homoduplexes showed that a percentage of molecules reannealed imperfectly, usually leaving two single stranded loops of equal size. This structure could be explained if the DNA was internally repetitious, allowing a duplex to form with the repeats annealed out of register. Further information was gained by examination of heteroduplexes. The extra DNA from the longer molecule could be seen looped out in a variety of positions and could form either one large single-stranded loop or a few smaller loops separated by regions of duplex. All the patterns found could be explained by re-association between molecules which contain different numbers of internal repeats. Wellauer et al (1976a) were able to define two repetitive regions, both within the non-transcribed spacer, and one of these regions seemed to account for the length differences between the four plasmids. More recently restriction analysis and sequence determination have

confirmed that much of the non-transcribed spacer is repetitive. In the region which is most heterogeneous in length there appears to be a higher order repeating unit superimposed on the short repeats (Botchan et al, 1977; Boseley et al, 1979). This longer repeating unit is defined by sites for BamHl and the region around each Bam site is very similar in sequence to the region in which transcription of 40S rRNA is initiated (Moss & Birnstiel, 1979).

Wellauer et al (1976b) and Buongiorno-Nardelli et al (1977) both examined the ribosomal DNA from individual frogs. The chromosomal rDNA from each frog displayed several size classes at various frequencies. The pattern of bands seen after electrophoresis of EcoRl digested material was distinctive to each frog. Examination of amplified rDNA from the ovaries of the same frogs again revealed that several size classes were present. Except in rare cases the same size classes were present in both chromosomal and amplified rDNA from an individual, but the frequency distribution of different size classes was often altered. Bird (1977) examined rDNA from individual oocytes by Southern blotting, using cloned rDNA as a probe. He showed that each oocyte selectively amplified only one or a few size classes, but different oocytes selected different size classes.

We can conclude from these studies that, while the general structure of each ribosomal repeat appears to be the same, variations certainly exist within the non-transcribed spacer DNA. We cannot rule out the possibility that small differences may also occur within the transcribed regions. Secondly amplified rDNA does represent rDNA repeats that exist in the chromosome, but probably not all chromosomal repeats will be proportionately represented.

1.3. Comparison between related species

To gain some understanding of how ribosomal DNA has evolved it is useful to compare rDNA from related species and see what kinds of differences exist. In this context comparisons have been made between Xenopus laevis and Xenopus borealis (wrongly identified as X. mulleri in early studies - Bisbee et al, 1977). The ribosomal DNA from X. borealis is judged to consist of about 500 tandem repeats on a single chromosome. The repeats are apparently homogeneous (by E.M. measurements), and they are of similar length to X. laevis repeats (Brown et al, 1972). Mature ribosomal RNA from the two species is indistinguishable in size and base composition. A mixture of rRNA from the two species (one labelled with ^{3}H and the other with ^{32}P) was hybridised to rDNA immobilised on a filter. The proportion of each rRNA hybridised was the same regardless of whether the rDNA came from X. laevis or X. borealis. The rRNA's were then melted from the hybrid and again no significant difference was seen in the release of either rRNA from either rDNA. These results indicate that the two species have very similar rRNA sequences, and it was estimated that there was a maximum of 3% mismatch in the heterologous hybrids (Brown et al, 1972). 40S rRNA precursors were also compared from the two species; the precursors were of the same length but rDNA from each species preferentially hybridised precursor RNA from the same species. This suggests that there are differences between the precursors, which must be located in the transcribed spacer regions of the molecule (Brown et al, 1972).

This study was followed up by electron microscopy of hetero-duplex molecules formed by reassociation between rDNA from X. laevis and X. borealis (Forsheit et al, 1974). Perfectly duplexed regions could be seen which were of about the length expected for the 18S and 28S genes.

The region between these genes, corresponding to part or all of the internal transcribed spacer, remained single stranded. Similarly the external transcribed spacer was single stranded for part or all of its length. In the non-transcribed spacer a variable pattern of duplex regions and single stranded loops was seen. This could represent short regions of homology repeated in the non-transcribed spacer but interspersed with non-homologous regions. Undoubtedly the repetitive nature of the non-transcribed spacer makes these results more difficult to interpret.

These studies show that the sequences coding for mature rRNA have changed very little since the divergence of X. laevis and X. borealis. (Indeed, Ford and Mathieson (1978) directly compared the sequence of 5.8S from the two species and found only one base change; the X. borealis sequence was also found in less than 1% of X. laevis 5.8S molecules.) In contrast the sequences of both transcribed and non-transcribed spacers appear to have changed markedly.

The divergence of transcribed and non-transcribed spacers in two related species suggests that their exact sequence is not critical to gene function. However within each species there is not great heterogeneity between the sequences of elements in different rDNA repeats (since non-transcribed spacers of different lengths differ in number of subrepeats rather than sequence). During the evolution of this multigene family some mechanism must prevent the rDNA repeats from diverging and yet allow them all to evolve simultaneously.

1.4. Further comparative studies on the transcription unit

A variety of methods has been used to compare the rRNA transcription unit over a wide range of species. Khan et al (1978) examined the sequences around methylated nucleotides in rRNA from human, hamster,

mouse, chicken and Xenopus. By comparing the fingerprints of methyl-labelled rRNA digested with T1 ribonuclease they estimated that the oligonucleotides were at least 95% homologous. However the sequences around methylated nucleotides might be more conserved than in other parts of the rRNA molecule.

Gerbi (1976) hybridised rRNA from six species, Gyrodinium cohnii, Drosophila hydei, Chironimus thummi, Sciara coprophila, Xenopus laevis and mouse, to heterologous rDNA. In general the extent of hybridisation reflected evolutionary distance, though some anomalies occurred. Hybridisation of rRNA to heterologous rDNA was also performed in the presence of unlabelled rRNA from a third species. It was observed that the unlabelled rRNA was in all cases able to compete for sites on the rDNA. Thus certain regions of the rRNA must be conserved in all the species tested. More recently Cox and Thompson (1980) hybridised 28S rRNA from X. laevis or 25S rRNA from Neurospora crassa to specific restriction fragments of rDNA from the two species. Conserved sequences were distributed through most of the 28S (25S) gene. Ribosomal DNA fragments with sequences common to the two species also hybridised with 28S rRNA from Drosophila melanogaster, again suggesting that certain regions are highly conserved in a wide range of organisms.

In recent years 5.8S rRNA from a number of species has been sequenced (see compilation by Erdmann, 1981). The vertebrate sequences are remarkably similar to each other but are only about 70% homologous to yeast, Neurospora, Vicia fabia (bean) and Drosophila (5.8S + 2S). Again it seems that certain parts of the molecule are more strongly conserved than others (Pavlakis et al, 1979).

Gourse and Gerbi (1980) have examined the whole transcription unit for highly conserved regions. They used cloned rDNA from yeast

(S. cerevisiae), Dictyostelium discoideum and E. coli to hybridise to specific X. laevis rDNA restriction fragments, including fragments from the transcribed spacers. Regions at the 3' end of both the 18S and 28S genes showed hybridisation between E. coli and Xenopus laevis. Other regions of the X. laevis genes hybridised strongly to yeast (and some also to D. discoideum) but not to E. coli. In the transcribed spacer regions no hybridisation was found between X. laevis and any of the heterologous rDNAs.

A particular limitation of hybridisation studies is that fairly long stretches of exact complementarity are required to give a positive result under stringent conditions (probably about 20 base pairs). This limitation is demonstrated by the lack of hybridisation between the 5.8S genes in the work of Gourse and Gerbi (1980). It would not be possible to detect short regions of homology such as putative control elements by this means.

A second problem in all these studies is that we are looking solely at primary sequence conservation, yet particular nucleic acid secondary structures almost certainly have a role in interactions with protein. Secondary structures can sometimes be inferred from a knowledge of the nucleotide sequence, but a change in sequence need not necessarily result in a corresponding change in secondary structure. This topic will be further discussed later, especially in relation to sequence comparisons arising from this project and other recently published work. However one study in which secondary structures were compared directly is that of Schibler et al (1975). They examined by electron microscopy partially denatured rRNA precursor molecules from a variety of vertebrates. Under the conditions used 18S rRNA showed no secondary structure but the 28S region of the molecule had characteristic loop structures. Two distinct multiple loop structures were conserved in 28S rRNA of all the species

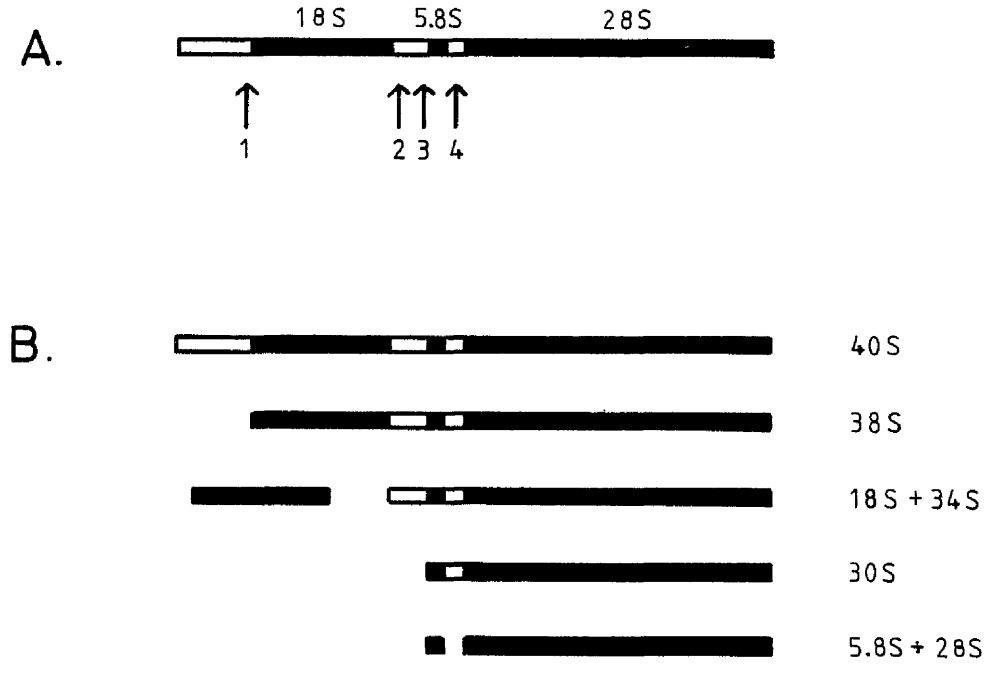


Figure 1.2. A. Approximate location of cleavages which take place during processing of precursor rRNA in eukaryotes. Cleavages sites are numbered from the 5' end. The order of cleavage can vary according to species and conditions.

B. Putative processing pathway in Xenopus laevis (Wellauer & Dawid, 1974).

studied. Characteristic secondary structures were also seen in the transcribed spacers; the position of structures in the internal transcribed spacer was comparable in all species but the size and complexity tended to increase with an increase in spacer length in the higher vertebrates.

1.5. Processing of the rRNA transcript

Ribosomal RNA is transcribed in the nucleolus by RNA polymerase 1 (Chambon, 1975). A number of processing steps then take place. Many nucleotides are specifically modified either by addition of a methyl group or by conversion of uridine to pseudouridine. All of these modifications seem to occur on nucleotides which are retained in mature rRNA even though most take place on the intact precursor molecule (for review see Maden et al, 1977). The second step is the elimination of the transcribed spacer RNA. This seems to involve at least four endonucleolytic cleavages, perhaps followed by exonuclease trimming. The approximate sites of cleavage are indicated in Fig. 1.2A.

The most detailed work on rRNA processing has been on vertebrate cells and yeast, but other eukaryotes examined appear to have similar pathways. Early kinetic studies showed that radioactive RNA precursors are incorporated first into 45S RNA in HeLa cells, and only later does radioactivity appear in 18S and 32S RNA, and later still in 28S rRNA (Penman, 1966). Thus it was concluded that ribosomal RNA is initially synthesised as a large precursor molecule which then has to be processed into mature rRNA.

Weinberg and Penman (1970) went on to analyse the processing pathway in greater detail. The use of polyacrylamide gels allowed the resolution of additional intermediate molecules and it was found that substantial amounts of minor intermediates were accumulated after infection of HeLa cells with poliovirus. Each intermediate was

characterised with respect to size and methyl content (from ratio of methyl label to phosphate label). Since methyl groups are conserved during the processing of 45S to 18S and 28S it was possible to deduce the nature of each precursor and to suggest a likely processing pathway. 45S contains as many methyl groups as 28S and 18S together. The second largest precursor (41S) has the same number of methyl groups but is smaller, so must have resulted from the excision of transcribed spacer RNA (ie, cleavage 1 in Fig. 1.2A). 32S has the same number of methyl groups as 28S rRNA but is longer, and 20S has (roughly) the same number of methyl groups as 18S but is again longer. The combined size of 32S and 20S is close to the size of 41S so these molecules could be generated by a single cleavage (3 in Fig. 1.2A). Subsequently cleavage 2 would generate 18S rRNA and cleavage at site 4 would generate 28S and 5.8S rRNAs. (Site 4 probably corresponds to two cleavages.) Two other intermediates were also found which could not fit this pathway and were judged to be the result of aberrant processing events (but see later).

Wellauer et al (1974b) examined the size and secondary structure of rRNA from mouse L-cells by electron microscopy. Under conditions of partial denaturation characteristic loop structures can be seen which allow the positions of 28S and 18S rRNA to be located in precursor molecules. Two large precursors were seen which both contain 28S and 18S but differ in that the smaller one does not have the external transcribed spacer. These are analogous to 45S and 41S in HeLa cells. No molecule corresponding to 20S was seen in mouse cells, but a 32S molecule was seen and also a 36S molecule. The 36S molecule contained the same elements as 32S but had additional internal transcribed spacer RNA. It could thus be deduced that in mouse cells cleavage at position 2 must occur before cleavage at position 3, generating 18S plus 36S instead of 20S plus 32S

as in HeLa cells. 36S would then be cleaved to generate 32S and subsequently 28S plus 5.8S. Wellauer and Dawid (1974) also examined Xenopus rRNA precursors by electron microscopy and concluded that cleavages occur in the same order as in mouse cell rRNA processing although the precursors were shorter (due to shorter transcribed spacers). The putative pathway for processing of Xenopus rRNA is shown in Fig. 1.2B.

More recently there have been several reports which have suggested that the order of processing is not as rigid as had been supposed. Winicov (1976) reported a line of hamster cells which was temperature sensitive in the cleavage of the 32S precursor. At the non-permissive temperature the sizes of precursors which could be isolated suggested that the predominant order of earlier cleavages was also altered. Similarly it is possible that the 'aberrant' precursors found by Weinberg and Penman (1970) in HeLa cells represent alternative processing pathways which could still yield mature rRNAs. A detailed kinetic study of processing in rat liver suggests that cleavages can occur in almost any order, but that some pathways are favoured. In a given precursor molecule it was possible to define the probability for each cleavage site that that site would be the next one cut (Dudov et al, 1978). It has also been proposed that rRNA processing in Drosophila can occur by two alternate pathways (Long & Dawid, 1980b). The technique used was to electrophorese RNA, transfer to DBM paper (diazobenzyloxymethyl-paper), then probe with specific rDNA restriction fragments to define which sequences were contained in a particular RNA molecule. Precursors were found which were apparently analogous to Xenopus precursors, but in addition a molecule containing 18S and both external and internal transcribed spacer RNA was detected. This could represent a processing pathway in which cleavage occurred at (or near) site 3 first, and only later at sites 1 and 2 (see Fig 1.2A). A novel cleavage site between sites 2 and 3 was also predicted in this study.

In summary we can conclude that the general processing steps are probably quite similar in all eukaryotes but that the order in which cleavage events occur can be flexible, with different pathways predominating in different situations. One of the objectives of this project was to locate more precisely the sites at which some of these cleavages occur in Xenopus laevis, having first determined the sequence of the DNA. This approach has also been used in yeast as I will describe in the discussion (Veldman et al, 1980a).

It should be added that ribosomal RNA is associated with proteins throughout its life-time. These will include processing enzymes as well as proteins of the mature ribosome. It has been shown that rRNA processing is dependent on protein synthesis. Addition of cycloheximide or starvation for valine both cause a rapid decrease in the rate of 45S processing (Willems et al, 1969; Maden et al, 1969). One interpretation is that processing enzymes may only correctly recognise RNA which is already associated with ribosomal protein in a precursor particle.

CHAPTER 2. OBJECTIVES AND EXPERIMENTAL APPROACH

2.1 Objectives

At the outset of this project I hoped to fully characterise the region of rDNA from Xenopus laevis which lies between the 18S and 28S genes. This region contains the two internal transcribed spacers, separated by the gene for 5.8S rRNA. The internal transcribed spacers are excised from the 40S precursor molecule by specific cleavages during rRNA maturation (see 1.5). The sites at which processing occurs must be correctly recognised by ribonucleases, presumably by virtue of either their primary sequence, secondary structure, tertiary structure, interaction with other proteins, or some combination of these features.

One way in which it may be possible to define critical characteristics of processing sites is to look for similarities between analogous sites. Relevant data could be obtained from within one species, for example by comparing characteristics around the 5' end of each mature rRNA in Xenopus laevis, or between different species, as in comparing the regions around the 5' end of 5.8S rRNA from several species. This approach has been used in examining features of the sequence which are necessary for splicing out introns from messenger RNA and a 'consensus' sequence has been defined (Chambon, 1981).

Hybridisation studies indicate that the overall nucleotide sequence of transcribed spacers in rDNA is not conserved through evolution (see 1.3 and 1.4). However short conserved sequences would not be detected in such studies. Furthermore the possible conservation of more complex structural features cannot be analysed by hybridisation.

In this project my approach has been to determine the nucleotide sequence of the internal transcribed spacers from Xenopus laevis and

to use this information in attempting to characterise processing sites in Xenopus, and as a basis for evolutionary comparisons.

Before the start of this project Boseley et al (1978) had published a detailed restriction map of the ITS region of rDNA and had sequenced the 5.8S gene and surrounding area. I have followed up their work by determining the complete nucleotide sequence of a 1,300 bp segment of DNA starting at the EcoRI site near the 3' end of the 18S gene and extending downstream into the 28S gene. To define the internal transcribed spacers I located the mature rRNA termini within the sequence; for 18S and 5.8S this was done using information from published rRNA sequences but for the 5' end of 28S further experiments were necessary. Radioactive 28S rRNA was hybridised to a small rDNA fragment containing the 5' end of the gene and the rRNA complementary to the fragment was analysed by fingerprinting.

Determination of the sequence of the 18S-28S intergene region has made it possible to examine the potential for RNA secondary structure formation within this region. A search was made for sequences which could base pair with each other in 'hairpin' structures. Several theoretically stable structures could be formed, especially in the central part of each internal transcribed spacer.

Recently the sequence of yeast ITS 1 has been published (Skryabin et al, 1979 and Veldman et al, 1980a), and also short sequences from the ITS 1 of some other eukaryotes (see Chapter 5). These studies enable one to look for features of the ITS which may be conserved in evolution. At the same time the gene sequences determined in this work can be compared to other published sequences and it is interesting to contrast the evolution of spacers with that of the genes.

The boundaries between gene and transcribed spacer define the

sites at which the final cleavages in rRNA processing take place to yield mature rRNA. It would also be of interest to find out where any earlier cleavages take place, by locating the ends of precursor molecules within the DNA sequence. One approach that I used was to hybridise 'nuclear' RNA from Xenopus cells to end-labelled rDNA from the ITS region and digest with S1 nuclease (single strand specific). The length of DNA protected from digestion (because it forms a duplex with RNA) will define the end of the RNA molecule. Theoretically this method can map an RNA terminus to within a few nucleotides. A second type of experiment used 'nuclear' RNA electrophoresed in an agarose gel and blotted onto DBM-paper ('Northern blotting'). Labelled DNA fragments from the ITS were then used to probe for precursor molecules. It should be possible with this technique to assess how many precursor species are present (above some threshold concentration) and to estimate broadly what part of the ITS is present in each type of molecule. It was concluded from these studies that one precursor exists containing 5.8S and probably 28S rRNA, and having the same 5' end as 5.8S rRNA. However several problems arose in the use of both techniques and other precursors could not be characterised conclusively.

Thus in the region sequenced there are still only four known processing sites. Taken together with other published data from Xenopus laevis and other eukaryotes extensive comparisons can be made of processing sites within and between species. However of the four criteria mentioned earlier only primary sequence, and to a limited extent second structure, can yet be compared in detail in a search for critical features of processing sites. No outstanding similarities have been revealed. It is hoped that the work reported here will serve as a basis for any future structural studies, as well as for more extensive sequence comparisons when data from other species becomes available.

A further aspect of the experimental work in this project was to assess the variability of internal transcribed spacers within Xenopus laevis. As discussed in 1.2 minor heterogeneity within the transcription unit would not have been detected in earlier experiments. If such variability does exist it could give some insight into the evolution of rDNA, both in terms of what part of the sequence is least critical to gene function and of how mutations can be propagated and spread in a multigene family. Several plasmids containing Xenopus rDNA are available (see Table 3.1). One of these was chosen for sequence analysis (2.2) and three other plasmids were selected for comparative work. Restriction mapping of the four plasmids revealed that sites for SmaI (CCCGGG) within ITS 1 are not in identical positions, though digestion with other restriction enzymes did not show such diversity. This comparison is currently being followed up by M.A. Stewart in this laboratory, by sequence analysis of the internal transcribed spacers from several rDNA plasmids.

2.2 Use of cloned ribosomal DNA

In order to determine the nucleotide sequence of the internal transcribed spacers it was obviously best to use cloned ribosomal DNA. Although pure amplified rDNA can be prepared from Xenopus oocytes, separated from chromosomal DNA by density gradient centrifugation, it would be hard work to make sufficient for extensive DNA sequencing. Furthermore any sequence variation within the rDNA would make the task impossible.

A large number of plasmids containing Xenopus ribosomal DNA have been constructed (Table 3.1). In this laboratory we chose to use pX1r101 for sequencing purposes. It contains a complete rDNA repeat

cleaved with HindIII (for which there is a single site at the extreme 3' end of the 28S gene). Therefore, unlike plasmids constructed from EcoRI digested rDNA, pX1r101 represents a single complete copy of the ribosomal transcription unit. The region of pX1r101 containing the internal transcribed spacers was subcloned into pBR322 to simplify the preparation of specific restriction fragments, as described in Chapters 3 and 4.

A detailed restriction map of the internal transcribed spacer region and beyond was available at the start of this project (Boseley et al, 1978) and was of great assistance throughout. The map was constructed from work on a different rDNA plasmid, pXL212 (Table 3.1).

If cloned ribosomal DNA is to be used it is important that the cloned copy is representative of bulk rDNA. However it is difficult to establish directly that this is the case. This work and other experiments in this laboratory have not revealed any anomalies between gene sequences from pX1r101 and either other rDNA copies (cloned or uncloned amplified rDNA) or oligonucleotide sequences from mature rRNA.

2.3 Choice of method for DNA sequencing

In late 1978 there were two methods for sequencing DNA from which to choose. Firstly there was the chemical method of Maxam and Gilbert (1977) which was well established and was in use in many laboratories. An updated protocol with more specific reactions for purines also became available and has now been published (Maxam and Gilbert, 1980). This protocol is extremely detailed and explains practical aspects of the procedures very thoroughly.

The second method was that of Sanger et al (1977) which involves enzymatic primer extension in the presence of chain-terminating

inhibitors such as dideoxynucleotides. This superseded earlier primer extension methods such as the 'plus and minus' method (Sanger and Coulson, 1975).

Several factors must be taken into account when deciding which method to use. Consider the critical stages in each method. For Maxam and Gilbert the starting material is an end-labelled restriction fragment; the ends must be separated either with a second restriction enzyme or by strand separation. Four separate modification reactions are then performed, each followed by strand scission. Finally the products are resolved on a polyacrylamide gel. For the Sanger method the starting material must be a single-stranded template. A restriction fragment is then annealed to the template as a primer, and the primer is extended by DNA polymerase 1 (Klenow fragment) in four separate reactions. Each reaction mixture contains a mixture of dNTPs, one or more being ^{32}P labelled at the α-P position, and a proportion of one dideoxy-nucleotide triphosphate. Again the products are resolved on a poly-acrylamide gel.

In each method the length of sequence which can be read depends on the resolution of the polyacrylamide gel so should in theory be the same for both. 'Piling-up' of bands in the gel due to secondary structure can occur in either system. However a second problem due to secondary structure which may occur in the primer extension method is that during primer extension the polymerase may not read through a double-stranded loop. Furthermore the Sanger method requires single-stranded template DNA whereas the chemical method may be done on double- or single-stranded DNA. On the other hand the chemical reactions take longer and require more separate manipulations than the primer extension reactions.

Of these considerations by far the most critical is the require-

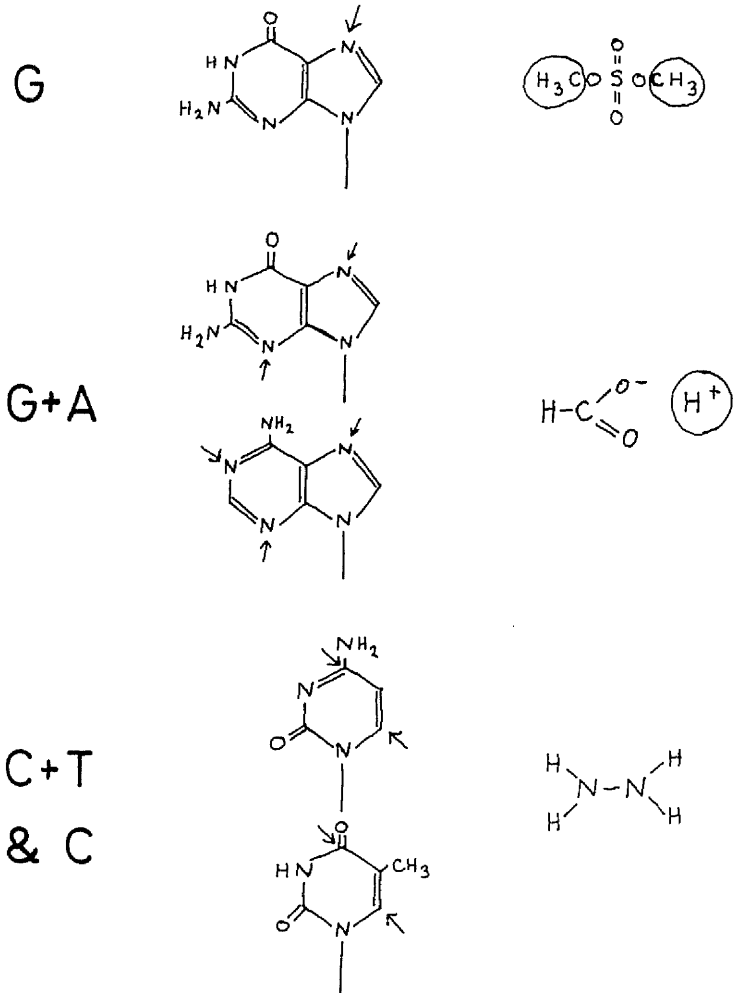


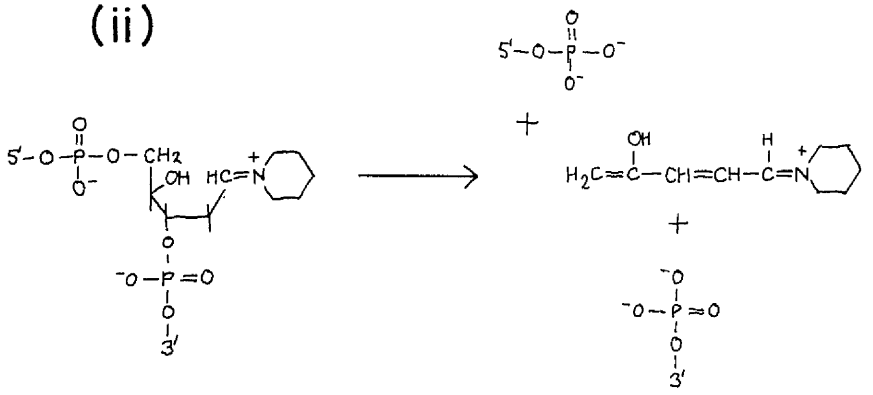
Figure 2.1 DNA sequencing reactions (Maxam & Gilbert, 1980)

(i) Base-specific modification. In each reaction a specific base is modified and subsequently displaced from its sugar.

G. Dimethylsulphate methylates guanine at N7. Addition of piperidine causes the ring to open between N7 and C8 in a base-catalysed reaction, and piperidine then displaces the ring-opened guanine.

G+A. Pyridinium formate protonates purine ring nitrogens, weakening the glycosidic bond and again

(ii)



piperidine displaces the modified purine.

C+T & C. Hydrazine attacks pyrimidines at C6 and C4 and recyclises with C4, C5 and C6, but further hydrazinolysis releases the newly formed ring. Again the modified bases are displaced by reaction with piperidine. In the presence of salt the reaction with thymine is inhibited giving a 'C only' reaction.

(ii) Strand scission. In each reaction after the modified base has been displaced piperidine proceeds to catalyse the β-elimination of both phosphates from the sugar, possibly by the route shown above.

ment for a single-stranded template in the Sanger method. Except in the case of some viruses the best method of producing single-stranded templates was probably by partial digestion with exonuclease III (Smith, 1979), but this was not a very well characterised procedure.

I finally chose to use the method of Maxam and Gilbert for two reasons. Firstly it could be carried out on double as well as single-stranded DNA. Secondly, but perhaps just as important, much more information was available on the chemical method from the detailed protocol and the method was in more general use. The reactions are summarised in figure 2.1.

I should add that since the start of this project the phage M13 has come into general use as a cloning vector. Cloning in M13 not only produces a single-stranded template but also alleviates the need to make lots of specific primers because a single primer complementary to the vector can be used (Anderson et al, 1980). For large sequencing objectives primer extension might now be the method of choice.

CHAPTER 3 MATERIALS AND METHODS

Common chemicals used were Analar grade, obtained from BDH Chemicals, Poole, Dorset or Fisons Scientific Apparatus, Loughborough, Leics. The suppliers of special chemicals, enzymes and other materials are stated in the text.

3.1 Maintenance of plasmids

3.1.a Plasmids

The plasmids used in this study are listed and described in Table 3.1 and Figure 3.1. The E. coli host is HB101 : K12, F^-, pro^-, leu^-, thi^-, lac Y^-, hsd R^-, end A^-, rec A^-, rps L20, ara-14, gal K2, xyl-5, mtl-1, sup E44 (Bolivar & Backman, 1979). Plasmids are stored both as naked DNA in 10 mM Tris HCl, 0.1 mM EDTA, pH 8.0, and in cultures of transformed HB101.

3.1.b Growth and storage of E. coli transformed with plasmids

L-broth: 1% tryptone (Difco, West Molesey, Surrey), 0.5% yeast extract (Difco), 0.5% NaCl, 10 mM Tris HCl pH 7.4, 1 mM $MgSO_4$.

M9 medium: 0.1% NH_4Cl, 0.013% $MgSO_4.7H_2O$, 0.3% KH_2PO_4, 0.6% Na_2HPO_4, 0.5% Casamino acids (Difco), 0.4% glucose, 2 μg/ml thiamine (Vitamin B1, Sigma, Poole, Dorset).

Autoclave L-broth and each organic component of M9 medium at 5 psi for 50 minutes, inorganic salts from M9 medium at 15 psi for 25 minutes (volumes of less than 500 ml, increase time for larger volumes).

Inoculate 10 ml of L-broth or M9 medium with transformed bacteria and incubate overnight at 37°C. Measure the absorbance at 650 nm. Use the overnight culture to inoculate fresh medium to an absorbance of 0.05 and incubate this until the absorbance reaches 0.5.

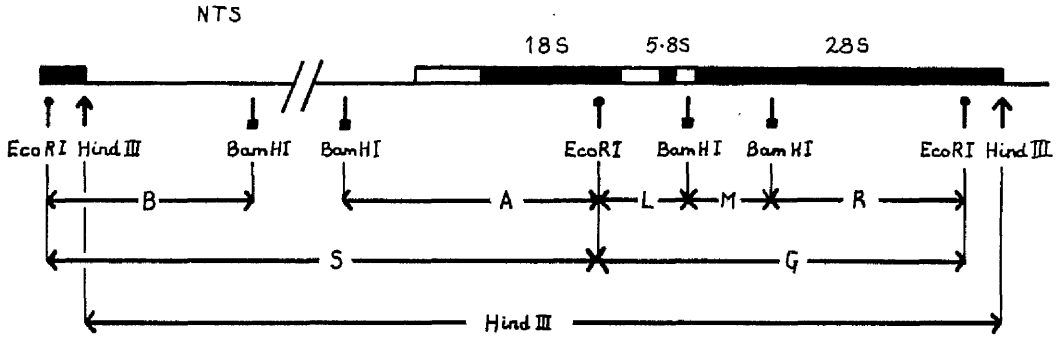


Figure 3.1 Map of restriction sites used to clone X.laevis rDNA. Restriction fragments which have been cloned are shown, see also Table 3.1.

Table 3.1. Plasmids containing ribosomal DNA from Xenopus laevis (or Xenopus borealis) used in this study.

Name	rDNA fragment+	Vector	Antibiotic resistance or other marker	Origin/Reference
pXlr 101	Hind III	pMB9	Colicin E1 immunity	cloned by R.H.Reeder from amplified X. laevis rDNA
pXlr 102	Hind III	pMB9	Colicin E1 immunity	
pXlr 103	Hind III	pMB9	Colicin E1 immunity	
pXlr 101L	L	pBR322	Ampicillin resistance	subcloned from above plasmids, this study
pXlr 101M	M	pBR322	Ampicillin resistance	
pXlr 102L	L	pBR322	Ampicillin resistance	
pXlr 103L	L	pBR322	Ampicillin resistance	
pXlr 11	G	Col E1	Colicin E1 immunity	Recloned (Dawid & Wellauer 1976) from CD18 (Morrow et al, 1974)
pXlr 11L	L	pBR322	Ampicillin resistance	subcloned from pXlr 11 by B.E.H. Maden
pXlr 11M	M	pBR322.	Ampicillin resistance	
pXlr 11R	R	pBR322	Ampicillin resistance	
pXlr 14	S	Col E1	Colicin E1 immunity	Recloned (Botchan et al, 1977) from pXlr 4 (Wellauer et al, 1976a)
pXlr 14B	B	pBR322	Ampicillin resistance	Subcloned from pXlr 14 by B. Sollner-Webb
pXL 212	G	pCR1	Kanomycin resistance	Subcloned (Boseley et al, 1978) from CD4 (Morrow et al, 1974)
pXL212L	L	pBR322	Ampicillin resistance	Subcloned from pXL212 by M. Stewart
*pXbr 101	Hind III	pMB9	Colicin E1 immunity	Cloned by R.H.Reeder from X.borealis rDNA
*pXbr101LM	L + M	pBR322	Ampicillin resistance	Subcloned from pXbr 101, this study

*ribosomal DNA from Xenopus borealis not X. laevis.

+ribosomal DNA fragments as shown in Fig 3.1.

Add 2 ml of the growing culture to 3 ml of sterile 80% glycerol and store at -20°C. This stock culture can be kept for many months.

3.1.c Analysis of antibiotic resistance

To check for antibiotic resistance markers conferred by the transforming plasmid test bacteria on nutrient agar plates containing the appropriate antibiotic. Nutrient agar contains L-broth (as above) with 15 g/l Difco 'Bacto' Agar, autoclaved then cooled to 65°C. Add antibiotics just before pouring plates (0.1 mg/ml ampicillin or 0.015 mg/ml Tetracycline, both from Sigma). Incubate the plates at 37°C overnight to remove excess moisture before use.

Check a liquid culture by streaking with a platinum loop onto appropriate antibiotic plates. Check colonies (eg from a transformation experiment) by picking with a wire or sterile cocktail stick onto plates marked with a grid. Incubate all plates upside down at 37°C for one or two days before analysis.

3.2 Preparation of plasmid DNA

3.2.a Cleared lysis

Derived from the method of Clewell and Helinski (1970). The procedure here is for a 3 litre culture (2 x 1.5 l or 4 x 750 l) of a single plasmid but can be adjusted for preparations of different volumes.

Brij solution : 1% Brij 58 (polyoxyethylene 20 cetyl ether, Sigma, Poole, Dorset), 0.4% sodium deoxycholate, 2.5 mM EDTA, 50 mM Tris HCl, pH 8.0.

Sucrose solution : 25% sucrose in 50 mM Tris HCl pH 8.0.

Lysozyme solution : 5 mg/ml lysozyme (Sigma) in 0.25 M Tris HCl pH 8.0, freshly made.

Inoculate 100 ml of L-broth or M9 medium with 0.1 ml from a 50% glycerol stock culture and incubate overnight at 37°C. Read

absorbance (650 nm) and add to 2 x 1.5 l (or 4 x 750 ml) M9 medium in 5 l (or 2 l) flasks so that the absorbance of the large cultures is 0.03. Stir at 37°C until absorbance reaches 0.5-0.8, add 300 mg (or 150 mg) chloramphenicol to each flask and stir at 37°C overnight.

Centrifuge cells at approximately 5,000 g for 10 minutes at 4°C then pour off supernatant. Resuspend in a total volume of 150 ml 10 mM Tris HCl, 1 mM EDTA, pH 8.0, combining pellets from the two (or four) flasks. Centrifuge a second time and discard the supernatant, keeping the pellet on ice. Resuspend in 25 ml cold sucrose solution, add 5 ml cold lysozyme solution, and swirl on ice for 5 minutes. (After addition of lysozyme it is essential to keep the mixture on ice to minimise contaminating DNase activity.) Add 10 ml 0.25 M EDTA, swirl on ice for 5 minutes then add 45 ml Brij solution and swirl on ice for a further 10 minutes. Centrifuge at 30,000 rpm for 45 minutes at 4°C in a Beckman 60 Ti rotor (90,720 g). Collect the supernatant (the 'cleared lysate') which contains the plasmid DNA and keep on ice, leaving behind the glutinous pellet which contains the lysed bacteria and most of the chromosomal DNA.

3.2.b Caesium chloride centrifugation

This step separates closed circular DNA from linear or open circular DNA on the basis that supercoiled DNA binds less ethidium bromide and so has a greater density than relaxed DNA in concentrated solutions of ethidium bromide (Bauer & Vinograd, 1968).

Carefully measure the volume of supernatant, add 1.03 g of caesium chloride per ml of supernatant and stir thoroughly to dissolve. Wrap in foil to keep out light, then add 10 µl of 10 mg/ml ethidium bromide per ml of supernatant (volume before adding CsCl). Adjust refractive index of solution to between 1.39989 and 1.39999. This value

was determined empirically to provide optimum density; components such as sucrose in the cleared lysis buffer alter the refractive index relative to that of a pure CsCl solution of known density.

Transfer to quick-seal polyallomer tubes (Beckman No. 342414, 1" x 3½") which must be filled completely, if necessary topping up with 1.03 g/ml CsCl solution. Seal tubes and centrifuge for 16 hours in a Beckman 50 V Ti rotor at 50,000 rpm (241,200 g) at 20°C. Alternatively use screw cap polycarbonate tubes (approximately two-thirds full) and centrifuge in a Beckman 60 Ti at 30,000 rpm (90,720 g) for 48 hours. After centrifugation visualise nucleic acid bands by U.V. illumination and collect lower, closed circular, band. This is done by carefully removing material from the top of the gradient with a pasteur pipette until most of the upper (sticky) band has been removed, then the lower band can be recovered directly with a clean pasteur pipette. Further purify the closed circular DNA by centrifuging in a Beckman 50 Ti at 33,000 rpm (98,550 g) for 24-48 hours and again collect the lower band. Avoid prolonged exposure of the DNA/ethidium bromide mixture to light.

3.2.c Recovery of DNA

To extract ethidium bromide from the DNA add an equal volume of isopropanol saturated with H_2O and CsCl, vortex, allow the phases to separate and discard the upper, isopropanol, phase. Repeat the extraction three times (until the colour has completely disappeared). Dilute the lower phase with 2 volumes of water and add a further 2 volumes (2 x total) of ethanol. Chill at -20°C for a few hours to precipitate the DNA. Centrifuge at 2,500 g for 30 minutes at -10°C then pour off the supernatant. Redissolve DNA precipitate in 5 ml of 10 mM Tris HCl, 0.1 mM EDTA, pH 8.0. Add 5 ml water-saturated phenol, vortex for 3-4 minutes, then separate phases by low speed centrifugation. Recover the upper (aqueous) phase, add sodium acetate to 0.3 M and

2.5 volumes ethanol. Again chill to precipitate DNA at -20°C overnight or at -70°C in a dry ice/methylated spirits bath for 1 hour, and recover by centrifugation as before. Wash pellet twice with 70% aqueous ethanol, centrifuging each time to ensure complete recovery. After the second wash pour away supernatant and dry off remaining ethanol under vacuum.

Dissolve the DNA in 2 ml 10 mM Tris HCl, 0.1 mM EDTA, pH 8.0. Read $O.D._{260}$ of 10 µl DNA in 1 ml of H_2O and calculate DNA concentration (1.0 $O.D._{260}$ unit is equivalent to 50 µg DNA per ml). Store DNA at 4°C over 0.5 ml chloroform in a screw cap tube. DNA is stable under these conditions for many months or even years. A small aliquot should also be stored at -20°C as a stock to transform fresh cultures of bacteria if necessary.

3.3 Construction of subclones

3.3.a Ligation

Prepare the restriction fragment to be cloned and the appropriate linear form of the plasmid vector by gel electrophoresis as described in section 3.4 and 3.5. (It is important to eliminate circular forms of the plasmid as these will transform with a high efficiency, and will be selected during subsequent antibiotic selection.) Use 5-10 times molar excess of 'insert' over plasmid. Dissolve DNA in 10 µl 10 mM Tris HCl, 0.1 mM EDTA, pH 8.0, heat to 50°C for 10 minutes then chill on ice. Bring the reaction mix to 30 µl total with a final concentration of 50 mM Tris HCl pH 7.6, 8 mM $MgCl_2$, 0.066 mM ATP, 10 mM dithiothreitol, then add 0.6 µl T_4 DNA ligase (Bethesda Research Laboratories, UK distributors Uniscience, Cambridge). Incubate the reaction mix at 15°C for 1 hour. Dilute the DNA with 10 volumes of reaction buffer (as above) and incubate at 15°C overnight. The rationale for this procedure is that

the initial concentrated reaction should favour ligation between molecules while subsequent dilution favours circularisation of linear molecules. However it is probably not necessary to use two different concentrations and various simpler modifications of the procedure are now used in many laboratories. For theoretical discussion see Dugaiczyk et al (1975).

3.3.b Transformation and selection

Grow an overnight culture of HB101 in L-broth (see 3.1). Inoculate fresh L-broth to 1% with the overnight culture and grow at 37°C until $O.D._{650}$ reaches 0.5. Pipette 2 x 20 ml of cells into 30 ml corex tubes, stand them on ice for 5 minutes, centrifuge at 2,500 g for 10 minutes then pour off the supernatant. Resuspend cells in 10 ml of ice cold 10 mM $MgSO_4$, maintain at 0°C for 30 minutes and again pellet cells. Resuspend in 10 ml 30 mM $CaCl_2$, hold on ice for a further 15 minutes, centrifuge once more and resuspend in 1.5 ml 30 mM $CaCl_2$. The cells are now ready for transformation (and can be stored as transformation competent cells at -20°C in 20% glycerol).

Bring the ligation mix to 0.03 M $CaCl_2$, transfer to a 10 ml tube, then add 0.7 ml of the prepared cells. Stand this mixture on ice for 15 minutes, then 37°C for 5 minutes, then room temperature for 10 minutes. Add 2.5 ml L-broth and shake at 37°C for 1 hour; this is necessary presumably to give time for expression of the transforming DNA (Cohen et al, 1972). Mix 1 ml of transformed cells at a range of concentrations (1 - 1/100) with 2.5 ml of soft agar (7.5 g/l agar kept molten at 37°C) and spread on appropriate antibiotic-containing agar plates to select for transformed cells (see 3.1). Check probable transformants on both ampicillin and tetracycline plates. Small scale 'cleared lysis' preparations of plasmid DNA can be made from several transformants to check size and restriction patterns (quicker

methods of preparing DNA on a small scale are now available, eg Birnboim & Doly, 1979).

3.4 Analysis of plasmid DNA

3.4.a Digestion with restriction enzymes

Restriction enzymes were obtained from New England Biolabs (U.K. distributors: CP Laboratories, Bishops Stortford, Herts), Bethesda Research Laboratories (Uniscience, Cambridge) or Boehringer Mannheim (Boehringer Corp. London, Lewes, East Sussex). Sma 1 was kindly given by B. Sollner Webb and Pst 1 by R. Peterson, both of Carnegie Institution of Washington, Baltimore.

Digestion buffers used are similar to those recommended by the suppliers. I have constructed a set of three buffers compatible with the requirements of almost all the enzymes used, see Table 3.2.

Volumes, quantities of enzyme, and time of digestion depend on the requirements of the particular experiment. For instance for an analytical experiment use enough enzyme to digest the DNA in one hour, but for large scale preparation use 1/10 as much enzyme (proportional to DNA) and digest overnight.

3.4.b Electrophoresis of DNA

Agarose Electrophoresis Buffer (1x) : 40 mM Tris HCl, 20 mM sodium acetate, 2 mM EDTA, pH 8.0.

Agarose loading Buffer : 1/5 electrophoresis buffer (as above), 2 M sucrose, few grains bromophenol blue.

1% Agarose tube gels. Dissolve 1 g agarose (Type II, Sigma, Poole, Dorset) in 100 ml agarose electrophoresis buffer by heating in a boiling water bath. When dissolved cool to around 65°C. Meanwhile prepare perspex tubes (10 cm x 0.6 cm, inner diameter) by covering the

Table 3.2 Set of buffers suitable for most restriction enzymes

No Salt Buffer (10x)		Enzymes
Tris HCl	100 mM	BglII, HaeII, HaeIII, HhaI,
$MgCl_2$	60 mM	HpaII.
Dithiothreitol	5 mM	
pH 7.5		

Low Salt Buffer (10x)		Enzymes
Tris HCl	100 mM	BclI, HaeIII, KpnI, TaqI
$MgCl_2$	60 mM	
NaCl	60 mM	
Dithiothreitol	5 mM	
pH 7.5		

High Salt Buffer (10x)		Enzymes
Tris HCl	100 mM	AluI, AvaI, AvaII, BamHI,
$MgCl_2$	60 mM	EcoRI, HincII, HindIII, HinfI,
NaCl	600 mM	PvuII, Sau3AI, SstI, TaqI,
Dithiothreitol	5 mM	XbaI.
pH 7.5		

bottom with dialysis tubing. Fill each tube to 0.5 cm from the top (about 3 ml per tube) and allow to set. Tube gels can be stored for a few weeks if kept immersed in electrophoresis buffer and are invaluable for quick analysis of restriction digests.

Load samples mixed with an equal volume of loading buffer. Electrophorese at 100 V for about 1 hour or until the dye has travelled three-quarters of the gel. Remove gels from tubes and stain in 10 μg/ml ethidium bromide for 10-20 minutes. Place the gels on a U.V. trans-illuminator (Model C-62, U.V. Products, Winchester, Hants) to visualise the fluorescent bands (wear U.V. safety goggles). Photograph the gel using a Polaroid CU-5 Land Camera fitted with a yellow filter and Polaroid 665 (positive/negative) film. If the negative is to be retained immerse it in 1 M sodium sulphite to remove all developer then wash it thoroughly in cold water.

1% Agarose slab gels. Either horizontal or vertical slab gels can be used. I used 20 x 20 cm horizontal gels (Model HO) and 16 x 16 cm vertical gels (Model V-16), both sets of apparatus from Bethesda Research Laboratories (U.K. distributors Uniscience, Cambridge).

For vertical gels set up the apparatus with 3 mm spacers arranged to converge slightly so that the gel is 1-2 cm wider at the top than at the bottom. This helps to prevent slippage of the gel once it has set. Prepare 150 ml of 1% agarose in agarose electrophoresis buffer (as above), melted then cooled to 65^{o}C. Tip back the gel apparatus to 45^{o} from vertical and pour 50 ml of the gel solution into the bottom tank so that it fills 2-3 cm of the gel. Allow this plug to set. Return the apparatus to vertical and pour in the remainder of the agarose from the top. Position the comb and allow the gel to set.

For horizontal gels make wicks of 2% agarose. The main gel

is 1% agarose in agarose electrophoresis buffer; prepare as for other gels and pour to a thickness of approximately 3 mm (requires 200-250 ml).

Before use fill both tanks with agarose electrophoresis buffer and carefully remove the comb. Prepare and apply samples as for tube gels and electrophorese at 100-150 V (vertical) or 150-200 V (horizontal). Electrophoresis time is varied according to the requirements of a particular experiment. The bromophenol blue marker dye runs with DNA molecules of 600-800 bp in this system. Stain and visualise bands as described for agarose tube gels.

T.B.E. Buffer (1x) : 0.1 M Tris borate, 1 mM EDTA, pH 8.3 (for 1 l of 10x buffer : 121.1 g Tris base, 51.35 g boric acid, 3.72 g EDTA)

Acrylamide Stock : 19% acrylamide, 1% N,N'-methylene bisacrylamide.

4% Acrylamide Gel Mix : 10 ml acrylamide stock, 5 ml glycerol, 5 ml 10x T.B.E. buffer, 30 ml H_2O.

Glycerol Loading Buffer : 50% glycerol, 0.05% xylene cyanol, 0.05% bromophenol blue.

4% Acrylamide gels. Acrylamide gels can be conveniently prepared in the vertical apparatus described for agarose slab gels (BRL V-16). For these gels use 1.5 mm spacers, including a spacer across the bottom (the gel is formed between the glass plates before they are fitted into the apparatus.) Make up 4% acrylamide gel mix as described above, and add 0.4 ml of 10% ammonium persulphate, freshly prepared. Remove 5 ml of gel mix to a separate tube, add 40 µl TEMED (NNN'N' tetramethylethylenediamine), and use to seal around the spacers and form a small plug at the bottom of the gel. When the plug has set (less than 5 minutes) add 30 µl of TEMED to the rest of the gel mix, stir and pour between the sealed glass plates. Put the comb in place and allow the gel to set. Once the acrylamide has set remove the bottom spacer and fit the gel into

the apparatus. Fill the tanks with T.B.E. buffer and remove the comb. Also ensure that the bottom of the gel is free of air bubbles, if necessary removing trapped air using a syringe with a bent needle.

Pre-electrophorese the gel at 150 V for 30-60 minutes. To each sample add 1/5 volume of glycerol loading buffer, then apply to gel. Electrophorese at 200 V for 3-4 hours. In this system bromophenol blue runs with DNA molecules of 50-100 bp, and xylene cyanol with molecules of around 400 bp. After electrophoresis remove one glass plate, leaving the gel attached to the back plate. Stain and view as described for agarose tube gels. It is not normally necessary to remove the gel from the glass plate in order to visualise the bands on the transilluminator.

3.5 Preparation of restriction fragments

3.5.a Preparative gel electrophoresis

Electrophorese samples, stain gel and visualise DNA bands as described above. In general use agarose gels for separation of fragments over 1000 bp long and acrylamide gels to separate fragments of less than 1000 bp. For large samples use broader wells as appropriate. With large scale preparations (greater than 100 μg) it is useful to electrophorese a small portion of the sample on a tube gel to check that digestion is complete before proceeding with separation of the whole sample.

3.5.b Elution of DNA from agarose gels

Agarose Elution Buffer : 0.15 M NaCl, 10 mM Tris HCl, 1 mM EDTA, pH 8.0.

Locate the required bands and cut out on the U.V. transilluminator. Transfer each gel piece to a plastic syringe and squeeze through a 21 g needle into a siliconised screw cap tube. Wash through syringe with agarose elution buffer and add wash to screw cap tube. Bring the final volume of buffer to 10x volume of gel piece. Shake tube overnight

at room temperature. Pass eluate into a corex tube through a syringe plugged at the tip with siliconised glass wool to trap agarose fragments. Add NaCl to 0.3 M, then 2.5 volumes of ethanol. Precipitate DNA for at least 16 hours, centrifuge (2,500 g, 30 minutes, -10°C), and pour off supernatant. Redissolve pellet in 0.4 ml 0.3M sodium acetate (pH 8.0) and transfer to 1.5 ml plastic eppendorf tube. Centrifuge sample for 5 minutes in an Eppendorf Centrifuge 5412 (9,950 g) to pellet residual agarose and glass wool. Transfer supernatant to a second tube, add 1 ml ethanol and chill at -70°C (dry ice/methylated spirits) for 15-30 minutes to precipitate DNA. Store the DNA as an ethanol precipitate until required. Before use centrifuge for 5 minutes (9,950 g), remove supernatant with a finely drawn pasteur pipette and dry off residual ethanol under vacuum. All subsequent manipulations take place in 1.5 ml eppendorf tubes unless otherwise specified.

3.5.c Elution of DNA from acrylamide gels

This procedure is taken from Maxam and Gilbert (1980).

Acrylamide Elution Buffer : 0.5 M ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA, 0.1% SDS.

Seal the point of a 1 ml (blue) pipette tip in a low flame, then pack siliconised glass wool into the narrow end. Cut the required band from the gel and transfer it to the sealed tip. Grind the gel fragment with a glass rod, add 0.6 ml acrylamide elution buffer and mix. Cover tip and incubate overnight at 37°C. Cut off point of tip and allow eluate to drain through the glass wool plug into a siliconised corex tube. Rinse tip through twice with 0.2 ml acrylamide elution buffer. Add 2.5 ml ethanol to the eluate. Precipitate DNA at -70°C for at least 30 minutes. Centrifuge, recover DNA and reprecipitate exactly as described for elution from agarose.

3.5.d Preparative sucrose gradient centrifugation

Sucrose gradient centrifugation is suitable for large scale (> 100 μg) preparation of restriction fragments where the molecules to be separated differ in size by at least 3-fold. (For example this was used to purify the rDNA insert from pXlr 101L after digestion with EcoR1 and BamH1.)

NET : 1 M NaCl, 25 mM Tris HCl, 1 mM EDTA, pH 8.0.

Prepare 37 ml gradients of 10-25% (w/w) sucrose in NET. Dilute sample (restriction digest) to 1 ml with NET to dilute out any glycerol present (from enzyme storage buffers), then layer on top of gradient. Centrifuge gradients in Beckman SW27 rotor at 26,000 rpm (92,444 g) for 24 hours at 25°C. Pump the gradients through a Gilford 240 recording spectrophotometer, monitoring continuously at 260 nm, and collect about 30 fractions per gradient. Pool the peak fractions and precipitate with 2.5 volumes of ethanol at -20°C for at least 24 hours. Pellet the DNA by centrifugation (2,500 g, 30 minutes, -10°C). Redissolve the pellet in 0.4 ml of 0.3M sodium acetate, transfer to a 1.5 ml eppendorf tube and again precipitate the DNA with 1 ml ethanol.

3.6 5'-end labelling of DNA

The following methods are based on the protocols of Maxam and Gilbert (1980).

3.6.a Phosphatase reaction

To remove 5' phosphate groups from DNA incubate the restriction fragment (usually about 1 μg but more for very large fragments) with 10 μl 'clean' phosphatase (see below) in 100 μl 10 mM Tris HCl, 0.1 mM EDTA, pH 8.0, at 37°C for 1-1½ hours. Add 100 μl redistilled phenol (water saturated) and vortex for 3-4 minutes. Centrifuge for 2 minutes (9,950 g

in Eppendorf 5412), transfer the upper, aqueous, phase to a second tube, rinse the phenol phase with a further 100 µl Tris/EDTA buffer and again transfer the aqueous phase to the second tube. Add sodium acetate to the combined aqueous phase to 0.3M final concentration then add 2.5 volumes of ethanol. Chill the sample at -70°C (dry ice/methylated spirits) for 15 minutes, centrifuge for 5 minutes (9,950 g), remove the supernatant, redissolve in 0.4 ml 0.3M sodium acetate and add 1 ml ethanol. Again chill and centrifuge the sample, discard the supernatant and dry the DNA pellet under vacuum.

NOTE: The following method, from B. Sollner-Webb (Carnegie Inst.), was used to further purify commercial bacterial alkaline phosphatase. Dialyse 5 mg phosphatase (Worthington, distributors Cambrian Chemicals, Croydon, Surrey) against 10 mM Tris HCl (pH 8.0) at 4°C overnight. Pass the dialysed enzyme through a DE52 column (1.5 x 10 cm) and elute with a 240 ml gradient of 0-0.2M NaCl in Tris HCl pH 8, collecting 2 ml fractions. Assay the fractions for phosphatase activity using p-nitrophenol phosphate (measure production of p-nitrophenol spectrophotometrically at 410 nm) and pool peak fractions. Dialyse against 10 mM Tris HCl, 0.05M NaCl, pH 8.0, then add glycerol to 50% (~ 16 ml total). Store at -20°C.

3.6.b Polynucleotide kinase reaction

If possible use $[\gamma\text{-}^{32}P]$ ATP, > 5,000 Ci/mmol, 10 mCi/ml in aqueous solution, either from Amersham International Ltd., Amersham, or New England Nuclear, Southampton. The concentrated aqueous solution can be used directly without lyophilisation and in my hands gives consistently much better (10x - 100x) incorporation of ^{32}P into restriction fragments than ethanol solutions (which have to be lyophilised before use).

Dissolve the DNA in 2 µl H_2O then bring the reaction mix to 10 µl with a final concentration of 50 mM Tris HCl pH 8.0, 10 mM $MgCl_2$, 5 mM dithiothreitol and 60 µCi $[\gamma\text{-}^{32}P]$ATP. Add 1 µl polynucleotide kinase (P-L Biochemicals, U.K. distributor Windsor Laboratories, Slough) and incubate at 37°C for 30 minutes. After incubation add 40 µl 2.5 M ammonium acetate and 160 µl ethanol, chill at -70°C for 15 minutes, centrifuge (9,950 g) for 5 minutes, then remove supernatant (to radioactive waste). Redissolve in 100 µl 0.3 M sodium acetate, add 300 µl ethanol, chill, centrifuge, remove supernatant and dry pellet under vacuum.

This protocol is best for DNA fragments which have 5' 'protruding' ends, for instance after cleavage with EcoR1 (5' G↓AATTC 3' / 3' CTTAA↑G 5'). If fragments have flush ends, eg SmaI (5' CCC↓GGG 3' / 3' GGG↑CCC 5') or 5' recessed ends, eg HhaI (5' GCG↓C 3' / 3' C↑GCG 5') the kinase reaction should be modified. In the reaction use Tris HCl pH 9.5 instead of pH 8.0 and also include 1 mM spermidine. Before adding polynucleotide kinase heat the reaction mix to 90°C for 2 minutes then chill on ice. Subsequent incubation and precipitations are exactly as described above.

3.6.c Separation of labelled ends

To produce DNA labelled at one 5' end only, for instance for DNA sequencing, the labelled fragment can either be cleaved with a second restriction enzyme or the two strands can sometimes be separated.

Digest the DNA with a restriction enzyme which gives two labelled fragments of different lengths. Separate the fragments on a 4% acrylamide gel, locate either by ethidium bromide staining or autoradiography (5-20 minute exposure) and elute as described (3.4).

The principle of strand separation is that two complementary DNA strands may form different secondary structures when allowed to renature.

Dissolve the DNA in 30% (v/v) DMSO, 1 mM EDTA, 0.05% xylene cyanol and bromophenol blue. Heat to 90°C for two minutes to ensure complete denaturation, chill to 0°C then load immediately on a non-denaturing gel. Use a 5% acrylamide gel with 0.5 x T.B.E. buffer and 40 cm long to improve resolution. Locate the strands by autoradiography (5-20 minutes exposure) and elute as before.

The radioactivity of the dry pellet can be estimated by Cerenkov radiation. Place the eppendorf tube with pellet in a glass scintillation vial and count using channel settings for tritium.

3.7 DNA Sequence Analysis

A very clear and detailed account of DNA sequencing by the chemical method is given by Maxam and Gilbert (1980). However some volumes and reaction times have been adjusted and the exact procedure used is described below.

3.7.a Base specific modification

Pyridinium formate : 4% (v/v) formic acid adjusted to pH 2.0 with pyridine (using 0.005M H_2SO_4 as pH 2.0 standard).

DMS buffer : 50 mM sodium cacodylate, 10 mM $MgCl_2$, 0.1 mM EDTA, pH 8.0.

'DMS stop' : 1.5M sodium acetate, 1M β-mercaptoethanol, 100 μg/ml yeast RNA.

'HZ stop' : 0.3M sodium acetate, 0.1 mM EDTA, 50 μg/ml yeast RNA.

The reactions used in this work are specific for guanine, guanine + adenine, cytosine + thymine and cytosine respectively. Carry out the reactions in siliconised 1.5 ml eppendorf tubes with the identification scratched on the tube (siliconise with 'Repelcote', Hopkin & Williams, Chadwell Heath, Essex). Take up the end-labelled DNA

(preferably at least 100,000 Cerenkov cpm) in 11 µl H_2O and mix thoroughly to dissolve. Pipette 2.5 µl of labelled DNA into each of 4 tubes and add 1 µl of carrier DNA (1 mg/ml, sonicated calf thymus DNA).

For the A + G reaction add 11 µl H_2O and 2.5 µl pyridinium formate and incubate at 30°C for 70 minutes. After the reaction freeze the sample, lyophilise, add 10 µl H_2O and lyophilise to dryness once more. This sample is now ready for the piperidine (strand cleavage) reaction.

Meanwhile add 98 µl DMS buffer to the 'G' tube, then 0.5 µl dimethylsulphate ('Gold Label', Aldrich Chemical Co., Gillingham, Dorset). Mix well and incubate at 20°C for 5 minutes, then add 24 µl 'DMS stop', and 400 µl ethanol (Analar grade). Mix and chill at -70°C for at least 15 minutes.

Add 6 µl H_2O to C + T tube, and 8 µl saturated NaCl to C tube. To each add 15 µl hydrazine (Kodak Ltd., Kirkby, Liverpool) and incubate at 20°C for 8 minutes (C + T) or 10 minutes (C only), then add 60 µl 'HZ stop' and 250 µl ethanol (Analar). Mix and chill at -70°C for 15 minutes, but do not allow the ethanol mix to solidify as hydrazine may precipitate into a separate phase, carrying the DNA with it.

Centrifuge the G, C + T and C tubes for 5 minutes at 9,950 g, remove the supernatants, redissolve in 60 µl 0.3M sodium acetate then add 200 µl ethanol. Again chill and centrifuge. Wash the pellet with 70% aqueous ethanol then 100% analar ethanol, chilling for 5 minutes and centrifuging for 5 minutes after each wash. These samples are now ready for the piperidine reaction.

NOTE: Dimethylsulphate and hydrazine are dangerous and should be handled with care in a fume hood. They are also labile and should not be exposed to air more than necessary. All waste DMS (including supernatants) must be inactivated with 5M NaOH and hydrazine waste with 3M ferric chloride.

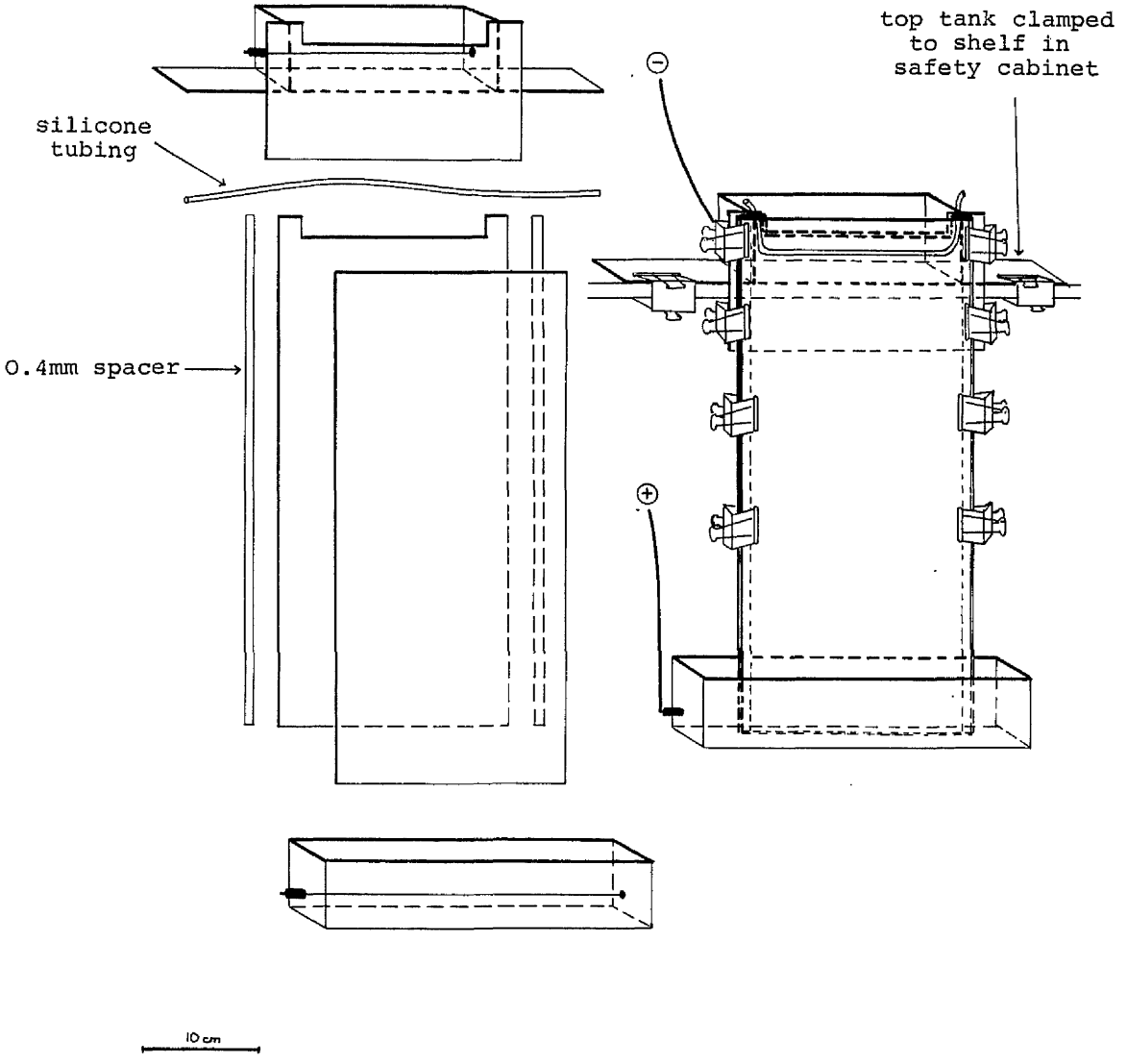


Figure 3.2. Scale drawing of apparatus for sequencing gels.

3.7.b Strand scission with piperidine

Dissolve the dry DNA pellets from all four modification reactions in 100 µl of 10% (v/v) piperidine (Koch-Light Laboratories, Colnbrook, Bucks). Heat to 90°C for 30 minutes. Weights must be placed on top of the tubes to prevent the caps from popping open. After 30 minutes chill the tubes on ice, centrifuge for 30 seconds, freeze and lyophilise. When dry dissolve in 20 µl water, freeze and lyophilise, and repeat. This step helps to remove residual piperidine and ensures that the sample is concentrated at the bottom of the tube. Cerenkov count the samples and dissolve in 99% formamide, 0.05% xylene cyanol.

3.7.c Sequencing gels (Sanger and Coulson, 1978)

Sequencing gel mix : (for two gels) 33.6 g urea (to 7M), 32 ml acrylamide stock (to 8%, see 3.4.b), 8 ml 10x T.B.E. buffer (see 3.4.b), 14 ml H_2O.

The apparatus used for sequencing gels is shown in Figure 3.2. Two gels may be run simultaneously using an LKB 2103 power supply. The gels were run inside a safety cabinet because of the high voltage to be applied.

Wipe both glass plates with 'Repelcote' to siliconise, then rinse with distilled water. Position the spacers between the two plates and carefully seal the sides and bottom with waterproof tape (Universal Scientific Ltd., London), then clamp firmly with bulldog clips. Make up the gel solution as above. Add 0.6 ml of freshly made 10% (w/v) ammonium persulphate then filter and degas the solution. Add 32 µl TEMED, mix, and carefully pour gel. I find it easiest to pour with the gel plates at an angle of around 20° from horizontal using a pipette to release a steady trickle of solution down the centre. Take care to eliminate any bubbles as soon as they appear by tilting the gel and tapping firmly in

the region of the bubble. Once filled, position the comb and lay the plates flat until the gel has set.

Pre-electrophorese the gel for about 2 hours (with T.B.E. buffer), limiting the current to 25 mA until the voltage reaches as near 2 kV as possible. The voltage is usually between 1.0 and 1.5 kV at the start.

Heat the samples to 90°C for 2 minutes, chill, and load 1-1.5 µl of each on the gel (ideally 10,000 cpm per track for subsequent autoradiographic exposure of 18-24 hours). Electrophorese at 25 mA, 1.5-2 kV. The length of run depends on requirements and up to three successive loadings may be made on one gel. In a typical experiment run xylene cyanol to 55 cm, 35 cm and 15 cm respectively.

After electrophoresis carefully remove one glass plate, cover the gel with cling film and autoradiograph at -70°C with Kodak X-Omat HI film and an intensifying screen (Cronex Lighting-Plus, Du Pont UK, Huntingdon, Cambs.). After 1-2 days, or more for low radioactivity, develop the film for 4 minutes in Kodak DX80, rinse and fix in Kodak FX40 for twice the clearing time. If the gel is to be autoradiographed a second time ensure that it does not thaw out, and wipe away any condensation before putting on the second film.

3.8 Preparation of ribosomal RNA

3.8.a Tissue culture cells and media

The cells used to prepare ribosomal RNA are derived from *Xenopus laevis* kidney. They were given to us by Dr. K. Jones, Edinburgh, and originally came from Dr. K. Rafferty, Johns Hopkins, Baltimore. The cells are maintained as monolayers in Roux bottles (20 oz medical flats) with 50 ml of medium, seeded with 5×10^6 cells and grown at 25°C.

The monolayer becomes confluent after 3-4 days. Cell cultures were maintained and contamination checks performed by the staff of the Wellcome Cell Culture Unit. (Cells are also stored in vials in liquid nitrogen, in 50% medium, 50% glycerol.)

Xenopus medium : Glasgow Modification of Eagles Medium (with glutamine, with 126.4 g/l arginine, without $NaHCO_3$, as obtained from Flow Laboratories, Irvine), with addition of 10% foetal calf serum (Gibco Bio-Cult, Paisley), 1% Non-Essential Amino Acids (Flow Laboratories), 0.22% $NaHCO_3$, 0.01% streptomycin, 100,000 units/l penicillin.

B.S.S. : 6.8% NaCl, 0.4% KCl, 0.2% $MgSO_4.7H_2O$, 0.14% $NaH_2PO_4.2H_2O$, 0.393% $CaCl_2$, 0.015% phenol red.

Trypsin : 0.25% trypsin in 10 mM trisodium citrate, 0.1M NaCl, 1.5 ml/l 1% phenol red, pH 7.8.

Versene : 0.02% EDTA, 0.8% NaCl, 0.o2% KCl, 0.115% Na_2HPO_4, 0.02% KH_2PO_4, 1.5 ml/l 1% phenol red.

Harvest cells with a 1:4 trypsin:versene mixture warmed to 37°C. Pour off the medium, rinse with trypsin/versene and pour off, add a further 10 ml of trypsin/versene and pour off leaving about 2 ml in the bottle. When the cells start to detach from the bottle add 20 ml of B.S.S. containing 0.22% $NaHCO_3$ and 10% calf serum. Shake off cells and transfer to a 30 ml corex tube on ice. Centrifuge cells at 800 g for 2 minutes at 4°C, pour off supernatant and resuspend cells in 20 ml B.S.S. with 0.22% $NaHCO_3$. Centrifuge again and resuspend in 4 ml BSS/$NaHCO_3$.

3.8.b Extraction of ribosomal RNA

Add 4 ml of water-saturated phenol to 4 ml cell suspension (see above), and vortex for 4 minutes. Centrifuge at 2,500 g for 10-15 minutes (room temperature) to separate the aqueous and phenol phases.

Recover the aqueous (upper) phase into a fresh corex tube, add 10 ml ethanol and chill at -20°C for 2 hours. The 'cold' phenol extraction should recover most cytoplasmic ribosomal RNA while most nuclear RNA (and residual cytoplasmic RNA) remains trapped at the interphase. To extract nuclear RNA add an equal volume of BSS to the phenol (plus interphase) layer, add SDS to 0.5%, and heat to 55°C. Vortex for 4 minutes, maintaining at 55°C. Centrifuge for 10 minutes, recover the aqueous phase, add 2.5 volumes of ethanol and chill to precipitate RNA. Alternatively the 'hot' phenol extraction may be performed directly on whole cells to recover all RNA simultaneously. Note that heating to 55°C will denature the RNA and release 5.8S from 28S rRNA. Pellet the RNA by low speed centrifugation (2,500 g, 30 minutes), remove the supernatant and redissolve pellet in 1 ml LETS (see below) prior to separation on sucrose gradients.

Differential phenol extraction is discussed by Markov and Arion (1973), but the exact conditions used have been adapted to suit these cells.

3.8.c Separation of ribosomal RNA

LETS : 0.1M LiCl, 10 mM EDTA, 10 mM Tris HCl, 0.2% SDS, pH 7.4.

Separate the ribosomal RNA species on 37 ml 10-25% (w/w) sucrose/ LETS gradients. Centrifuge in a Beckman SW27 rotor at 23,000 rpm for 16 hours at 20°C. Pump the gradients through a Gilford 240 recording spectrophotometer, monitoring continuously at 260 nm, and collect about 30 fractions per gradient. Pool the peak fractions and precipitate RNA with 2.5 volumes of ethanol at -20°C overnight. To release 5.8S from 28S RNA redissolve in LETS and heat to 55° for 4 minutes. Run the RNA on a second sucrose gradient to separate (18S may also be further purified by running on a second sucrose gradient).

3.8.d Preparation of ^{32}P-labelled RNA

Grow cells in normal medium as above. After two days of growth pour off medium and rinse with 'low phosphate' medium ($\frac{1}{10}$ normal concentration of phosphate). Add 25 ml of low phosphate medium and 5 mCi $^{32}PO_4$ (Amersham International, Amersham). Be careful not to inject $^{32}PO_4$ directly onto cell layer. Mix and incubate for a further 2 days. Harvest RNA as described above.

3.9 Filter hybridisation

^{32}P ribosomal RNA is hybridised to DNA restriction fragments immobilised on a filter. This method can be used analytically or to prepare specific small ribosomal RNA fragments for subsequent analysis (Salim & Maden, 1980).

20x SET : 3M NaCl, 1M Tris HCl, 20 mM EDTA, pH 8.0.

3.9.a Filtration of DNA

Use 2.5 cm nitrocellulose filters (0.45 μm pore, Schleicher and Schuell, distributors Anderman & Co., East Molesey, Surrey). Soak filters in 5x SET before use. Meanwhile denature DNA in 100 μl of 0.1M NaOH then add to 10 ml of 5x SET and pass the solution through a filter at a rate of around 3 ml per minute. Allow the filter to dry in air then bake at 65^{o}C for 4 hours. The DNA should now be firmly attached to the filter. (Fragments less than 300 base pairs are not efficiently retained.) Also closed circular DNA must be linearised before use to allow stable denaturation.

3.9.b Hybridisation

Dissolve molar excess of ^{32}P RNA in 0.4 ml of 5x SET and incubate this with the filter in a scintillation vial at 65^{o}C overnight. After hybridisation drain off the SET and wash the filter thoroughly with

at least four changes of 2x SET. For preparation of specific RNA fragments the unhybridised RNA 'tails' must be trimmed away. Place the filter in 0.5 ml of 2x SET with 25 units T_1 ribonuclease (Calbiochem, Bishops Stortford, Herts) and incubate at 37°C for 15 minutes. Rinse four more times with 2x SET, removing washes to an RNase wash bottle (care must be taken not to contaminate glassware etc with T_1RNase).

3.9.c Recovery of hybridised RNA

Rinse the filter briefly with 0.1x SET then add 2 ml 0.1x SET, heat in a boiling water bath for 15 minutes, then chill on ice. Transfer the 2 ml of 0.1x SET (now containing the RNA) to a tube with 4 ml of 2M NaCl, 0.1M $MgCl_2$ and 20 µg carrier RNA. Rinse the filter with 1.5 ml 0.1x SET and transfer the rinse to the same tube. Add 10 µl DNase 1 (1 mg/ml, Worthington, distributors Cambrian Chemicals, Croydon, Surrey); incubate at 20°C for 15 minutes. Add an equal volume of water-saturated phenol, vortex 2-3 minutes and separate the phases by centrifugation. Recover the aqueous phase and add 2% volumes of ethanol. Chill to precipitate the RNA, if possible at -20°C for 36 hours then -70°C for 20 minutes to ensure recovery from this large volume. Centrifuge at 2,500 g for 30 minutes to pellet the RNA.

3.10 Analysis of RNA

3.10.a 'Fingerprinting'

Methods for RNA fingerprinting are described in detail by Brownlee (1972), see also Salim (1972) and Maden and Salim (1974).

1st dimension buffer : 5% (v/v) acetic acid, 7M urea, adjusted to pH 3.5 with pyridine.

2nd dimension buffer : 7% (v/v) formic acid.

Liquid scintillant : 5 g/l PPO (2,5-diphenyloxazole) in toluene.

Digest RNA (at least 20 μg, including carrier RNA) in 5 μl of 20 mM Tris HCl, 2 mM EDTA, pH 7.4, with 2 μg T1 ribonuclease at 37°C for 30 minutes. The first dimension separation is by electrophoresis on a cellulose acetate strip (3 x 95 cm) in 1st dimension buffer. Apply the sample at 15 cm from one end (nearest cathode) and electrophorese at 4.7 kV for 2½ - 3½ hours. Transfer the required region to DEAE cellulose paper (43 x 94 cm) by blotting through with water. Remove excess urea from the paper by washing with methylated spirits. Run the second dimension in 2nd dimension buffer at 1.2 kV for 16 hours with constant cooling. After electrophoresis dry the DEAE paper, mark with ink containing ^{35}S, then autoradiograph with Kodak NS-5T film. Individual oligonucleotide 'spots' can be located on the paper by reference to the autoradiograph and can be cut out and counted. Place the 'spot' in a vial with 10 ml liquid scintillant and analyse in a scintillation counter. (Note - this is a summary of a more detailed protocol which is available from B.E.H. Maden.)

3.10.b Further analysis of separated oligonucleotides

Elute oligonucleotides from DEAE cellulose paper with 30% (v/v) triethylamine carbonate, pH 10.0, then evaporate to dryness. Redissolve and digest in 10 μl of 10 mM Tris HCl, 1 mM EDTA, pH 7.4 with 2 μg pancreatic ribonuclease (RNase A, Calbiochem, Bishops Stortford, Herts) at 37°C for 30 minutes. Separate products by electrophoresis on Whatman 52 paper at pH 3.5 for 40 minutes at 4.7 kV. Dry and autoradiograph as before.

3.11 Mapping RNA termini by S1 nuclease protection

The principle of this technique, first described by Berk and Sharp (1978), is explained in Chapter 6. This protocol was taken from Moss and Birnstiel (1979).

3.11.a Hybridisation and S1 digestion

Take up DNA (end labelled, approx 100,000 cpm) and RNA (3-30 μg) in 15 μl of 80% formamide, 0.25M NaCl, 25 mM MES (2-[N-morpholino]ethane-sulphonic acid), 0.6 mM EDTA, pH 6.7. Hybridise at 60°C for 3 hours, then add 135 μl S1 buffer (0.25M NaCl, 1 mM $ZnCl_2$, 30 mM sodium acetate buffer, 5% glycerol, pH 4.6), cool to 45°C and digest for 1 hour with S1 nuclease (Sigma, Poole, Dorset). Add 2 μl carrier DNA (1 mg/ml) and 400 μl ethanol (analar). Chill at -70°C for 15 minutes then centrifuge for 5 minutes at 9,950 g. Remove the supernatant, redissolve the sample in 10 μl H_2O and lyophilise to dryness. This step makes it easier to dissolve the sample in a small volume for electrophoresis. Estimate the radioactivity by counting Cerenkov radiation.

3.11.b Electrophoresis

Electrophorese the samples on a 'sequencing' type gel, alongside a sequencing ladder of the same DNA fragment that was used for the hybridisation. This ensures accurate sizing of the protected DNA fragment.

3.12 Northern transfers

Essentially the method of Alwine et al (1977) was used, modified mainly as described in the data-sheet for 'Transa-Bind' (Schleicher and Schuell).

3.12.a Electrophoresis of RNA

Electrophoresis buffer : 0.5M boric acid, 50 mM $Na_2B_4O_7.10H_2O$, 0.1M Na_2SO_4, 0.01M EDTA, pH 8.2.

Best results were obtained by electrophoresis in denaturing gels containing methyl mercuric hydroxide. Make up a 1% agarose vertical slab gel as described (3.4.b) but using the above buffer. After cooling

to 65°C add methyl mercuric hydroxide (Lancaster Synthesis Ltd., Lancaster) to 7.5 mM, then pour gel. Methyl mercuric hydroxide is highly toxic so the gel should be poured in a fume cupboard. Rinse pipette tips and glassware in 5M ammonium acetate before disposal. Load sample in running buffer with 10% glycerol and 0.05% bromophenol blue and xylene cyanol. Electrophorese at 40V for about 16 hours.

Wash the gel in 50 mM NaOH, 5 mM β-mercaptoethanol, 2-5 μg/ml ethidium bromide, for 40 minutes. This treatment nicks the RNA to smaller pieces for the transfer, inactivates methylmercuric hydroxide, and stains the RNA. Neutralise by washing the gel three times for 5 minutes in 0.1M sodium phosphate buffer, pH 7.0, containing 7 mM iodo-acetic acid. Then wash twice for five minutes with 0.2M sodium acetate buffer, pH 4.0 (14.8 g anhydrous sodium acetate, 47.6 ml glacial acetic acid, H_2O to 1 litre). View with U.V. illumination and photo-graph.

3.12.b Activation of NBM paper and Transfer

Nitrobenzyloxymethyl (NBM) paper was obtained from Schleicher and Schuell (Transa-Bind, U.K. distributors, Anderman & Co., East Molesey, Surrey). Start the activation about 20 minutes before the NaOH treatment of the gel. Incubate the paper for 30 minutes at 60°C in 20% (w/v) sodium dithionite ($Na_2S_2O_4$), 0.4 ml per cm^2 of paper. This incubation should be done in a fume cupboard. Wash the paper several times with water, once with 30% acetic acid for 5 minutes and again several times with water. Place the paper in ice-cold 1.2M HCl (0.3 ml per cm^2 paper) with 0.03 volumes of fresh sodium nitrite solution (10 mg/ml). Incubate on ice for at least 30 minutes and leave until the gel is ready for transfer. Immediately before transfer wash the paper twice with water and twice with 0.2M sodium acetate pH 4.0 (paper turns orange).

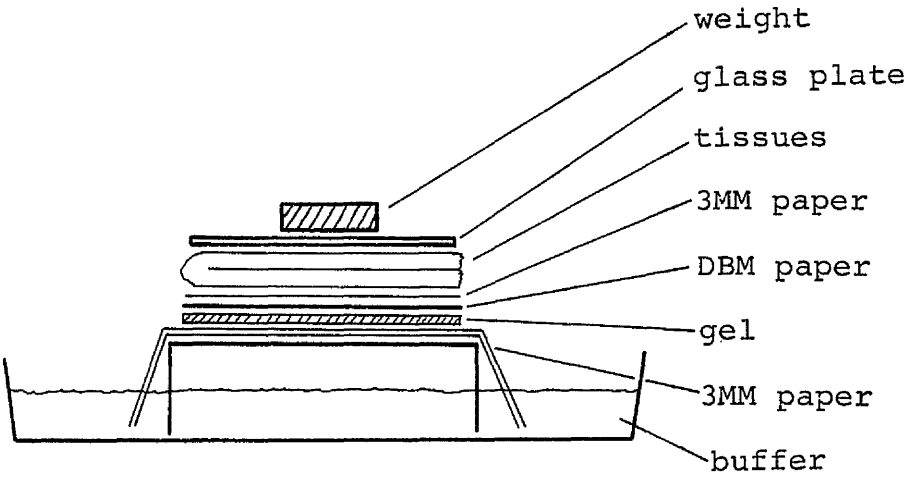


Figure 3.3 Apparatus for a 'Northern' transfer.

Set up the gel and activated paper as shown in Fig 3.3. Allow 0.2M sodium acetate buffer to blot the RNA from the gel onto the paper. Change the paper towels several times and leave overnight to complete the transfer. After transfer incubate the paper at 37°C for at least 3 hours in hybridisation buffer (see below) plus 1% glycine. Store the paper at 4°C in the same buffer.

3.12.c Preparation of DNA Probes

Radioactive probes must be made to a high specific activity. Label DNA fragments by nick-translation (Rigby et al, 1977). Use [α-^{32}P] dCTP (~2,000-3,000 Ci/mmol, Amersham International, Amersham), dry down before use. Nick-translation is carried out with 1 µg (or less) DNA, 50 mM potassium phosphate buffer, pH 7.4, 5 mM $MgCl_2$, 20 µM each dATP, dGTP and dTTP, 50-100 µCi [α-^{32}P] dCTP (~25-50 pmoles), all in 8 µl, plus 1 µl DNase 1 (10 µg/ml, Worthington) and 1 µl DNA polymerase (BCL, Lewes, East Sussex). Incubate at 14°C for 1-2 hours, then add 1 µl carrier DNA (1 mg/ml), 90 µl 0.3M sodium acetate, 4 µl 0.25M EDTA, 300 µl ethanol. Chill at -70°C for 15 minutes, centrifuge at 9,950 g for 5 minutes, then remove supernatant. Redissolve DNA in 100 µl 0.3M sodium acetate and again precipitate with 300 µl ethanol.

3.12.d Hybridisation

Hybridisation buffer : 50% formamide, 50 mM sodium phosphate, pH 6.5, 0.75M NaCl, 75 mM trisodium citrate, 0.02% ficoll (type 400 DL, Sigma), 0.02% polyvinyl pyrrolidone (PVP 360, Sigma), 0.2% SDS, 0.02% albumin (bovine), 1 mg/ml yeast RNA.

Dissolve the DNA probe in 10 µl H_2O, heat to 90°C for 2 minutes to denature, then add hybridisation buffer (50 µl/cm^2 paper). Place the paper and hybridisation mix in a small polythene bag and seal with air

excluded. Hybridise for 16-24 hours at 37-42°C. After hybridisation wash the paper for at least three hours in three or more changes of buffer (50% formamide, 50 mM sodium phosphate, pH 6.5, 3.75 mM NaCl and 0.375 mM trisodium citrate). Blot the paper dry with 3MM paper, cover with cling film, then autoradiograph with Kodak X-Omat HI film and an intensifying screen at -70°C.

The hybridised probe can be washed off by incubating the paper in several changes of 99% formamide at 65°C. The paper can then be reused.

3.13. Computer Analysis

All programs were run on a PDP 11/34 computer with the help of Dr. R. Eason. Programs HAIRPN and TRNA were obtained from R. Staden, MRC, Cambridge (Staden, 1978 and 1980).

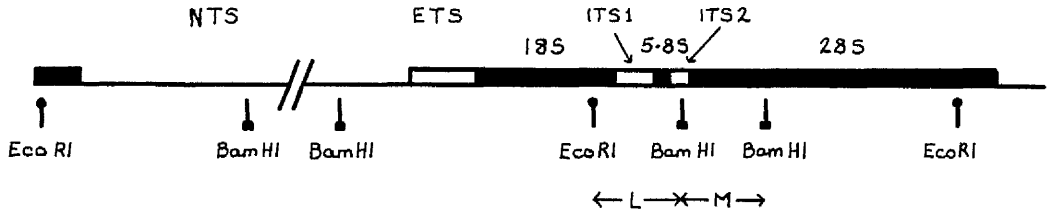


Figure 4.1 Restriction sites for EcoRI and BamHI in a unit of ribosomal DNA

CHAPTER 4 DETERMINATION OF THE NUCLEOTIDE SEQUENCE

4.1 Subcloning the 18S-28S intergene region

The restriction sites for enzymes EcoRI and BamHI are shown in figure 4.1. These two enzymes divide the transcription unit into several sections of which one, denoted region L, contains the 3' end of the 18S gene and most of the internal transcribed spacer region. Initially it was not clear whether the 5' end of the 28S gene also lay within region L or not. Experiments by B.E.H. Maden failed to detect hybridisation between 28S rRNA and DNA from region L (Maden, 1980). Brand and Gerbi (1979) also reported that the start of the 28S gene was approximately 50 bp to the right of the BamHI site, within region M.

In order to simplify preparation of restriction fragments for sequence determination I decided to subclone regions L and M from pXlrl01 into pBR322. Analogous subclones had previously been made from pXlrl1 (Maden, 1980). I also wished to subclone region L from pXlrl02, pXlrl03, and pXbrl01 in preparation for comparative experiments.

4.1.a Subcloning region L

Region L is bounded at one end by an EcoRI site and at the other by a BamHI site. It is thus suitable for insertion between the single EcoRI and BamHI sites in pBR322 (Sutcliffe, 1978). Insertion of foreign DNA between these two sites in pBR322 destroys the activity of the tetracycline resistance gene but leaves the ampicillin resistance gene intact. Each of the 'parent' rDNA plasmids was digested with both EcoRI and BamHI together, to check that the products were as expected, and could be separated from each other. The DNA was separated on a 1% agarose gel as described in 3.4, and the gel obtained is shown in figure 4.2. Region M is clearly separated from region L but region L cannot be

Figure 4.2 Analytical electrophoresis of restriction fragments for subcloning.

1. 2 μg pXbr101 digested with BamHI + EcoRI
2. " pXlr101 " " " "
3. " pXlr102 " " " "
4. " pXlr103 " " " "
5. " pXlr11 " " " " (length marker)
6. 2 μg pXlr103 digested with BamHI + HindIII
7. " pXlr102 " " " "
8. " pXlr101 " " " "
9. 1 μg pBR322 " " " " (length marker)

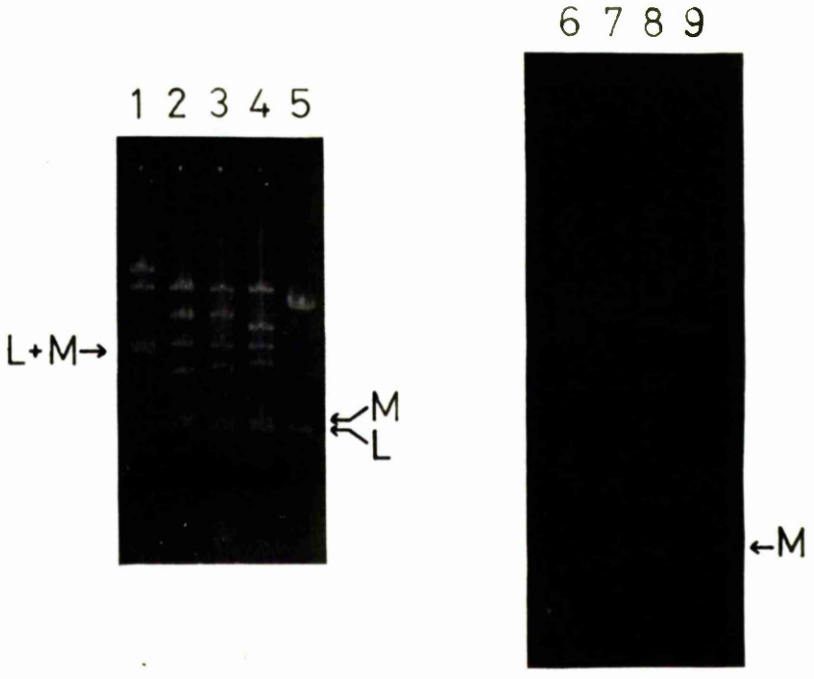
The samples were electrophoresed on 1% agarose horizontal slab gels (3.4). Maps are shown with the predicted positions of sites for EcoRI (RI), BamHI (B) and HindIII (H) in an rDNA unit and pMB9 (from Boseley et al, 1978; Boseley et al, 1979; Bolivar et al, 1977).

Expected fragment sizes: BamHI + EcoRI, 4750, 3350, 2250, 1990 or 2040, 1180(M), 1120(L), 1100, 840 or 790.

BamHI + HindIII, 5050, 4470, 2990, 1690, 1180(M), 1100, 350.

Fragments underlined are derived from the non-transcribed spacer so are variable in length as seen on the gels. Where alternative lengths are given the length depends on the orientation of the rDNA with respect to pMB9. It can be deduced that Xbr101 (which has the same EcoRI sites as X. laevis plasmids) does not have BamHI sites between L and M or in the non-transcribed spacer.

Note: In this and subsequent figures the gels are illustrated with the cathode of the bottom of the gel. Thus the direction of migration is always from top to bottom.


6 7 8 9
1 2 3 4 5
L+M→
M
L
←M

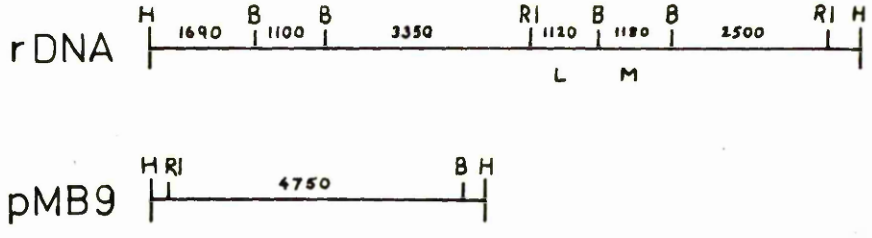

rDNA
H
1690
B
1100
B
3350
RI
1120
B
1180
B
2500
RI
H
L
M
pMB9
HRI
4750
B H

Figure 4.2

separated from a BamHI fragment derived from the non-transcribed spacer. In the X. borealis plasmid fragments corresponding to L and M were not seen, but an additional large fragment of the size expected for L plus M was present. This suggests that the site for BamHI between regions L and M is absent in X. borealis rDNA, but that the positions of other sites within the transcription unit are the same as in X. laevis.

The band containing region L was prepared from a large scale BamHI plus EcoRI digest of each plasmid (50 μg DNA). Similarly pBR322 was digested with BamHI plus EcoRI and the larger fragment purified from a gel. The two fragments were ligated as described in 3.3, and the ligation mixture was used to transform HB101. Bacteria which had become resistant to ampicillin were selected for. It would be expected in this experiment that almost all transformants to ampicillin resistance should contain the L subclone. It is not possible for pBR322 to re-circularise because the ends have been cut with different enzymes. Furthermore the contaminating BamHI spacer fragment could not form a recombinant with pBR322 because it does not have an EcoRI end. Thus only region L could be ligated at both ends with pBR322, although it would also be possible for more complex recombinants to be formed containing both region L and the spacer fragment.

Ligation was performed with approximately 4 μg of L from pXlrl01 and 0.5 μg of the pBR322 fragment. Many hundreds of ampicillin resistant colonies were obtained. Of these, twenty colonies were picked onto an agar plate containing ampicillin and a plate containing tetracycline, and all grew only on the ampicillin plate as expected. Six colonies were then grown in 100 ml of L-broth and the plasmid prepared by cleared lysis, as described in 3.2. After cleared lysis the DNA was further purified by phenol extraction but density gradient centrifugation was omitted. A sample of DNA from each colony was run on a 1% agarose gel as shown in

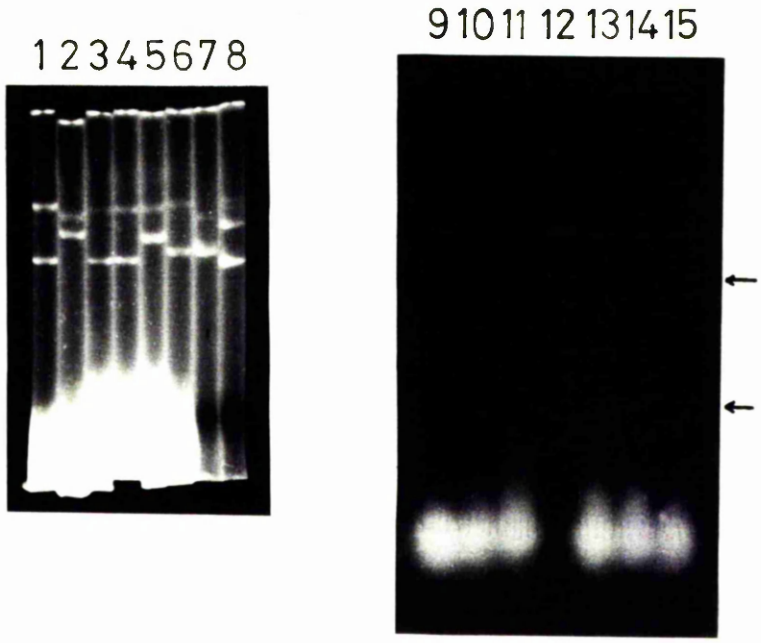
Figure 4.3 Test cultures of pXlr101L

DNA prepared by cleared lysis from 100 ml cultures of transformants (expected to contain pXlr101L) was dissolved in 500 µl of 10 mM Tris HCl, 0.1 mM EDTA, pH 8.0. The following samples were electrophoresed on 1% agarose gels:

1. 20 µl transformant 1
2. " " 2
3. " " 3
4. " " 4
5. " " 5
6. " " 6
7. 1 µg pXlr11L
8. 1 µg pBR322
9. 5 µl transformant 1 digested with BamHI
10. " " 3 " " "
11. " " 2 " " "
12. 1 µg pXlr11L digested with BamHI
13. 5 µl transformant 4 digested with BamHI
14. " " 5 " " "
15. " " 6 " " "

In 1-8 the lower band represents closed circular DNA and the upper band represents open circular DNA. Transformants 1, 3, 4 and 6 are the same length as pXlr11L. In 9-15 the arrows point to extra BamHI fragments in transformants 2 and 5 (tracks 11 and 14). In all cases the brightly staining area at the bottom of the gel is bacterial RNA which would normally be removed during caesium chloride density gradient centrifugation.

The structure expected of pXlr101L is shown below the gels.



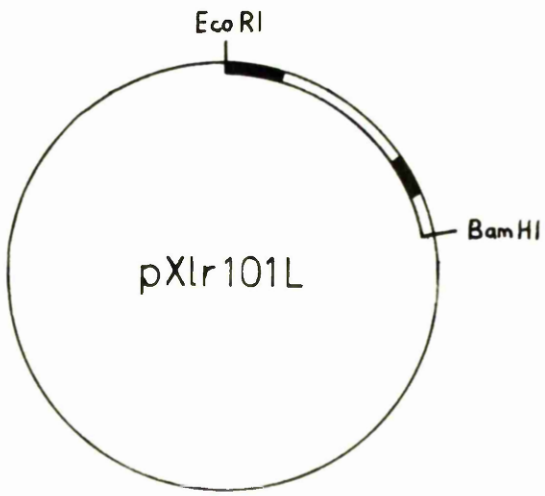


Figure 4.3

Figure 4.4 Test cultures of pXlr101M

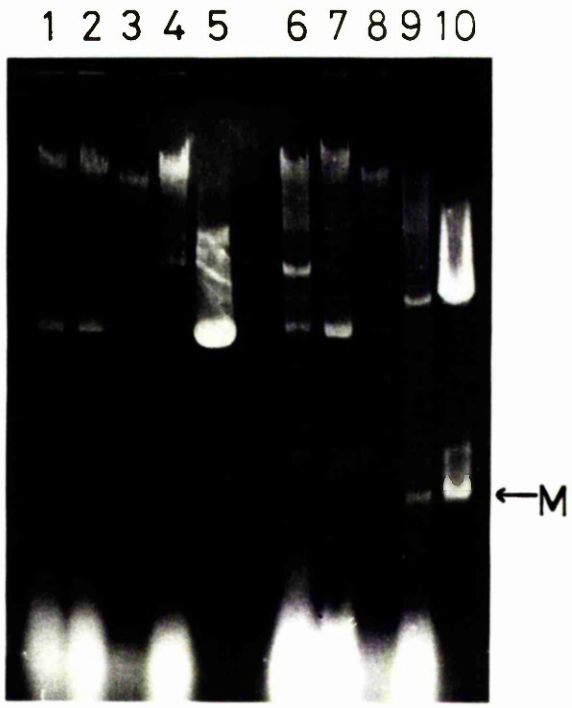
DNA prepared by cleared lysis from 50 ml cultures of transformants (expected to contain pXlr101M) was dissolved in 200 µl of 10 mM Tris HCl, 0.1 mM EDTA, pH 8.0. The following samples were electrophoresed on a 1% agarose gel:

1. 10 µl transformant 1
2. " " 2
3. " " 3
4. " " 4
5. 2 µg pXlr11M
6. 10 µl transformant 1 digested with BamHI
7. " " 2 " " "
8. " " 3 " " "
9. " " 4 " " "
10. 2 µg pXlr11M digested with BamHI.

The brightly staining area at the bottom of the gel is bacterial RNA, and the irregular bands at the top are bacterial 'chromosomal' DNA. Both of these contaminants would normally be removed by density gradient centrifugation.

Transformants 1, 2 and 3 are all the same length as pXlr11M. Digestion with BamHI is incomplete, possibly because of the large amount of bacterial DNA present. Transformants 1 and 4 both produce a BamHI fragment corresponding to M, but transformant 4 is larger than pXlr11M so probably contains more than one copy of M, so transformant 1 was selected for further use.

The expected structure of pXlr101M is shown. The rDNA fragment may also be inserted in the opposite orientation but subsequent experiments revealed that the orientation shown here is correct for pXlr101M from transformant 1.



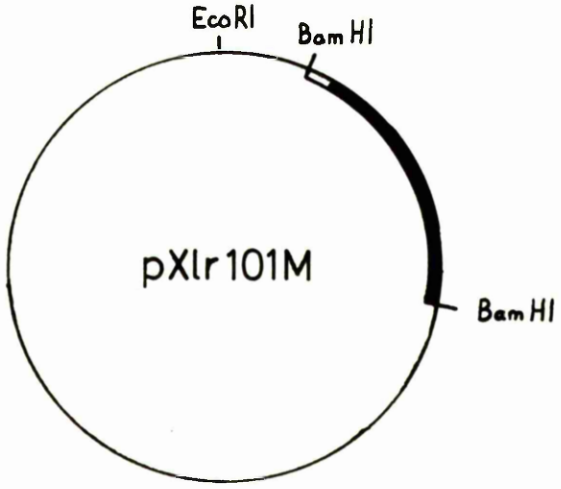


Figure 4.4

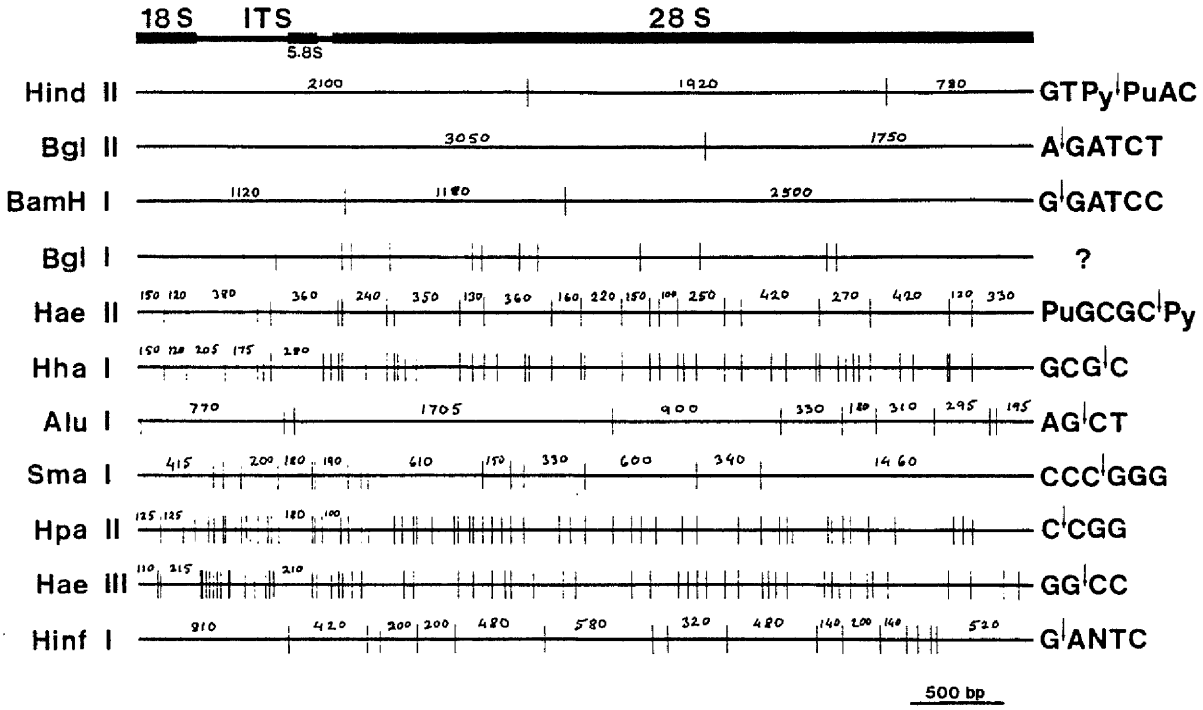


Figure 4.5 Restriction map of pXL212

Restriction sites determined by Boseley et al, (1978). Sizes of restriction fragments from P.Boseley, personal communication.

figure 4.3. Four of the plasmids were of the expected length but two were longer. Digestion with BamHI confirmed that two of the plasmids contained extra fragments while the remaining four had one site for BamHI and were of identical length to pXlrllL as expected. A large scale preparation was then made from colony 1, and the plasmid was denoted pXlrl01L. Region L from pXlrl02 and pXlrl03, and region 'LM' from pXbrl01 were also subcloned in exactly the same manner.

4.1.b Subcloning region M

Region M was purified from the same gel of BamHI and EcoRI digested pXlrl01 as above. This region was to be inserted into the BamHI site of pBR322, again inactivating the tetracycline resistance gene. pBR322 was digested with BamHI and the linear fragment purified on a gel to ensure that no circular molecules remain. Ligation was performed with approximately 4 μg of M and 0.5 μg of pBR322. In this case it is possible for pBR322 to recircularise and produce transformants but such colonies can be distinguished because they are resistant to tetracycline as well as to ampicillin. Again many hundreds of transformants were obtained on ampicillin. Thirty colonies were tested on an ampicillin plate and a tetracycline plate, and while all grew on ampicillin only 16 grew on tetracycline. The fourteen colonies which grew only on ampicillin should be genuine recombinants. Four colonies were chosen and cleared lysis performed on 50 ml cultures. The DNA was electrophoresed on 1% agarose gels, undigested or after digestion with BamHI, as shown in figure 4.4. Colony 1 was selected for further use, and the plasmid was denoted pXlrl01M.

4.2. Mapping restriction sites in regions L and M

Boseley et al (1978), using the plasmid pXL212, mapped the sites for several restriction enzymes (figure 4.5). Of these enzymes the most

Figure 4.6 Mapping pX1r11L by partial digestion with restriction enzymes.

Approximately 1 μg (15,000 cpm) of region L, 5'-end labelled at the BamHI site, was added to 7 μg unlabelled pBR322 (or pX1r11L for SmaI digestion). The DNA was digested with approximately 4 units of either HaeII, TaqI or SmaI, taking time points at 2, 5, 15, 30 and 60 minutes. The reaction was stopped by adding agarose loading buffer and chilling to $0^{o}C$. Samples were electrophoresed on a 1% agarose horizontal gel, along with unlabelled marker DNA. The gel was stained to visualise the marker bands, then dried onto DE81 paper (Smith & Birnstiel, 1976) and autoradiographed.

1. 2 minute digestion.
2. 5, 15 and 30 minute digestion, pooled.
3. 60 minute digestion.

The distance travelled by the marker DNA fragments is plotted (O), and the size of digestion products can be estimated from the graph (H marks the position of a HaeII band, T is TaqI, and S is SmaI). Below is shown the restriction map which was constructed from these results.

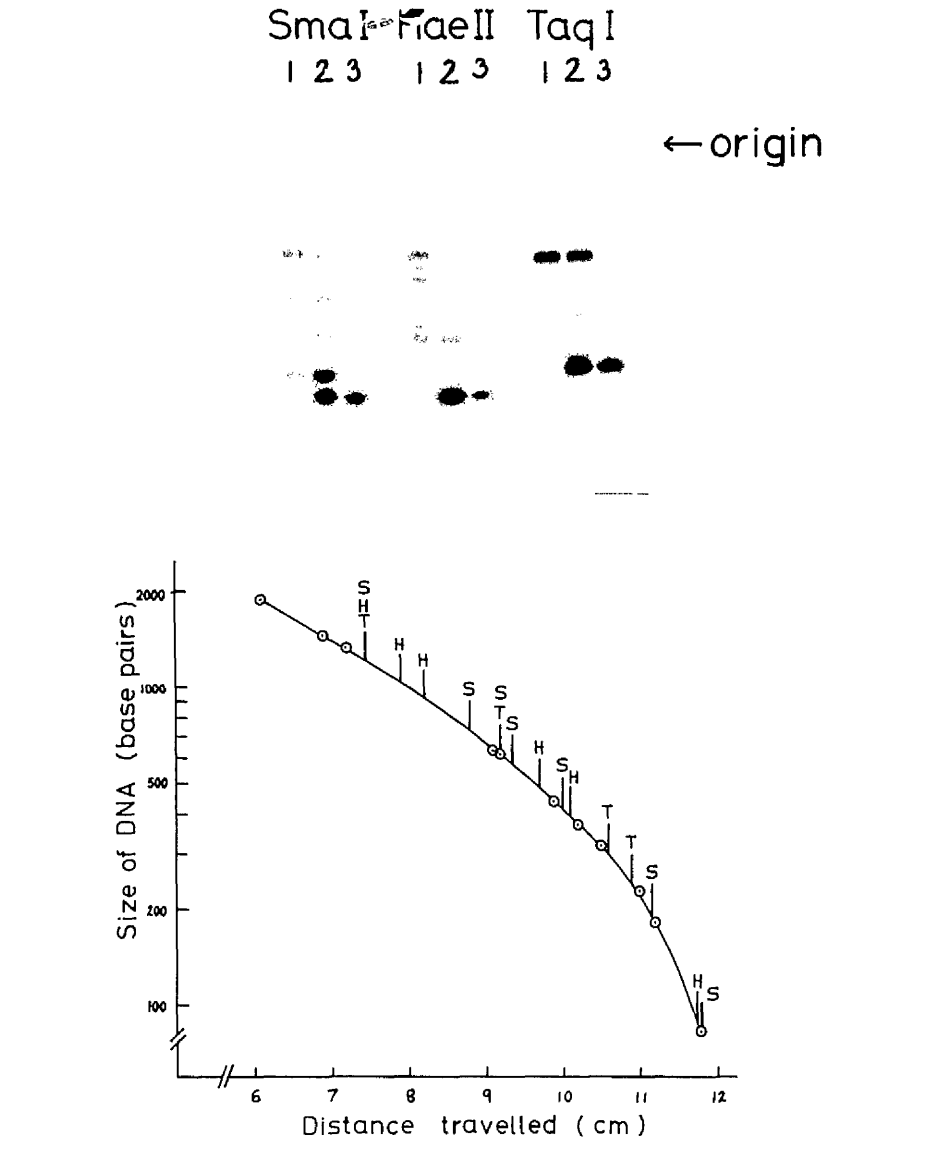

SmaI HaeII TaqI
1 2 3 1 2 3 1 2 3
←origin
Size of DNA (base pairs)
2000
1000
500
200
100
6
7
8
9
10
11
12
Distance travelled (cm)
HaeII
Taq I
SmaI
Bam HI
Bam HI
Bam HI

Figure 4.6

useful for sequencing are those which give several fragments of over 100 bp long, such as SmaI or HhaI. However 5' end-labelling is most efficient if the DNA is cut to yield a 5' 'protruding' end, unlike SmaI (CCC↓GGG) or HhaI (GCG↓C). It was therefore useful to map sites for AvaI (C↓PyCGPuG) and TaqI (T↓CGA), which both yield 5' protruding ends. All SmaI sites will be sites for AvaI, but there could also be sites for AvaI which are not sites for SmaI. Therefore to map sites for AvaI the first step was simply to compare the restriction fragments produced by digestion with the two enzymes. This experiment showed that the same fragments were produced when region L from pX1r11 was digested with either SmaI or AvaI. (At this time in the project I was planning to sequence pX1r11L; it was later decided to use pX1r101L, and pX1r101M, to fit in with other sequencing objectives in this laboratory, see 2.2, but restriction data from pX1r11L was still useful.)

Restriction sites for TaqI were mapped by partial digestion of end-labelled DNA, and at the same time the sites for HaeII and SmaI were mapped to see if these were the same as in pXL212. If DNA labelled at one end is subjected to partial digestion a series of labelled molecules are produced representing a cut at each of the restriction sites in the molecule. Gel electrophoresis followed by autoradiography reveals the length of each labelled fragment and hence the restriction sites can be mapped. To establish the conditions for partial digestion with each enzyme unlabelled material was used. A small quantity of region L, labelled at the BamHI site, was then added to unlabelled DNA and digested under the same conditions. Several time points were taken and electrophoresed on a 1% agarose gel, along with marker DNA of known lengths. The results are shown in figure 4.6.

This experiment confirms that sites for HaeII are probably

Figure 4.7 Mapping restriction sites in pXlr101M

Approximately 1 μg (20,000 cpm) of region M, 5'-end labelled at both ends, was digested with restriction enzymes as described below. The products were separated by electrophoresis in a 4% acrylamide gel (3.4) and stained with ethidium bromide. Lengths were estimated by comparison to marker bands. Bands were excised and the radioactivity estimated with a Mini-Monitor to check for the presence or absence of a labelled end.

1. 2 μg pBR322 digested with HpaII (markers)
2. 1 μg M digested with Sau3AI (1st & 3rd bands are radioactive)
3. " " " " TaqI (1st band radioactive)
4. " " " " Sau3AI + TaqI (1st & 3rd bands radioactive)
5. " " " " Sau3AI + HincII (2nd & 4th bands radioactive)
6. " " " " TaqI + HincII (1st & 2nd bands. radioactive)

The results show that TaqI does not have a recognition site within region M. Sau3AI has two sites yielding three fragments of which the smallest and largest are labelled so must be terminal. The larger fragment is cut with HincII, for which the position of the single recognition site is known, so the orientation of the Sau3AI sites must be as shown in the map. (Faint bands in all tracks are from contaminating vector DNA in the preparation of region M, which was separated from the vector on a sucrose gradient.)

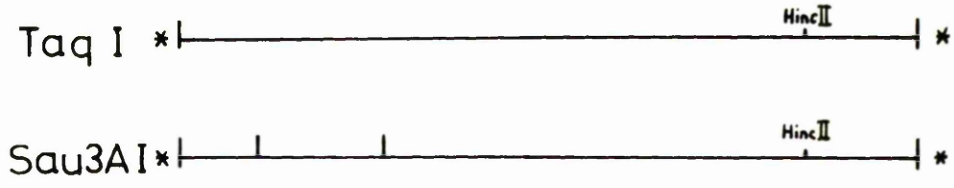


Figure 4.7

identical in pXlrll and pXL212, but suggests that the Sma sites may not be quite the same. Three sites for Taql were shown, two corresponding to predicted sites within the 5.8S gene and one lying in ITS1. The difference in sites for SmaI was later followed up as described in 5.6.

In the course of sequencing pXlrl01L sites for other enzymes were discovered and became useful for later experiments, notably the enzyme Sau3AI (see 4.4).

Before sequencing pXlrl01M further enzymes were again tested. Complete digestion of unlabelled plasmid with SmaI or AvaI showed that AvaI recognizes an extra site in pXlrl01M, cutting the 600 bp SmaI fragment into two shorter fragments. Sites for Sau3AI and TaqI were mapped by digesting end-labelled region M (labelled at both ends) with combinations of these enzymes and HincII (one site at a known position). The use of end-labelled material allowed the terminal fragments to be identified, and because few sites were present complete characterisation of the fragments was possible, as shown in figure 4.7.

4.3 Sequence determination

The complete sequence was derived from a large number of individual determinations on overlapping restriction fragments. In each experiment the first step was to purify a single restriction fragment. With a detailed restriction map of the region each fragment could be identified and treated separately. Thus an appropriate enzyme could be chosen for secondary digestion after labelling, making full use of every fragment. In the following sections preparation of fragments for sequencing is described.

The procedure for sequencing end-labelled DNA is discussed in 2.3 and described in 3.7. The length of sequence which can be read from

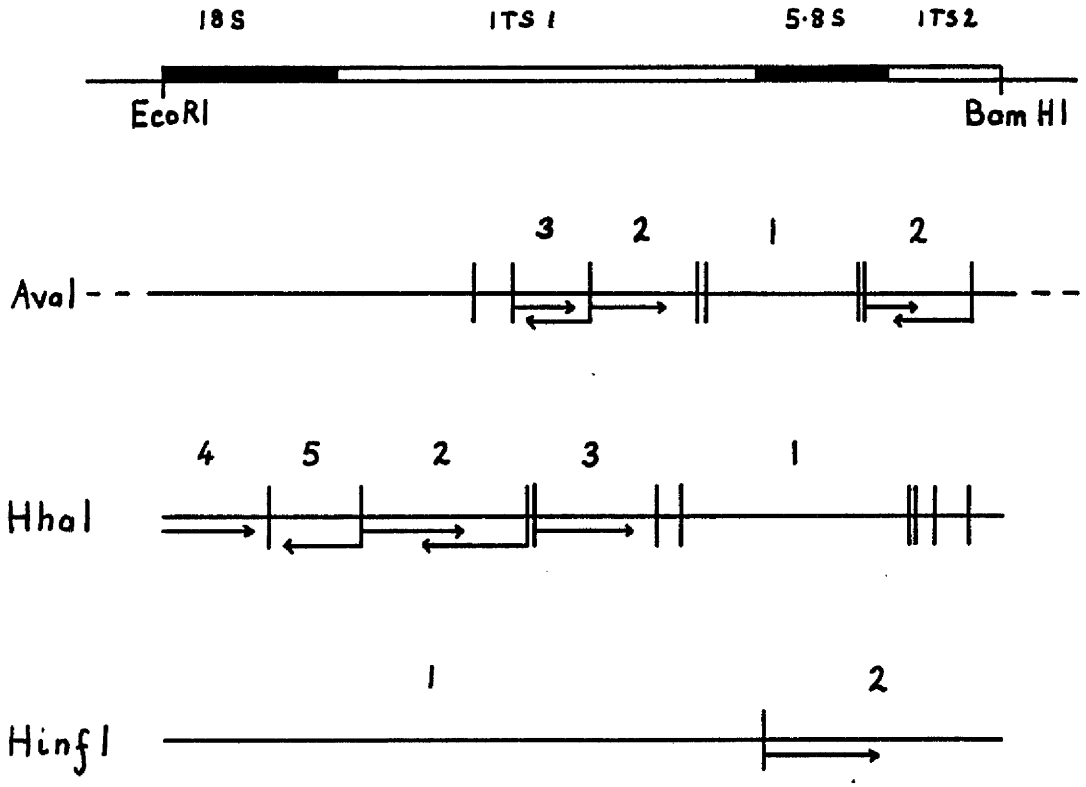


Figure 4.8 Strategy for sequencing region L: strand separation

The fragments obtained by digestion with various restriction enzymes are shown. Restriction fragments were 5'-end labelled then subjected to strand separation. The arrows show the length of sequence which could be determined from each fragment.

each gel depends on the resolution of individual gels and on the extent of attack of each chemical modification (as well as on the length of fragment). Problems which occurred in reading sequencing gels are discussed in 4.3.c.

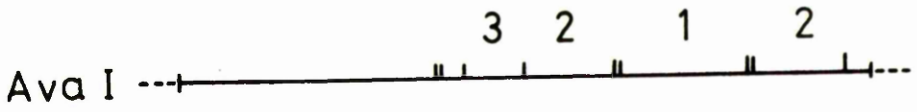
4.3.a Strand separation preparations

In the first of these preparations whole pXlr101L was digested with AvaI and the products separated on a 5% acrylamide gel (in later preparations 4% acrylamide slab gels were used). The three largest non-vector bands were excised and eluted. Band 2 apparently contained two fragments of the same size (see figure 4.9A). The DNA from each band was 5'-end-labelled, denatured, and separated on a non-denaturing 'strand separation' gel (3.6.c). After autoradiography it was seen that fragment 3 had separated into two bands, fragment 2 into four bands, but fragment 1 ran as a single band with the two strands failing to separate (figure 4.9). Each of the bands derived from fragments 2 and 3 was eluted and sequenced; the length of sequence which was determined from each band is shown in figure 4.8. Sequencing gels from the two strands of fragment 3 are shown in figure 4.9B, illustrating how the sequence is read and where the two strands match up.

Figure 4.8 shows other preparations in which strand separation was employed with some success. Region L, purified from pXlr101L, was digested with HhaI and the five major bands eluted. Each was labelled and strand separated. As shown in figure 4.10 the strand separation produced a complex pattern. Again the largest fragment ran as a single band, though a faint fast-moving band was also seen. Fragment 2 showed narrow separation between two major bands, with a third slower moving band which could have been double stranded material. Fragment 3 showed five predominant bands of variable intensity. Fragments 4 and 5 also produced several bands. After elution some bands were successfully

Figure 4.9 Preparation of AvaI fragments for sequencing : strand separation.

4.9A. 30 μg of pXlr101L was digested with AvaI. The map shows sites for AvaI within region L. (pBR322 has only one AvaI site.) The products were separated on a 5% acrylamide tube gel (i), and bands 1, 2 and 3 were excised and eluted. The DNA from each band was 5'-end labelled, then denatured and electrophoresed in a strand separation gel (3.6). The products of strand separation are shown (ii).



(ii)

(i)

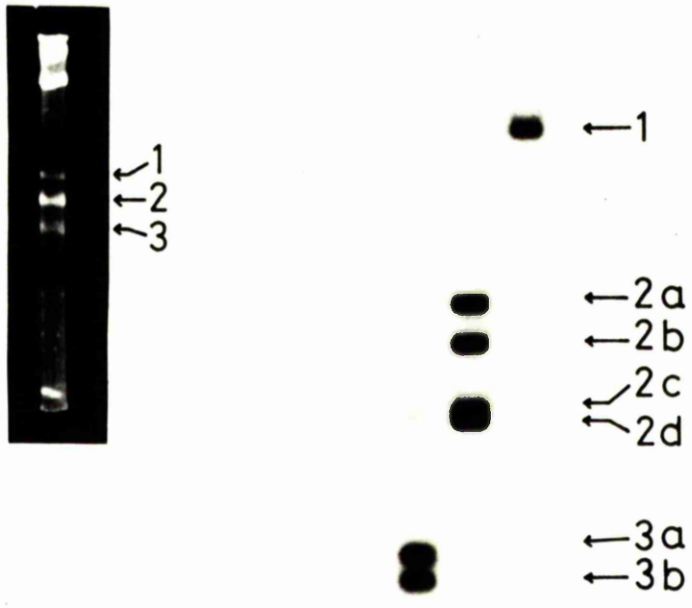


Figure 4.9A

4.9B. Each of the separated bands (from 4.9A(ii)) was subjected to sequence analysis by the method of Maxam and Gilbert (1980), as described in 3.7. The figure shows autoradiographs of sequencing gels from fragments 3a and 3b. In each case samples from the four reactions were loaded on a sequencing gel and electrophoresed until the xylene cyanol marker dye had travelled 20 cm (1). A second loading was made in the four adjacent wells and again the marker was allowed to travel 20 cm (2). A third and final loading was made and electrophoresis continued until the marker dye had travelled for 15 cm (3). Thus in (1) the marker dye has travelled for a total of 55 cm, and in (2) for 35 cm. The shortest molecules, representing the 5' end of the fragment, are resolved in (3), while in (1) the shortest fragments have moved off the end of the gel but the longer molecules are more clearly resolved.

The sequence which can be determined is shown for each loading. The numbers given correspond to the distance from the EcoRI site at the end of region L, and are derived from the completed sequence (figures 4.20 and 4.21). The overlap between successive loadings is illustrated by the numbers. Furthermore it can be seen that the two sequences are complementary, being derived from the opposite strands of a single restriction fragment.

N.B. Code for uncertain nucleotides - see figure 4.20.

Figure 4.9B

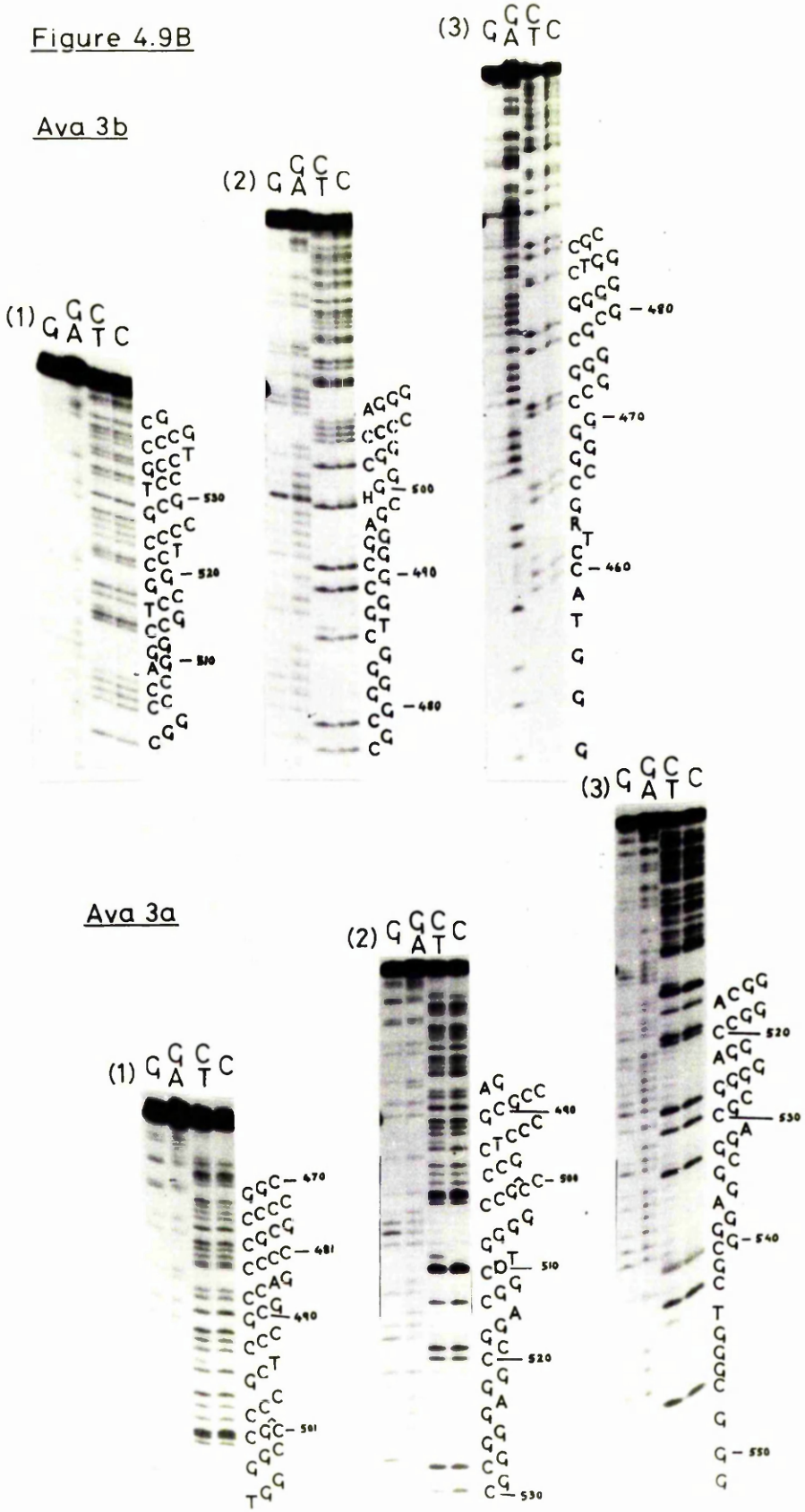


Figure 4.10 Preparation of HhaI fragments for sequencing : strand separation.

Approximately 5 μg of region L was digested with HhaI and the products separated by electrophoresis on a 4% acrylamide gel (i). The position of HhaI sites in region L is shown. (pBR322 also has many HhaI sites so region L was purified before digestion with HhaI.) Each fragment was eluted, 5'-end labelled and subjected to strand separation. The products of fragments 1, 2 and 3 are shown (ii).

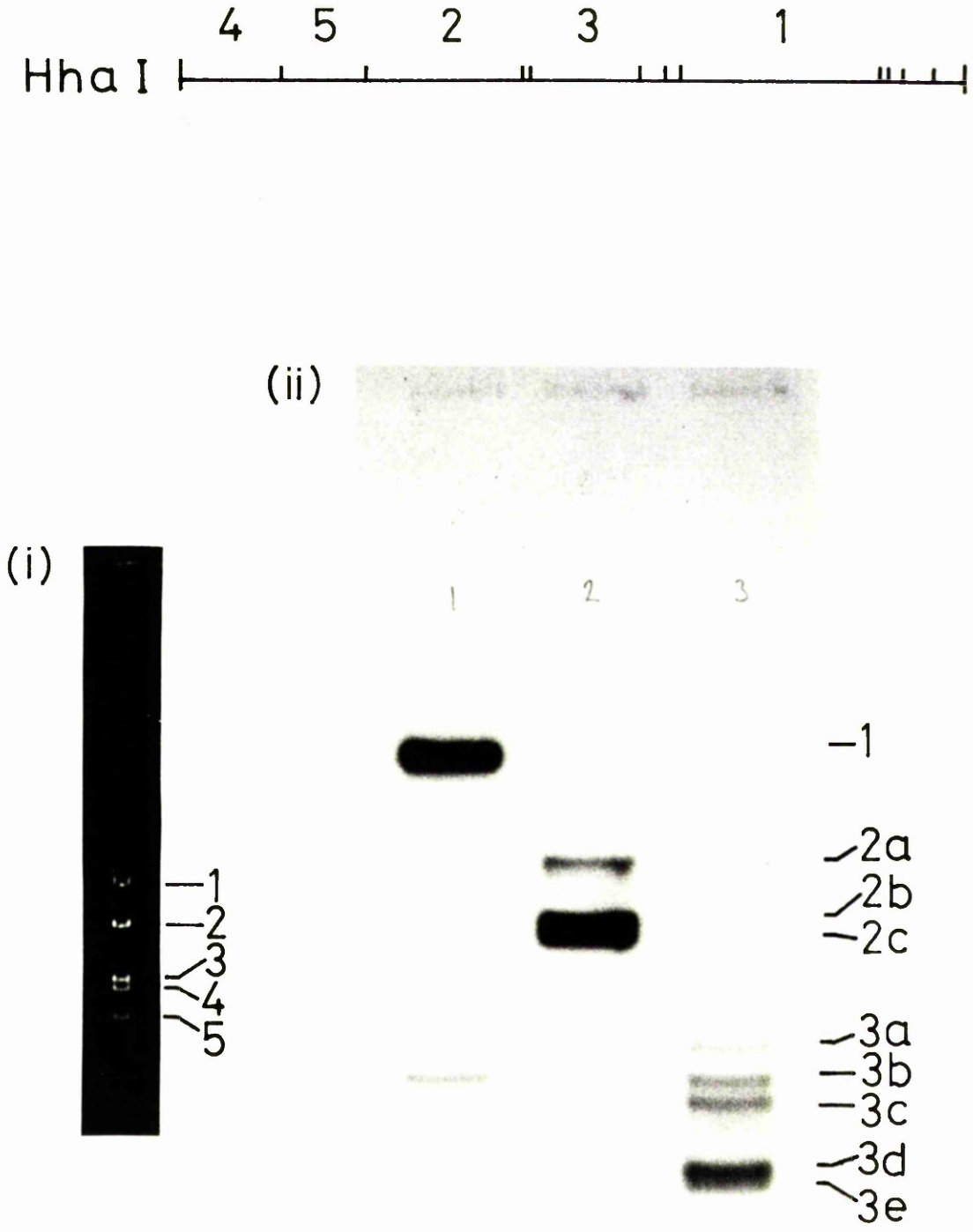


Figure 4.10

Figure 4.11 Preparation of HinfI fragments for sequencing : strand separation.

Approximately 5 μg of region L was digested with HinfI. Fragment 2 was prepared from a 4% acrylamide gel, then 5'-end labelled and subjected to strand separation. The products of strand separation are shown. Sequencing gels (15 cm loading only) from band 2a (i) and 2b (ii) are shown. Band 2a gives the sequence extending rightwards (as drawn) from the Hinf site. By comparison (iii) is a sequencing gel of Hinf2/Taq which is the sequence extending leftwards from the BamHI site. Careful examination shows that (ii) contains a mixture of the two strands. Numbering is according to the completed sequence (figure 4.21).

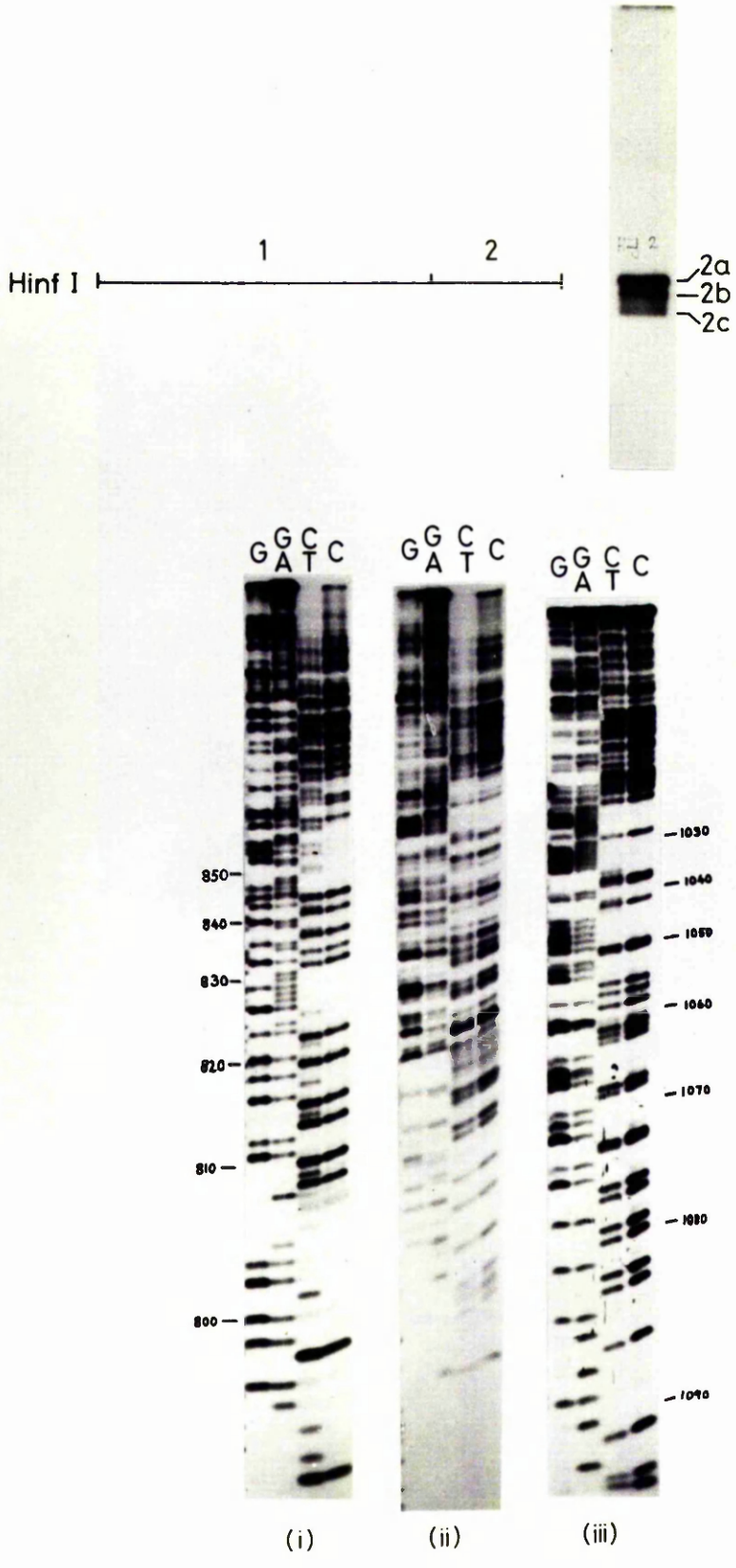

Hinf I
1
2
2a
2b
2c
G A
G T
C C
850
840
830
820
810
800
1030
1040
1050
1060
1070
1080
1090
(i)
(ii)
(iii)

Figure 4.11

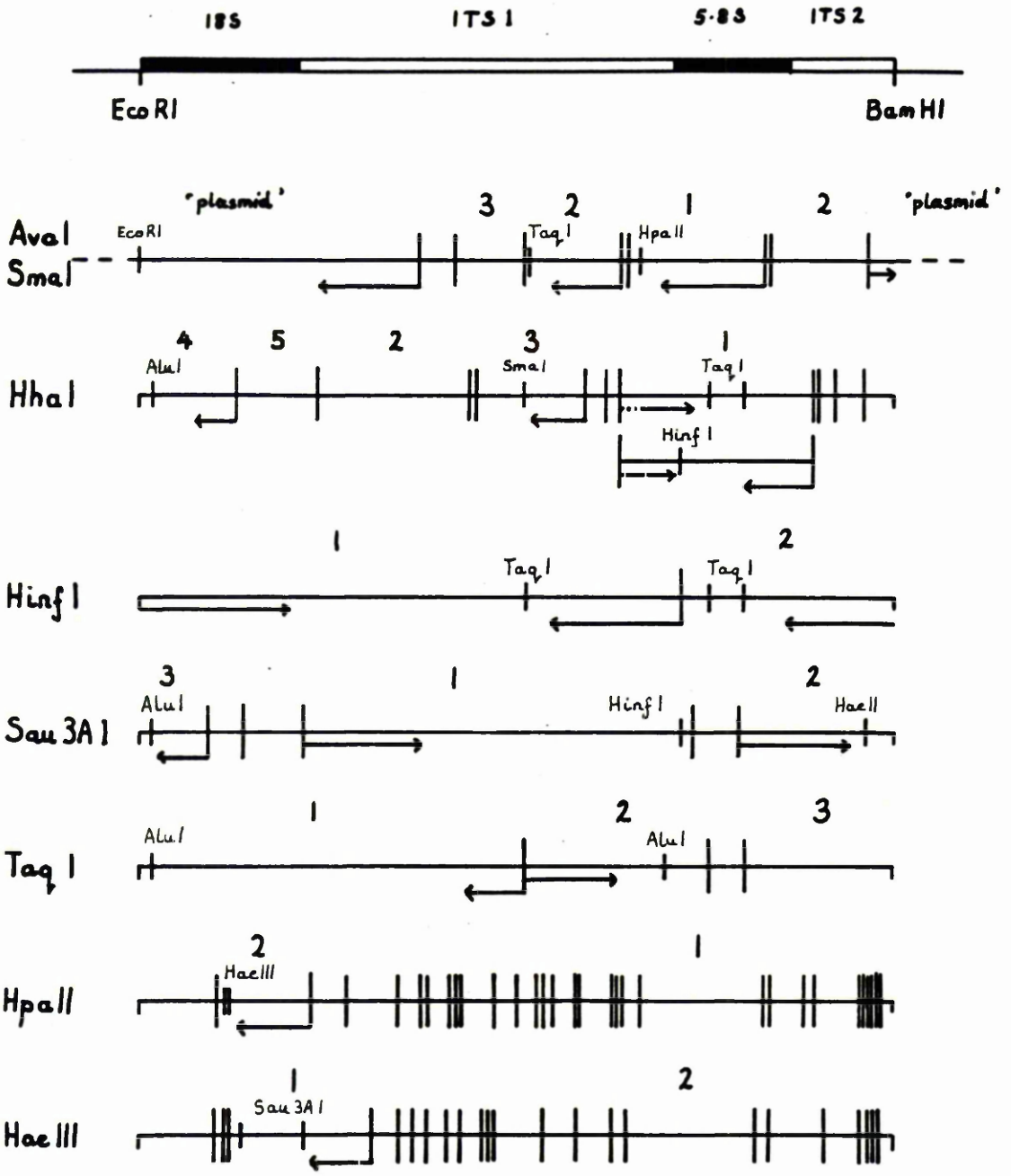


Figure 4.12 Strategy for sequencing region L: secondary digestion.

The fragments obtained by digestion with various restriction enzymes are shown. Restriction fragments were 5'-end labelled then digested with a second restriction enzyme as specified. The arrows show the length of sequence which could be determined from each fragment.

sequenced, others gave a sequence pattern which suggests that the DNA was a mixture of fragments, and the remainder were not sufficiently radioactive to sequence.

Strand separation of HinfI fragment 2 results in the resolution of three bands of different intensities (figure 4.11). In this case the slowest moving band gave a clear sequencing gel, while the second and third bands appeared to be impure. Figure 4.11 shows part of the sequencing gel from the slowest band together with a gel from the complementary strand derived from a later experiment. By comparison the gel from the second band is shown, and it can be seen that this band contained a mixture of both strands as the two gel patterns are superimposed.

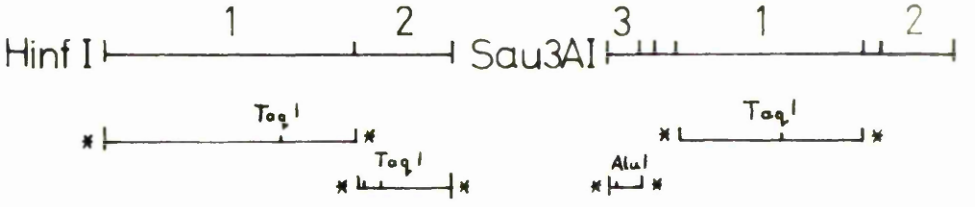
Since strand separation frequently gave unexpected results and involved the sequence analysis of several bands which turned out to contain a mixture of molecules I decided to concentrate on secondary restriction in later preparations. The reason for the appearance of several bands after strand separation of a single double-stranded molecule is not clear but it may be that the extremely high content of G + C in this region of the DNA creates some unusual secondary structure effects. Furthermore the location of bands in the gel relied on comparison to an autoradiograph so that when two bands were close together any inaccuracy in orientation of the autoradiograph could result in mixing of the bands.

4.3.b Secondary restriction preparations

The majority of fragments for sequence analysis were prepared by digesting a pure end-labelled DNA fragment with a restriction enzyme which cuts the fragment asymmetrically. After secondary restriction the fragments can again be separated on a 4% acrylamide gel and the bands

Figure 4.13 Preparation of Sau3AI and HinfI fragments for sequencing : secondary digestion.

4.13A. Approximately 5 μg of region L was digested with HinfI (H) or Sau3AI (S) and the products separated on a 4% acrylamide gel (i). The background bands are due to contaminating vector DNA. The required bands were eluted and 5'-end labelled. They were then digested with a second enzyme as illustrated, and the products separated on a second 4% acrylamide gel (ii). The products were located by staining with ethidium bromide (Sau3/Alu is small and faint but could just be seen on the original gel.)


Hinf I
1
2
Sau3AI
3
1
2
Taq I
Taq I
Taq I
AluI

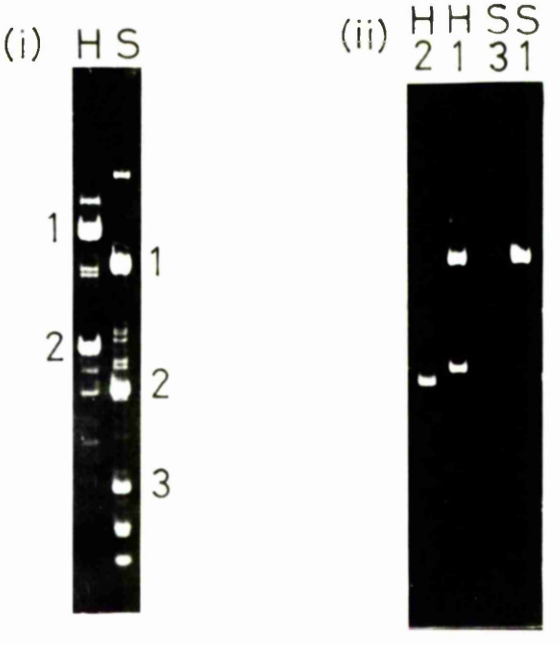

(i) H S
1
2
1
2
3
(ii) H H S S
2 1 3 1

Figure 4.13A

4.13B. Sequencing gels of the two products of HinfI/Taq are shown. Three successive loadings of each fragment were made. Numbers correspond to the completed sequence (figure 4.21), and show the extent over which each gel could be read (see also figure 4.20).

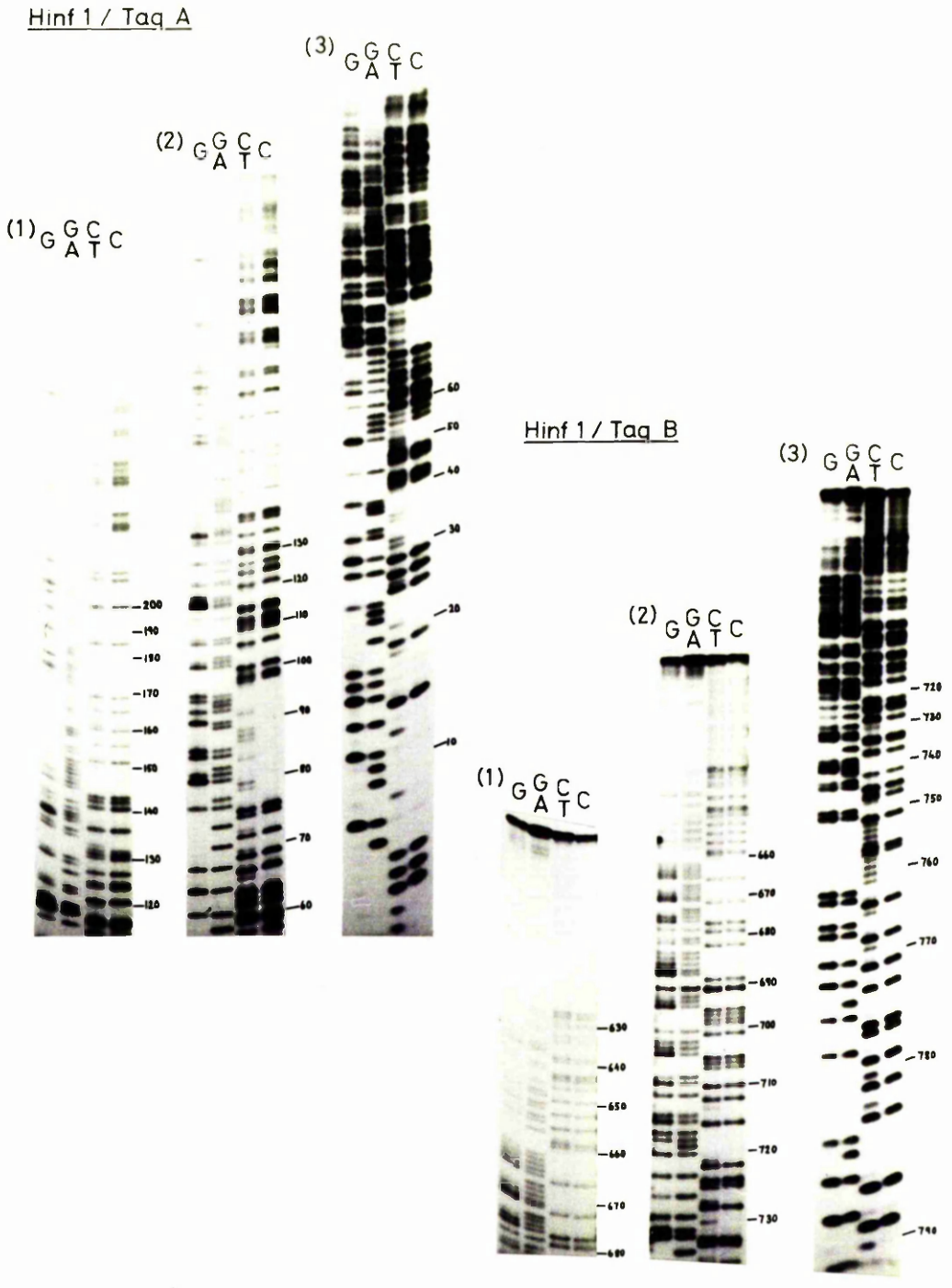

Hinf 1 / Taq A
(1) G A T C C
(2) G A T C C
(3) G A T C C
Hinf 1 / Taq B
(1) G A T C C
(2) G A T C C
(3) G A T C C

Figure 4 13B

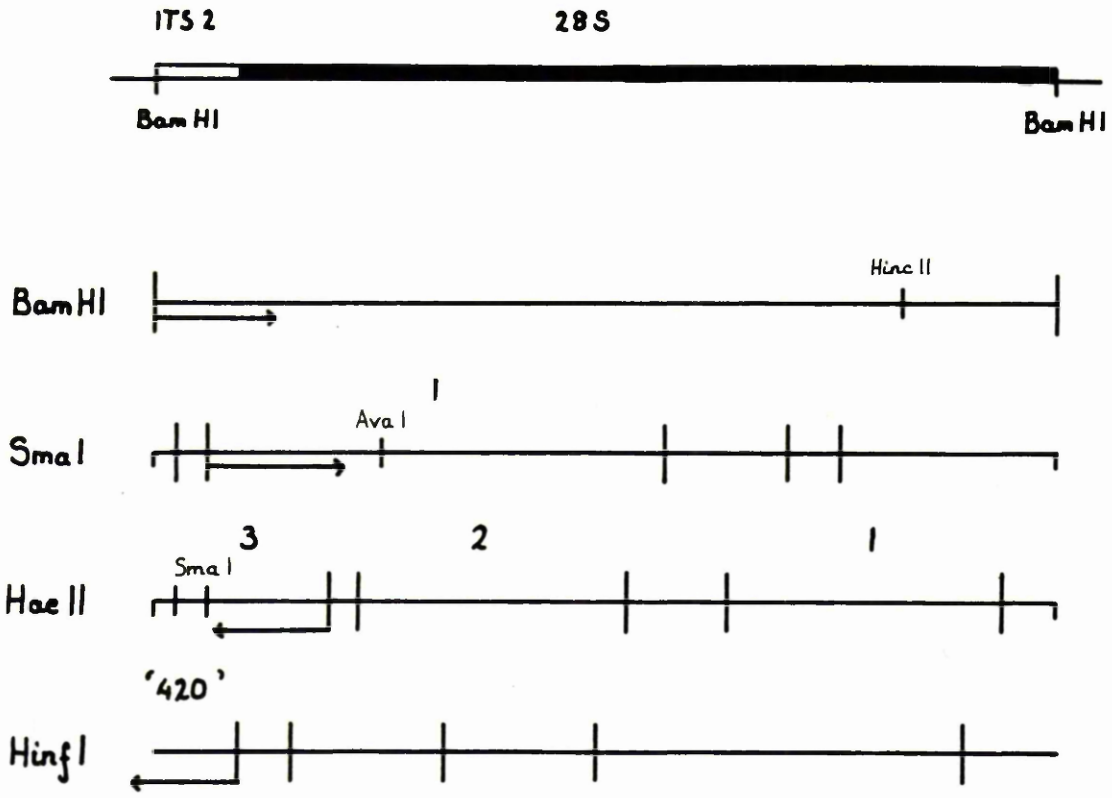


Figure 4.14 Strategy for sequencing region M.

The fragments obtained by digestion with various restriction enzymes are shown. Restriction fragments were 5'-end labelled then digested with a second restriction enzyme as shown. The arrows show the length of sequence which could be determined from each fragment. For explanation of Hinf '420' see figure 4.15.

located directly by staining with ethidium bromide. Figure 4.12 shows fragments which have been prepared in this way and the length of sequence determined.

In one experiment region L was digested with Sau3AI or with HinfI, and the products separated on a 4% acrylamide gel. Each of the required fragments was end-labelled, digested with an appropriate restriction enzyme, and separated on a second gel as shown in figure 4.13A. Two of the sequencing gels obtained are also shown (figure 4.13B). As more of the sequence was determined the exact positions of restriction sites became known and sites for previously untested enzymes were discovered. Thus it was possible to find ways of filling in remaining gaps.

The left hand part of region M was sequenced using the same approach, as shown in figure 4.14.

The sequence of regions L and M was derived from two plasmids containing adjacent DNA fragments from pXlr101. The two regions are separated by a site for BamHI. To ensure that the two regions really are adjacent and are separated by a single BamHI site I decided to sequence through this site using a DNA fragment derived from the parent plasmid pXlr101. Examination of the restriction map suggested that a HinfI fragment of 420 bp would be suitable, if it could be purified. Whole pXlr101 was digested with HinfI, and separated on a 4% acrylamide gel with appropriate marker fragments. By a stroke of good fortune one fragment of 420 bp was clearly separated in spite of the large number of fragments produced (figure 4.15). This fragment was eluted, end-labelled, digested with AluI and the larger product sequenced. Figure 4.15 shows that regions L and M must be contiguous, linked by a single BamHI site.

Figure 4.15 Bridging the BamHI site

The map shows sites for HinfI within regions L and M, and the position of the BamHI site (marked B) between them.

60 μg of pXlrl01 was digested with HinfI and the products separated by electrophoresis on a 4% acrylamide gel (1). pBR322 digested with HpaII was also electrophoresed to provide markers (2). A marker of 400 bp can be used to locate the 420 bp HinfI fragment. The fragment was eluted, labelled and digested with AluI. The larger product was subjected to sequence analysis. A detail of the sequencing gel is shown, illustrating the position of the single BamHI site. Unfortunately bands in the 'G-only' track are very faint, but the sequence can still be fitted unambiguously with the sequences of L and M (see also figure 4.20).

N.B. Code for uncertain nucleotides - see figure 4.20. Numbers correspond to the completed sequence (figure 4.21).

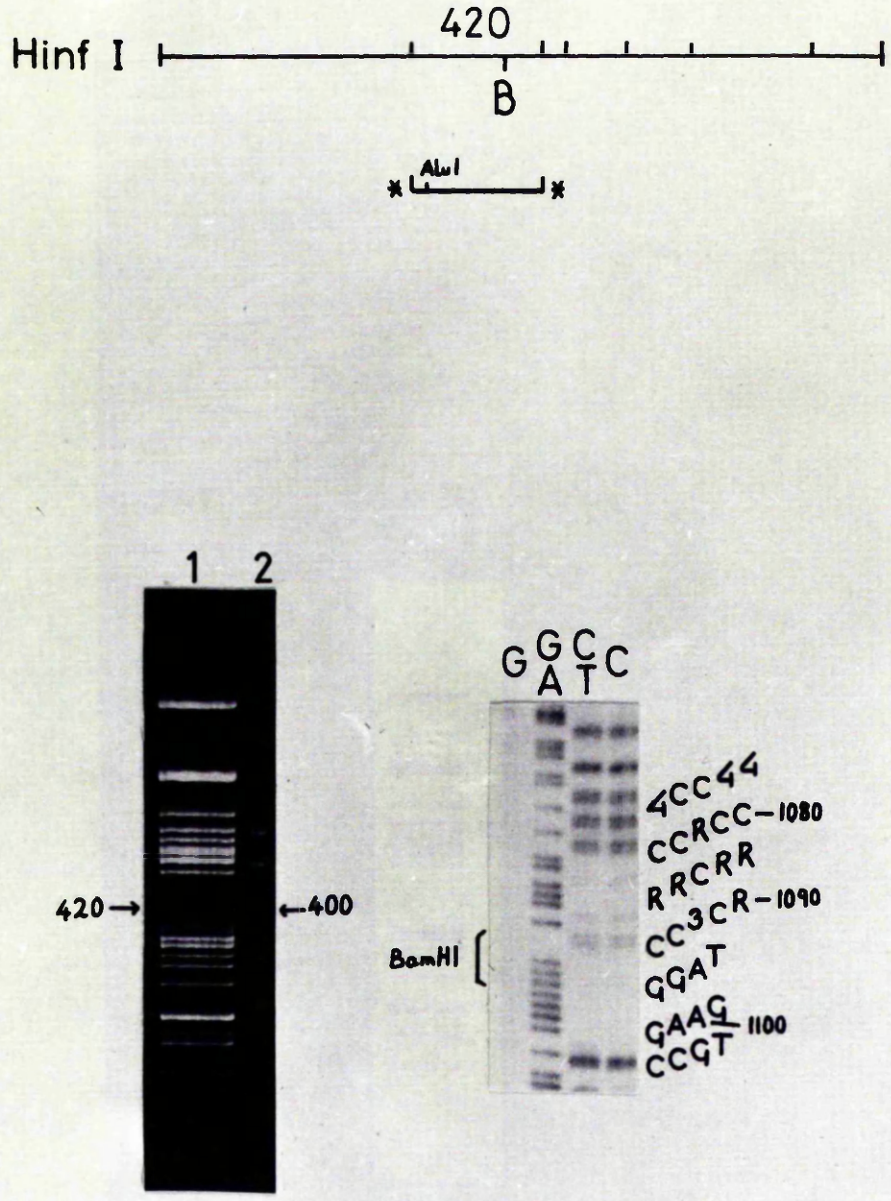

Hinf I
420
B
AlwI
1
2
420
400
BamHI
G
G
A
C
T
C
4CC4
CCRCC-1080
RRCRR
CC3CR-1090
GGAT
GAAG
CCGT
1100

Figure 4.15

4.3.c Interpretation of sequencing gels

A number of sequencing gels are shown in figures 4.9B, 4.11 and 4.13B. In many cases the gels can be read unambiguously up to the point where the bands are no longer resolved. However in several sequence determinations problems have occurred which make parts of the gel difficult to interpret.

In the C + T reaction hydrazine frequently reacts more strongly with cytosine than with thymine, even in the absence of sodium chloride. The result is that T bands are fainter than C bands in the C + T track. Furthermore faint bands are sometimes seen in the pyrimidine tracks at the position of every G residue; this is because hydrazine reacts with guanine under conditions of lowered pH, probably when sodium acetate and ethanol are added to precipitate the DNA after modification (Maxam & Gilbert, 1980). The problem can be minimised by using fresh hydrazine. Once these apparent inconsistencies are understood they do not generally affect the interpretation of a sequencing gel.

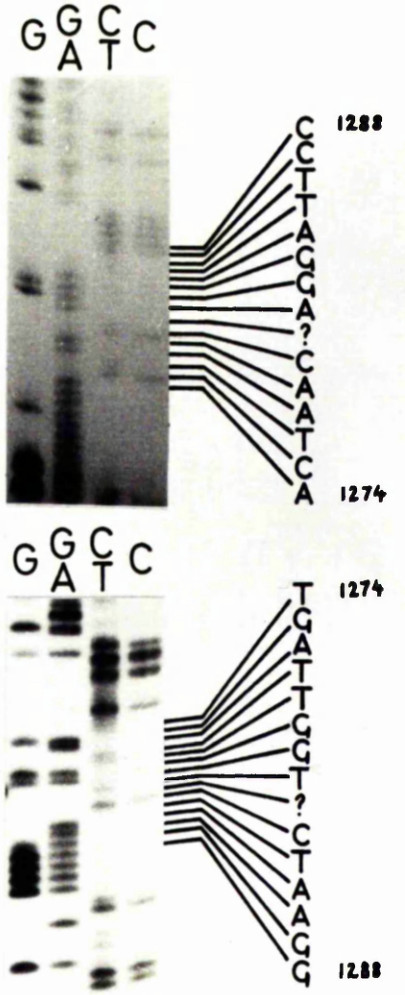
On several occasions reaction of dimethylsulphate with guanine was more extensive than the other reactions so that bands in the 'G only' track became faint at higher molecular weights. This problem occurred mainly in early determinations when a longer reaction time was used, but continued to occur occasionally, perhaps due to the difficulty of accurately measuring 0.5 µl of dimethylsulphate. Thus in parts of some gels it is not possible to distinguish between guanine and adenine. Where this was the case the identity was always confirmed on at least one other gel.

A second problem in early experiments was caused by incorporation of bromophenol blue in the loading dye for sequencing gels. At the position of the bromophenol blue dye band on a gel the DNA bands are blurred, occasionally causing one band to spread out on each side of the

Figure 4.16 Location of methylated cytosines

The two details from sequencing gels are from the opposite strands of the same region of DNA. On each strand there is a gap in the pattern of bands, where on the opposite strand a guanine is found. When the sequence of the two strands is put together it is seen that the gaps must correspond to methylated cytosines in an EcoRII recognition site.

Details are from fragments M Sma/Ava and M Hae/Sma (figures 4.14 and 4.20). Numbers correspond to the completed sequence (figure 4.21).


G G C C
A T
1288
CCTTAGGA?CAATCA
1274
G G C C
A T
1274
TGATTCGT?CTAAGG
1288


1274
1288
5' ACTAACCAGGATTCC 3'
3' TGATTGGTCCTAAGG 5'

Figure 4.16

dye band. In figure 4.11, gel (i), bands 805 and 806 are both blurred, and slightly displaced, by the bromophenol blue. Subsequently the loading buffer was made up without bromophenol blue and the problem was eliminated.

A further anomaly occurs when a methylated cytosine residue is present in the DNA, as this fails to react with hydrazine. It is usually possible to see from the spacing that a band is missing, but often it is first detected by the presence of an extra G band on the opposite strand (see figure 4.16). The methyl cytosines found were at a site which if unmethylated would be recognised by EcoRII; this could be confirmed by digesting with BstNI which recognises the same site whether or not it is methylated ($\overset{me}{C}CAGG$ / $GGT\overset{me}{C}C$). Two other recognition sites for BstNI were not methylated.

However the most persistent and unavoidable problem was the appearance in specific regions of certain gels of unevenly spaced bands, often with several bands compressed together, making it difficult or impossible to tell the number and order of nucleotides. This distortion is due to strong local secondary structure which is not fully eliminated even under the denaturing conditions of the gel. For example within ITS1 there is a sequence 5'CCCCCCCCCCACGGGGGGGGGG 3' which will form a very stable hairpin. On a sequencing gel the G bands become very compressed and cannot be counted. When the other strand is sequenced it is again the G bands which are compressed, but the C bands corresponding to G on the opposite strand can be counted (in practice the situation is slightly more complicated, see figure 4.19). In other words the position of the secondary structure effect is different on the two strands (occurring with bands on the 3' side of a hairpin) and so the complete sequence can be deduced. Figures 4.17, 4.18 and 4.19 illustrate three of the most persistent and extensive regions of secondary structure and show how the sequence for these regions was resolved.

Figure 4.17 Secondary structure effects : 1

Details from sequencing gels from opposite strands of the same region of DNA are shown. The gels are orientated in opposite directions so that they can be compared directly. The sequence of the DNA is written out and related to the gels. It can be seen that bands which are very compressed on one strand are well resolved on the opposite strand. Numbers correspond to the completed sequence (figure 4.21). Details are from fragments Hha3ss and Hha3/Sma (see figures 4.8, 4.12 and 4.20).

Figure 4.18 Secondary structure effects : 2

As for figure 4.17. Details are from fragments Taq2/Alu and Hinf1/Taq (see figures 4.12 and 4.20).

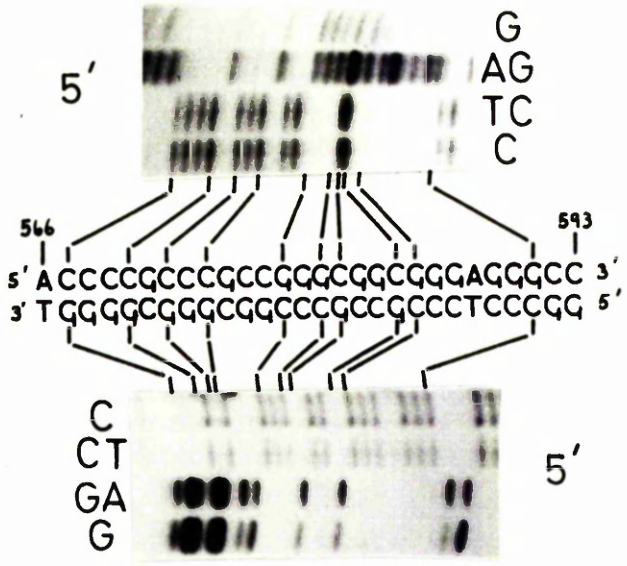


Figure 4.17

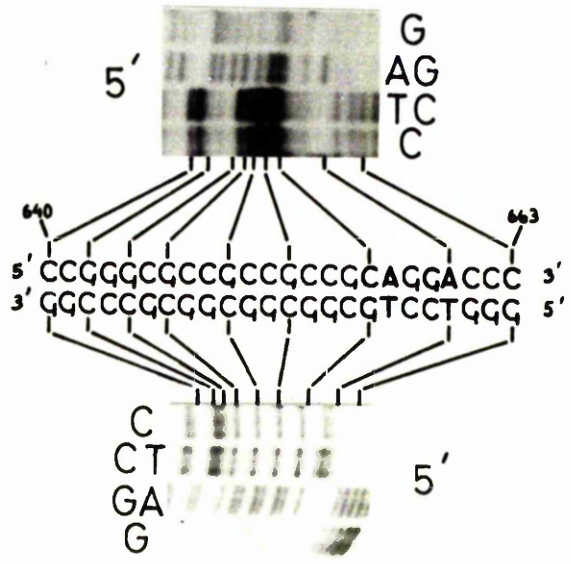


Figure 4.18

Figure 4.19 Secondary structure effects : 3

Details of four sequencing gels are shown, two from each strand of the DNA of one region. Again the gels are orientated in opposite directions so that they can be compared. In each gel a slightly different pattern of compression of bands is seen. From a combination of the four gels the sequence of the region can be deduced. One problem remains : gel (c) shows two G bands at position 310, while the three remaining gels (plus a fourth from a later experiment - see figure 6.6) clearly have only one band. The reason for this anomaly is not understood, but it was considered most likely that the extra band is an artifact as it could not be confirmed on either the same or the opposite strand.

Details are (a) from Hha2ss, (b) from Saul/Hinf, (c) from Avap1/RI and (d) from HaeIII1/Sau (see figures 4.8, 4.12 and 4.20). Numbers are as in the completed sequence (figure 4.21).

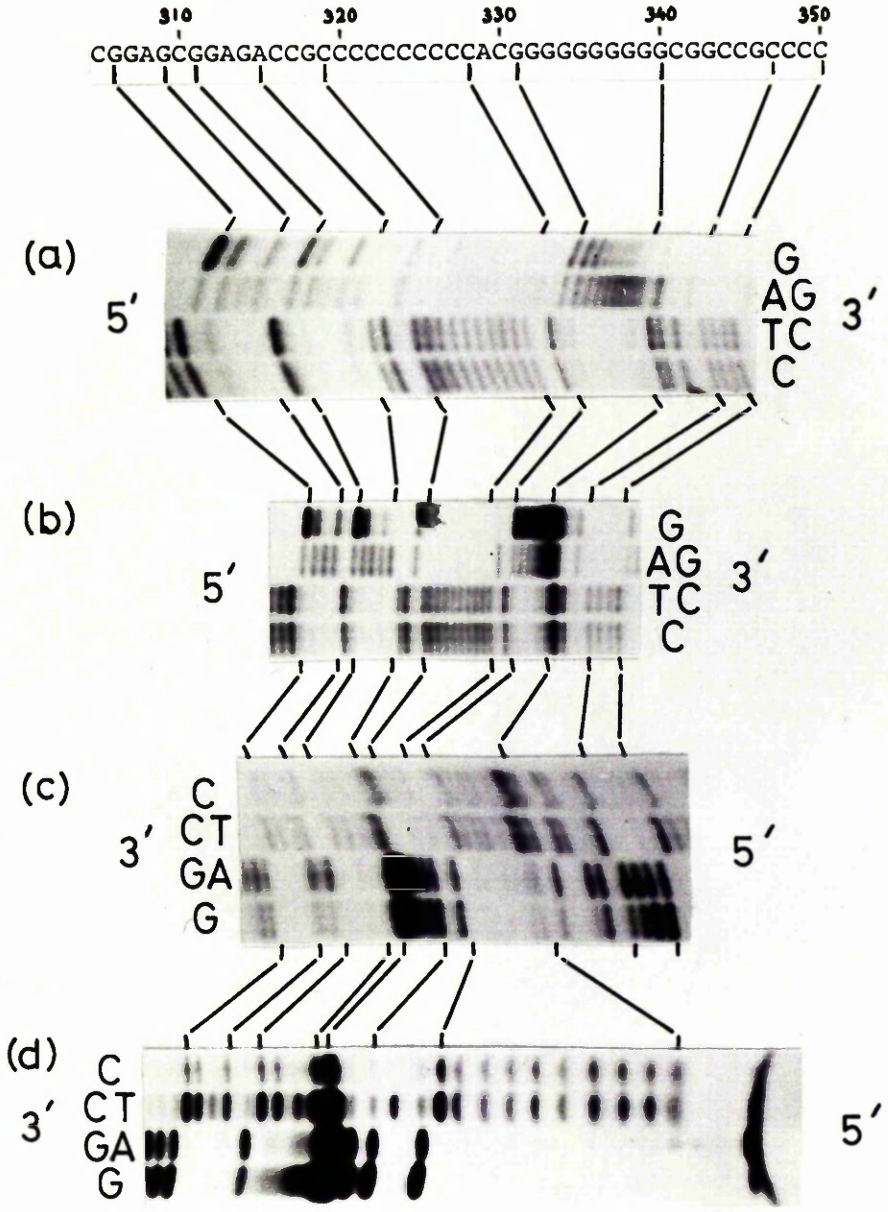

310
320
330
340
350
CGGAGCGGAGACCGCCCCCCCCCCACGGGGGGGGGGCGGCCGCCCC
(a)
5′
G
AG
TC
C
3′
(b)
5′
G
AG
TC
C
3′
(c)
3′
C
CT
GA
G
5′
(d)
3′
C
CT
GA
G
5′

Figure 4.19

As parts of the sequence were determined they were lined up with previous determinations from the same and the opposite strand. Figure 4.20 shows how the individual determinations were fitted together to produce the complete sequence. It can be seen that except for a few very short stretches the sequence was determined on both strands throughout, and with extensive overlaps.

4.4 Locating the gene-spacer boundaries

4.4.a The 18S gene

The 3' terminal T1 oligonucleotide of 18S rRNA is known to be (G)AUCAUUA$_{OH}$ in many eukaryotes, including Xenopus (Vass & Maden, 1978; Shine & Dalgarno, 1974; Eladari & Galibert, 1976). The corresponding DNA sequence, GATCATTA, occurs only once, at 220-227 nucleotides from the EcoRI site (see figure 4.21). The preceding nucleotides also agree exactly with the published sequence of 38 nucleotides at the 3' end of 18S rRNA from rat liver, confirming this location (Alberty et al, 1978).

Brand and Gerbi (1979) had previously estimated that the 3' end of the 18S gene was 310 bp from the EcoRI site (determined as the length of single stranded DNA protected from S1 nuclease digestion by hybridisation to 18S rRNA). The plasmid used by Brand and Gerbi was pX1r11, which contains rDNA from a different source from pX1r101 (see Table 3.1), so it was possible that one of the plasmids contained an insertion or deletion in the 18S gene. The presence of a recognition site for Sau3AI (GATC) in the terminal octanucleotide of the 18S gene enabled me to test this hypothesis. Digestion of region L from the two plasmids with Sau3AI yielded identical restriction fragments, so a major length difference cannot exist.

4.4.b The 5.8S gene

The 5.8S gene has previously been sequenced by Boseley et al

Figure 4.20 Compilation of all sequence determinations

The sequence which was read from each gel is written out and matched up with other sequence determinations. The sequence which was deduced from the combined determinations is written above in larger letters, showing the sequence of the strand synonymous to RNA (ie the non-coding strand).

Fragment nomenclature: Fragments are named according to the restriction enzyme with which they were cut, the number of the fragment and the second enzyme if used. Thus Sau3/Alu was produced by labelling the third largest Sau3AI fragment and digesting with AluI (restriction enzyme numbers are omitted except for HaeII and HaeIII). Hha4ss was produced by strand separation of the fourth largest Hha fragment. Fragment names preceded by M are fragments derived from region M.

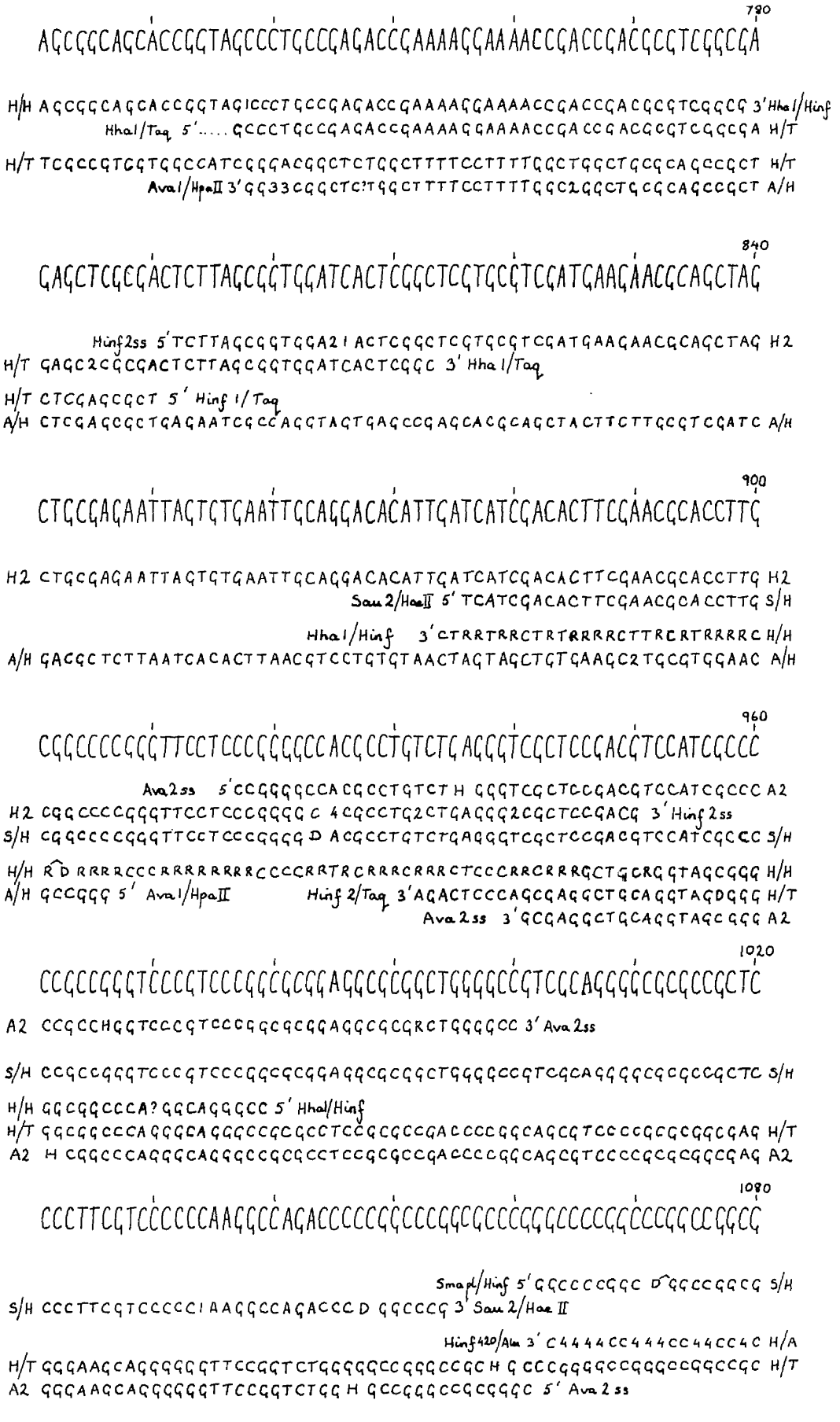
The origin of each fragment can be traced by reference to figures 4.8, 4.12 or 4.14.

Symbols for uncertain nucleotides: Where there was uncertainty about the identity of a band various symbols were used as described below (adapted from Staden, 1979):-

1 = probably C	R = A or G	X = any nucleotide
2 = probably T	Y = C or T	? = presence uncertain
3 = probably A	D = C, possibly 2 or more	^ = order uncertain
4 = probably G	H = G, possibly 2 or more	


60
TTCCCAGTAAGTGCGGGTCATAAGCTCGCGTTGATTAAGTCCCTGCCCTTTGTACACACC
Hha4ss 5' TTCCCAGTAAGTGCGGGTCATAAGCTCGCGTTGATTAAGTCCCTGCCCTTTGTACACACC H4
Hinf1/Taq 5' TTCCCAGTAAGTGCGGGTCATAAGCTCGCGTTGATTAAGTCCCTGCCCTTTGTACACACC H/T
Sau3/Alu 3' GCGCAACTAATTCAGGGACGGGAAACATGTGTGG S/A
120
GCCCGTCGCTACTACCGATTGGATGGTTTAGTGAGGTCCTCGGATCGGCCCCGCCGGGGT
H4 GCCCGTCGCTACTACCGATTGGATGGTTTAGTGAGGTCCTCGGATCGGCCCCGCC H GGT H4
H/T GCCCGTCGCTACTACCGATTGGATGGTTTAGTGAGGTCCTCGGATCGGCCCCGCCGGGGT H/T
S/A CGGGCAGCGATGATGGCTAACCTACCAAATCACTCCAGGAGCCT 5' Sau3/Alu
Hha4/Alu 3' AACCTACCAAATCAGTCCAGGAGCCTAGCCGGGGCGGCCCCA H/A
180
CGCCCACGGCCCTGGCGCGACGCCCGACGAAGACGATCAAACTTGACTATCTACAGGAAGTA
H4 CGGCCAC 3' Hha 4 ss
H/T CGGCCACGGCCCTGGCGGAGCGCCGAGAAGACGATCAAACTTGACTATCTAGAGGAAGTA H/T
Hpa2/HaeIII 3' GA D H CCTCGCGGCTCTTCTGCTAGTTTGAACTGATAGATCTCCTTCAT H/H
H/A GCCGGTGCCGGGACCGCC 5' Hha 4/Alu Hha5ss 3' GYYYGAACTGATAGA2CYCCTTCAY H5
240
AAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTAACGAGACCCCCCT
Sau1/Hinf 5' CATTAACGAGACCCCCCT S/H
H/T AAAGTCGTAACAAGGTTTCCGTAGGTGAACC2GCGGAA?GGATCATTAACG 3' Hinf1/Taq
H/H TTTCAGCATTGTTCCAAAGGCATCCACTTGGACGCCTTCCTAGTAATTGCTCTGGGGGGA H/H
H5 Y22CAGCAY2G?YTCC3333GCA22CCACTTGGACGCCTTCCTAGT4ATTGCTCT?GGGGGGX H5
300
CACCCGGAGAGAGGGAAGGCGCCCGCCGCACCCTCCCCGCGGAGAGAGAGAGAGACGCCC
S/H CACCCGGAGAGAGGGAAGGCGCCCGCCGCACCCTCCCCGCGGAGAGAGAGAGAGACGCCC S/H
Hha2ss 5'CCC??CCGXCXCCCTCCCCHCGGAGAGAGAGAGAGACGCCC H2
HaeIII1/Sau
H/H GTGG 5' Hpa2/HaeIII 3'TTCCGCGGGCGGCGTGGGAGGGGCGCCTCTCTCTCTCTCTGCGGG H/S
H5 42GGGCCTC2CTCCC22C 5' Hha5ss
AvaI/RI 3' 2GGG D CCCDTCCCT CCGCGGGCGGCGTGGGAGGGGCGCCTCTCTCTCTCTCTGCGGG A/RI
360
GCCCCGCAGCCGAGACCGCCCCCCCCCACGCGCGGGGGGGCGCCCCCCCGAAAGGGACG
S/H GCCCCGCAGCGGAGACCGCCCCCCCCCACGGGGGG H D G CGCCCCGAAAGGGACG S/H
H2 GCCCCGGAGDGGAGACCGCCCCCCCCCACGGGGGGG44C H CCGCCCCG333GGG3CG H2
H/S CGGGGCCTCGCCTC D Hn GGTGCCCCCCCCCCGCC 5' HaeIII1/Sau
Hha2ss 3'CCCGCCGGGGGGC ? CCC?GC H2
A/RI CGGGGCCTHCGCCTC C Hn GGGGTGCCCCC D CCGCCGGCGGGGCTTTCCCTGC A/RI


420
480
540
600
660
720
S/H
H2
A/RI
A3
T/A
H/T
H/S
S/T
Sau I/Hinf
Ava3ss
Hha3ss
Ava2ss
Taq2/Alu
Hinf I/Taq
Hha3/Sma
Sma2/Taq
Hha I/Hinf


780
AGCGGCAGCACCGGTAGCCCTGCCGAGACCGAAAAGGAAAACCGACCGACGCGTCGGCGA
H/H AGCGGCAGCACCGGTAGICCCTGCCGAGACCGAAAAGGAAAACCGACCGACGCGTCGGCG 3'HhaI/Hinf
HhaI/Taq 5'..... GCCCTGCCGAGACCGAAAAGGAAAACCGACCGACGCGTCGGCGA H/T
H/T TCGCCGTCGTGGCCATCGGGACGGCTCTGGCTTTTCCTTTTGGCTGGCTGCGCAGCCGCT H/T
AvaI/HpaII 3'GG33CGGCTC?TGGCTTTTCCTTTTGGC2GGCTGCGCAGCCGCT A/H
840
GAGCTCGCGACTCTTAGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCAGCTAG
Hinf2ss 5'TCTTAGCGGTGGA2!ACTCGGCTCGTGCGTCGATGAAGAACGCAGCTAG H2
H/T GAGC2CGCGACTCTTAGCGGTGGATCACTCGGC 3' HhaI/Taq
H/T CTCGAGCGCT 5' Hinf I/Taq
A/H CTCGAGCGCTGAGAATCGCCAGCTAGTGAGCCGAGCACGCAGCTACTTCTTGCGTCGATC A/H
900
CTGCGAGAATTAGTGTGAATTGCAGGACACATTGATCATCGACACTTCGAACGCACCTTG
H2 CTGCGAGAATTAGTGTGAATTGCAGGACACATTGATCATCGACACTTCGAACGCACCTTG H2
Sau2/HaeIII 5' TCATCGACACTTCGAACGCACCTTG S/H
HhaI/Hinf 3'CTRRTRRCTRTRRRRCTTRCRTRRRRC H/H
A/H GACGCTCTTAATCACACTTAACGTCCTGTGTAACTAGTAGCTGTGAAGC2TGCGTGGAAC A/H
960
CGGCCCCGGGTTCCTCCCGGGGCACGCCTGTCTGAGGGTCGCTCCGACGTCCATCGCCC
Ava2ss 5'CCGGGGCCACGCCTGTCT H GGGTCGCTCCGACGTCCATCGCCC A2
H2 CGGCCCCGGGTTCCTCCCGGGG C 4CGCCTG2CTGAGGG2CGCTCCGACG 3' Hinf2ss
S/H CGGCCCCGGGTTCCTCCCGGGG D ACGCCTGTCTGAGGGTCGCTCCGACGTCCATCGCCC S/H
H/H R^D RRRRCCCRRRRRRRRCCCCRRTRCRRRCRRRCTCCCRRCRRRGCTGCRGGTAGCGGG H/H
A/H GCCGGG 5' AvaI/HpaII Hinf 2/Taq 3'AGACTCCCAGCGAGGCTGCAGGTAGDGGG H/T
Ava2ss 3'GCGAGGCTGCAGGTAGCGGG A2
1020
CCGCCGGGTCCCGTCCCGGCGCGGAGGCGCGGCTGGGGCCGTCGCAGGGCCGCGCCGCTC
A2 CCGCCHGGTCCCGTCCCGGCGCGGAGGCGCGRCTGGGGCC 3' Ava2ss
S/H CCGCCGGGTCCCGTCCCGGCGCGGAGGCGCGGCTGGGGCCGTCGCAGGGCGCGCCGCTC S/H
H/H GGCGGCCCA?GGCAGGGCC 5' HhaI/Hinf
H/T GGCGGCCCAGGGCAGGGCCGCGCCTCCGCGCCGACCCCGGCAGCGTCCCCGCGCGGCGAG H/T
A2 H CGGCCCAGGGCAGGGCCGCGCCTCCGCGCCGACCCCGGCAGCGTCCCCGCGCGGCGAG A2
1080
CCCTTCGTCCCCCCAAGGCCAGACCCCCGCCCCGGCGCCCCGCGCCCCCGCCCGCCGGCG
SmaI/Hinf 5'GCCCCCGGC D^GCCCGGCG S/H
S/H CCCTTCGTCCCCCIAAGGCCAGACCC D GGCCCG 3' Sau2/HaeII
Hinf420/Alu 3' C4444CC444CC44CC4C H/A
H/T GGGAAGCAGGGGGGTTCCGGTCTGGGGGCCGGGCCGC H G CCCGGGGCCGGGCCGGCCGC H/T
A2 GGGAAGCAGGGGGGTTCCGGTCTGG H GCCGGGCCGCGGGC 5' Ava2ss

1140

GCGGCTGTCTGTGGATCCCTTCACGGCTGCCGCCCCGGCCGGCCCCCCGGGGCCCCGGCC

S/H GCGGCTGTCTGTGGATCC 3' Sma pL/Hinf
M Bam/Hinc 5' TCCCTTCACGGCTGCCGCCCCGGCCGGCCCCCCGGGGCCCCGGCC B/H

H/A CRCCRRCRRRC3CCTAGGGAAGTGCCRRCGGCGGGGCCGGCCGGGGGG1CCC4GGGCCGG H/A
H/T CGCCGACAGACACC 5' Hinf 2/Taq

1200

CGCCGGCGGCAGCGGGCCCGGCCCCCCCCCCGGGCCGCGGCCCCGCGCCCCCCCCCC

M Sma/Ava 5' GGCCGCG4C D GCGCCCCCCCCCCC S/A
B/H CGCCGGCGGGAGCGGGCGCCGGCCCCCCCCCCGGGCCGCGGCCCCGCGCCCCCCCCCC B/H

H/A GCGGCCGCCCTCGCCCGGGCCGGGGGGGGGGD CGGCGCCGGGGCGCGGGGGGGGGG H/A
M HaeII/Sma 3' GG GGGGGGGGG H/S

1260

CCCACGACTCAGACCTCAGATCAGACGCGGCGACCCGCTCAATTTAAGCATATTACTAAG

S/A CCCACGACTCAGACCTCAGATCAGACGCGGCGACCCGCTCAATTTAAGCATATTACTAAG S/A
B/H CCCACGACTCAGACCTCAGATCAGACGCGGCGACCCGCTCAATTTAAGCATATTACTAAG B/H

H/A GGGTGCT 5' Hinf 420/Alu
H/S GGGTGCTGAGTCTGGAGTCTAGTCTGCGCCGCTGGGCGACTTAAATTCGTATAATGATTC H/S

1320

CGGAGGAAAAGAAACTAACCAGGATTCCCCCAGTAACGGCGAGTGAAGAGGGA3GAGCCC

S/A CGGAGGAAAAGAAACTAAC?AGGATTCCCCCAGTAACGGCGAGTGAAGAGGGA3GAGCCC S/A
B/H CGGAGG 3' M Bam / HincII

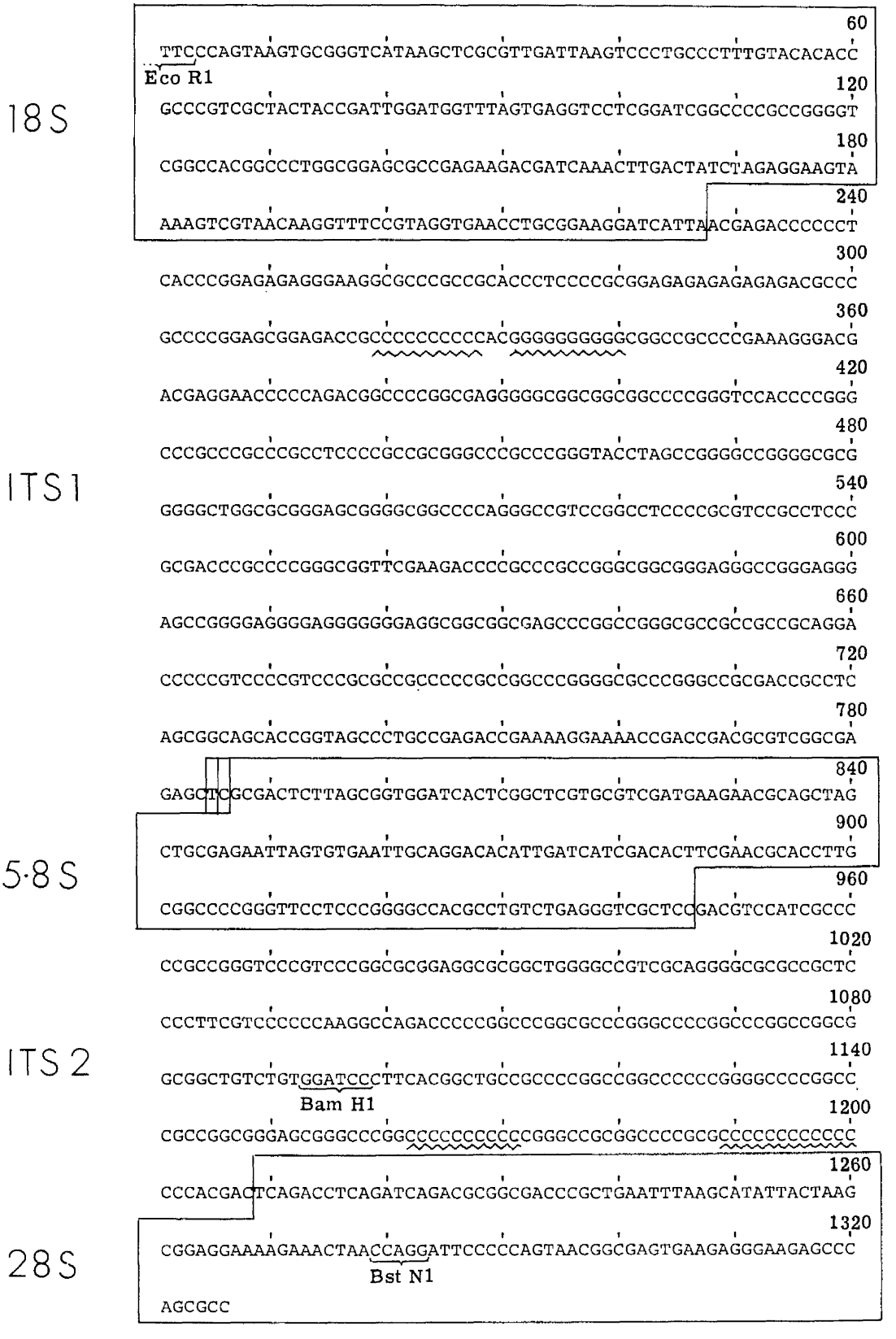
H/S GCCTCCTTTTCTTTGATTGGT?CTAAGGGGGTCATTGCCGCTCACTTCTCCCTTCTCGGG 5'
M HaeII/Sma

AGCGCC

S/A AGCGCCGAACCCCGCCCGCCCGGC 3' M Sma/Ava

Figure 4.21 Nucleotide sequence of the 18S-28S intergene region

The sequence shown is of X. laevis rDNA (from pX1r101) from the centre of the EcoRI site in the 18S gene to the first HaeII site in the 28S gene. The strand synonymous to RNA is shown (ie non-coding strand). Boxed regions denote 18S, 5.8S and 28S genes. Serrated lines indicate homopolymeric tracts of 10 or more C or G residues.


18S
ITS1
5.8S
ITS2
28S
Eco R1
Bam H1
Bst N1
60
120
180
240
300
360
420
480
540
600
660
720
780
840
900
960
1020
1080
1140
1200
1260
1320
TTCCCAGTAAGTGCGGGTCATAAGCTCGCGTTGATTAAGTCCCTGCCCTTTGTACACACC
GCCCGTCGCTACTACCGATTGGATGGTTTAGTGAGGTCCTCGGATCGGCCCCGCCGGGGT
CGGCCACGGCCCTGGCGGAGCGCCGAGAAGACGATCAAACTTGACTATCTAGAGGAAGTA
AAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTAACGAGACCCCCCT
CACCCGGAGAGAGGGAAGGCGCCCGCCGCACCCTCCCCGCGGAGAGAGAGAGACGCCC
GCCCCGGAGCGGAGACCGCCCCCCCCCCACGGGGGGGGGGCGGCCGCCCCGAAAGGGACG
ACGAGGAACCCCCAGACGGCCCCGGCGAGGGGGCGGCGGCGGCCCCGGGTCCACCCCGGG
CCCGCCCGCCCGCCTCCCCGCCGCGGGCCCGCCCGGGTACCTAGCCGGGGCCGGGGCGCG
GGGGCTGGCGCGGGAGCGGGGCGGCCCCAGGGCCGTCCGGCCTCCCCGCGTCCGCCTCCC
GCGACCCGCCCCGGGCGGTTCGAAGACCCCGCCCGCCGGGCGGCGGGAGGGCCGGGAGGG
AGCCGGGGAGGGGAGGGGGGGAGGCGGCGGCGAGCCCGGCCGGGCGCCGCCGCCGCAGGA
CCCCCGTCCCCGTCCCGCGCCGCCCCCGCCGGCCCGGGGCGCCCGGGCCGCGACCGCCTC
AGCGGCAGCACCGGTAGCCCTGCCGAGACCGAAAAGGAAAACCGACCGACGCGTCGGCGA
GAGCTCGCGACTCTTAGCGGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCAGCTAG
CTGCGAGAATTAGTGTGAATTGCAGGACACATTGATCATCGACACTTCGAACGCACCTTG
CGGCCCCGGGTTCCTCCCGGGGCCACGCCTGTCTGAGGGTCGCTCCGACGTCCATCGCCC
CCGCCGGGTCCCGTCCCGGCGCGGAGGCGCGGCTGGGGCCGTCGCAGGGGCGCGCCGCTC
CCCTTCGTCCCCCCAAGGCCAGACCCCCGGCCCGGCGCCCGGGCCCCGGCCCGGCCGGCG
GCGGCTGTCTGTGGATCCCTTCACGGCTGCCGCCCCGGCCGGCCCCCCGGGGCCCCGGCC
CGCCGGCGGGAGCGGGCCCGGCCCCCCCCCCCGGGCCGCGGCCCCGCGCCCCCCCCCCCC
CCCACGACTCAGACCTCAGATCAGACGCGGCGACCCGCTGAATTTAAGCATATTACTAAG
CGGAGGAAAAGAAACTAACCAGGATTCCCCCAGTAACGGCGAGTGAAGAGGGAAGAGCCC
AGCGCC

(1978), and was aligned with the published 5.8S RNA sequence (Khan & Maden, 1977). More recently the 5' end of 5.8S rRNA was shown to be heterogeneous, starting with pG, pCG or pUCG (Ford & Mathieson, 1978). This heterogeneity can only be accommodated if the 5.8S gene is assumed to start slightly to the left of the previously assigned position, incorporating an extra CG doublet in the 5.8S sequence. Thus the 5' end of the 5.8S gene would start at 785 (figure 4.21) with the sequence TCGCGACT..., instead of the previously assigned start of CGACT.... This and other anomalies between the DNA and RNA sequences will be discussed in 5.3. The 3' end of 5.8S rRNA matches clearly with the DNA sequence, so the 5.8S gene is located at 785-946 nucleotides from the EcoRI site (see figure 4.21).

4.4.c The 28S gene

In order to locate the 5' end of the 28S gene it was first necessary to determine the 5' oligonucleotide of 28S rRNA; this was done by B.E.H. Maden in this laboratory. It was already known that the first nucleotide is pUp (Slack & Loening, 1974; Khan & Maden, 1976). 28S rRNA (^{32}P labelled) was hybridised to pX1r11M, and the hybridised RNA was eluted and fingerprinted. Each oligonucleotide containing uridine was further hydrolysed with alkali and screened for the presence of pUp. The only nucleotide containing pUp was deduced, by digestion with pancreatic ribonuclease, to have the sequence pUCAG.

The tetranucleotide TCAG occurs three times in the DNA sequence of region M, starting at positions 1209, 1216 and 1221 in figure 4.21. To identify which, if any, of these three sites was the start of the 28S gene I hybridised ^{32}P-labelled 28S rRNA to a restriction fragment from pX1r101M. The fragment was obtained by digestion with BstNI and contained part of pBR322 plus 180 bp from the 'left hand' end of region M, extending

60-70 bp to the right of the putative start positions, as shown in figure 4.22. This fragment was chosen because it is long enough to be attached to a nitrocellulose filter yet contains a relatively short piece of the 28S gene. Any hybridised 28S rRNA should yield a simple and predictable fingerprint. After hybridisation the 28S rRNA was trimmed with T1 ribonuclease to remove 'tails' of unhybridised RNA. The hybridised RNA was then eluted and fingerprinted. The pattern of the fingerprint would differ in a predictable way according to which of the three TCAG sites was the start of the gene (see figure 4.22). pUCAG would be present in all cases. In addition AUCAG and ACCUCAG would also be present if the gene started at the first TCAG, or only AUCAG for the second. The only other products with a single uridine residue would be CUG and AAACUAACCAG which would migrate at the bottom and top of the 'one uridine' region of the fingerprint respectively. Figure 4.22 shows the fingerprint which was obtained. Characteristic products, such as AAAAG, CUG and AAACUAACCAG, could be identified by their mobilities, by comparison to other fingerprints. pUCAG could also be identified; this product runs more slowly than other 'one uridine' products because of the presence of an extra phosphate. Between CUG and AAACUAACCAG there are clearly two spots which have the mobilities expected of AUCAG and ACCUCAG. The experiment was repeated and the same result obtained. In both cases the amount of hybridised material recovered was very low, possibly due to the shortness of the region of complementarity between RNA and DNA. This made it impossible to calculate the molar yields of individual oligonucleotides with any accuracy.

In one preparation the major oligonucleotides were eluted from the fingerprint, digested with pancreatic ribonuclease and the products separated. Figure 4.22 shows the predicted products of each oligo-

Figure 4.22 5' end of the 28S gene

(a) shows detail of a sequencing gel (M Sma/Ava) extending through the three TCAG sites. (b) shows a T1 ribonuclease fingerprint of the region of 28S rRNA that hybridises to the BstNI fragment (see text). C and CG have run off the bottom of the fingerprint. Below is shown the predicted RNA sequence of the hybridised fragment if the 28S gene starts at the first TCAG site.

The major oligonucleotides were eluted and digested with pancreatic ribonuclease. In each case all the predicted products (see below) were found.

AAUUUAAG : U(2), AAU, AAG.

CAUAUUACUAAG : U(2), C, AU(2), AC, AAG.

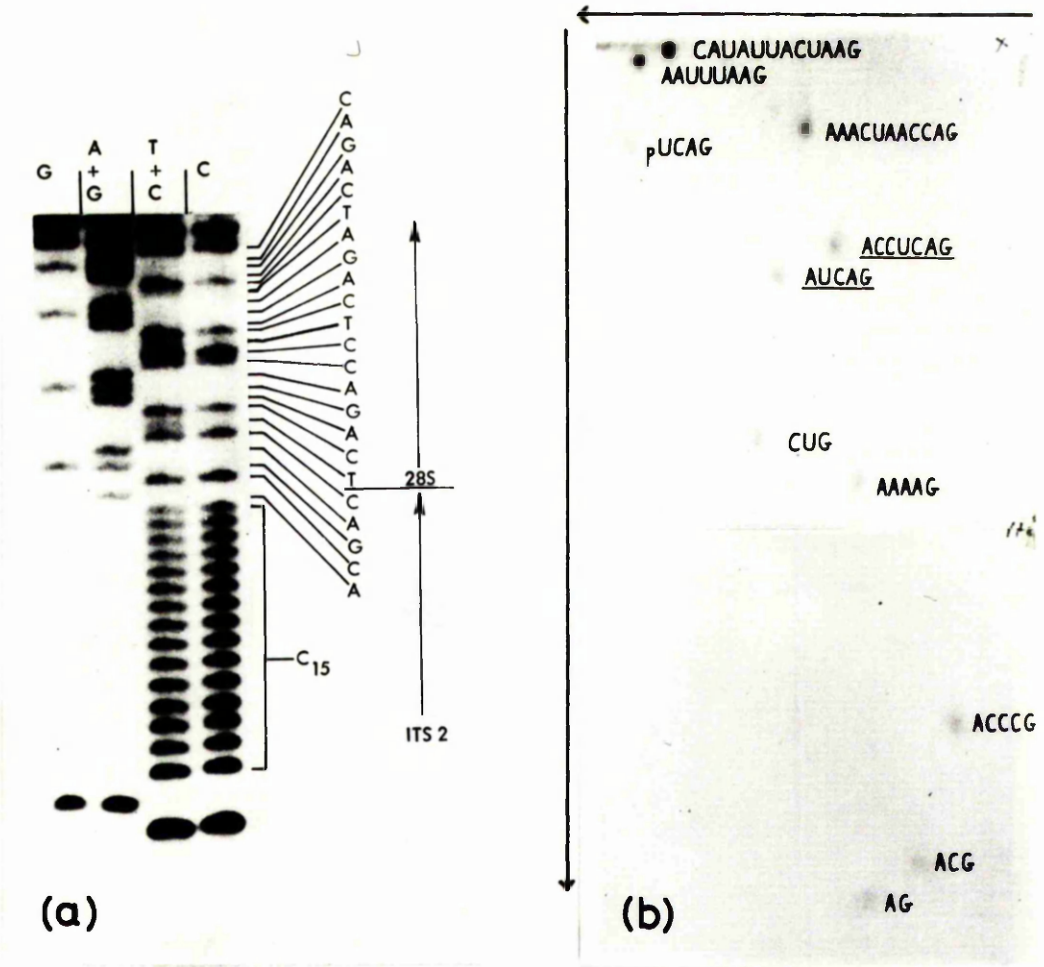
AAACUAACCAG : U, C, AG, AAC, AAAC.

ACCUCAG : U, C(2), AG, AC.

AUCAG : C, AU, AG.

CUG : U, C, G.

AAAAG : AAAAG.


G
A + G
T + C
C
C A G A C T A G A C T C C A G A C T C A G C A
28S
ITS 2
C15
(a)
CAUAUUACUAAG
AAUUUAAG
pUCAG
AAACUAACCAG
ACCUCAG
AUCAG
CUG
AAAAG
ACCCG
ACG
AG
(b)

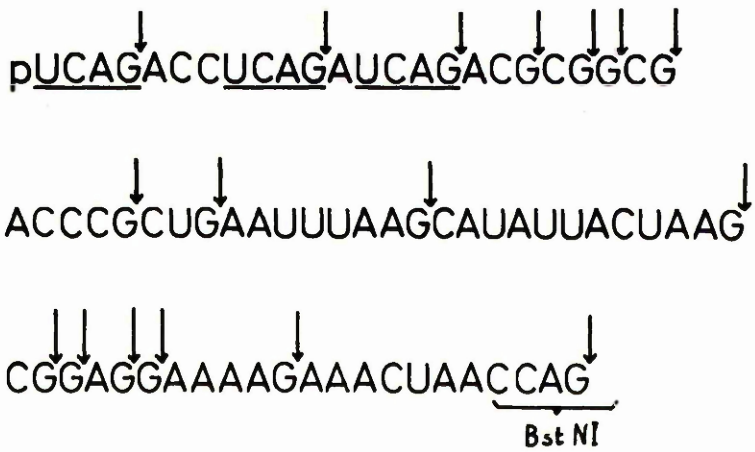

pUCAGACCUCAGAUCAGACGCGGCG
ACCCGCUGAAUUUAAGCAUAUUACUAAG
CGGAGGAAAAGAAACUAACCAG
Bst NI

Figure 4.22

nucleotide, and in all cases the observed result was consistent with these predictions. The presence of oligonucleotides AUCAG and ACCUCAG shows that the gene for 28S rRNA starts at position 1209 (from the EcoRI site), at the first of the three TCAG sites.

REGION	LENGTH	T%	A%	G%	C%
18S (3')	227	22.0	22.5	29.0	26.5
ITS1	557	3.5	12.5	41.0	43.0
5.8S	162	21.5	18.5	30.0	30.0
ITS2	262	7.0	5.0	35.0	53.0
28S (5')	118	13.5	33.0	27.0	26.5

Table 5.1 Base composition of gene and spacer regions.

Base composition of strand synonymous with RNA is shown.

CHAPTER 5. CHARACTERISTICS OF THE SEQUENCE

The complete nucleotide sequence of the 18S-28S intergene region from pX1r101 is shown in figure 4.21. This sequence has already been reported (Hall & Maden, 1980), but is discussed here in greater detail. For the purposes of this thesis I shall retain the numbering system used so far: to number the sequence from the centre of the EcoRI site in the 18S gene.

The region is divided into five components; the 3' end of the 18S gene, the first internal transcribed spacer, the 5.8S gene, the second internal transcribed spacer, and the 5' end of the 28S gene. What features of nucleotide sequence characterise these components of the ribosomal transcription unit? Table 5.1 summarises the base composition of each part of the sequence. In the following sections I shall describe features of the components, relating them to sequences from other systems where applicable. Finally I shall discuss experiments comparing pX1r101 to other cloned copies of the Xenopus transcription unit and discuss how variable the sequence might be.

5.1 3' end of the 18S gene

It has long been known that the ribosomal RNA of Xenopus and other vertebrates has a relatively high G + C content. Within the region of the 18S gene sequenced in this study the encoded RNA would be 55% G + C. Closer examination of the sequence shows that there is a short region of exceptionally high G + C content (106 - 146 in fig 4.21), but that the rest of the sequence is actually slightly A + U rich.

The sequence was examined for regions which could yield stable hairpins in the RNA molecule. This was done by running a computer program to search for perfect hairpins with specified stem and loop sizes

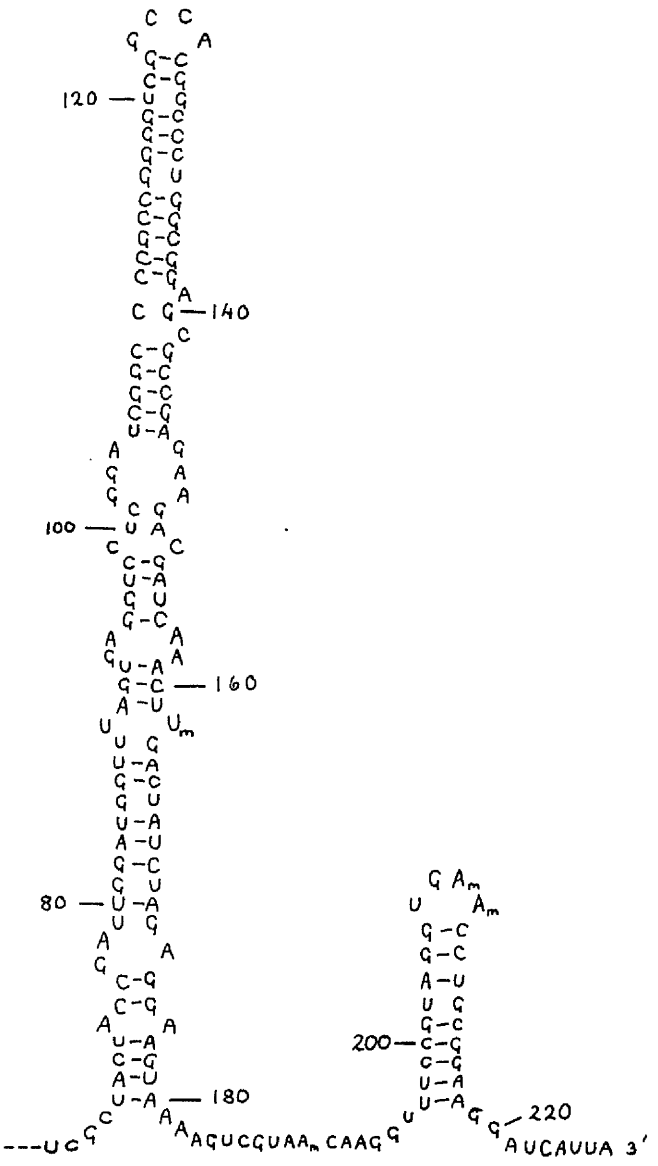


Figure 5.1 Model of possible secondary structures at the 3' end of 18S rRNA

These structures would be stable according to the rules of Tinoco et al (1973), with free energies calculated as -54.0kcal and -13.8 kcal for the large and small hairpins respectively. Sequence is numbered from the EcoRI site.

(HAIRPN; Staden, 1978). The sequence around each hairpin was then examined by eye to see if it could form a larger structure with some mismatches allowed. Two plausible structures were found. One of these occurs between 197 and 218 (figure 4.20), forming a hairpin stem of 9 base pairs (see figure 5.1). The loop enclosed by this hairpin contains two A residues which are each doubly methylated in 18S rRNA of both eukaryotes (Khan & Maden, 1976) and *E. coli* (Ehresmann *et al*, 1971). It was already known that this is an area of extremely high sequence conservation and the hairpin structure has previously been postulated (Alberty *et al*, 1978).

A larger imperfect hairpin structure can be formed in the sequence between 70 and 180, as shown in figure 5.1. This hairpin incorporates the region of high G + C content.

The sequence has now been published for the 3' region of 18S rRNA from three other eukaryotes, *S. cerevisiae* (Skryabin *et al*, 1979), *Bombyx mori* (Samols *et al*, 1979) and *D. melanogaster* (Jordan *et al*, 1980). The four sequences are compared in figure 5.2. It can be seen that there are extensive regions of close homology, including almost complete identity in the 50 nucleotides at the 3' terminus. In striking contrast the region which is 90% G + C in *Xenopus* (106-146) has a totally unrelated sequence in all four species. Furthermore this variable sequence incorporates slight length heterogeneity between the species.

As I have mentioned, the region with a high G + C content in *Xenopus* can form part of a large hairpin structure, so it was of great interest to look at the potential for base-pairing in this variable region from other species. In each case the unrelated primary sequence is able to fold into an analogous hairpin secondary structure, as shown in figure 5.3. The variable sequence forms the upper part of a larger

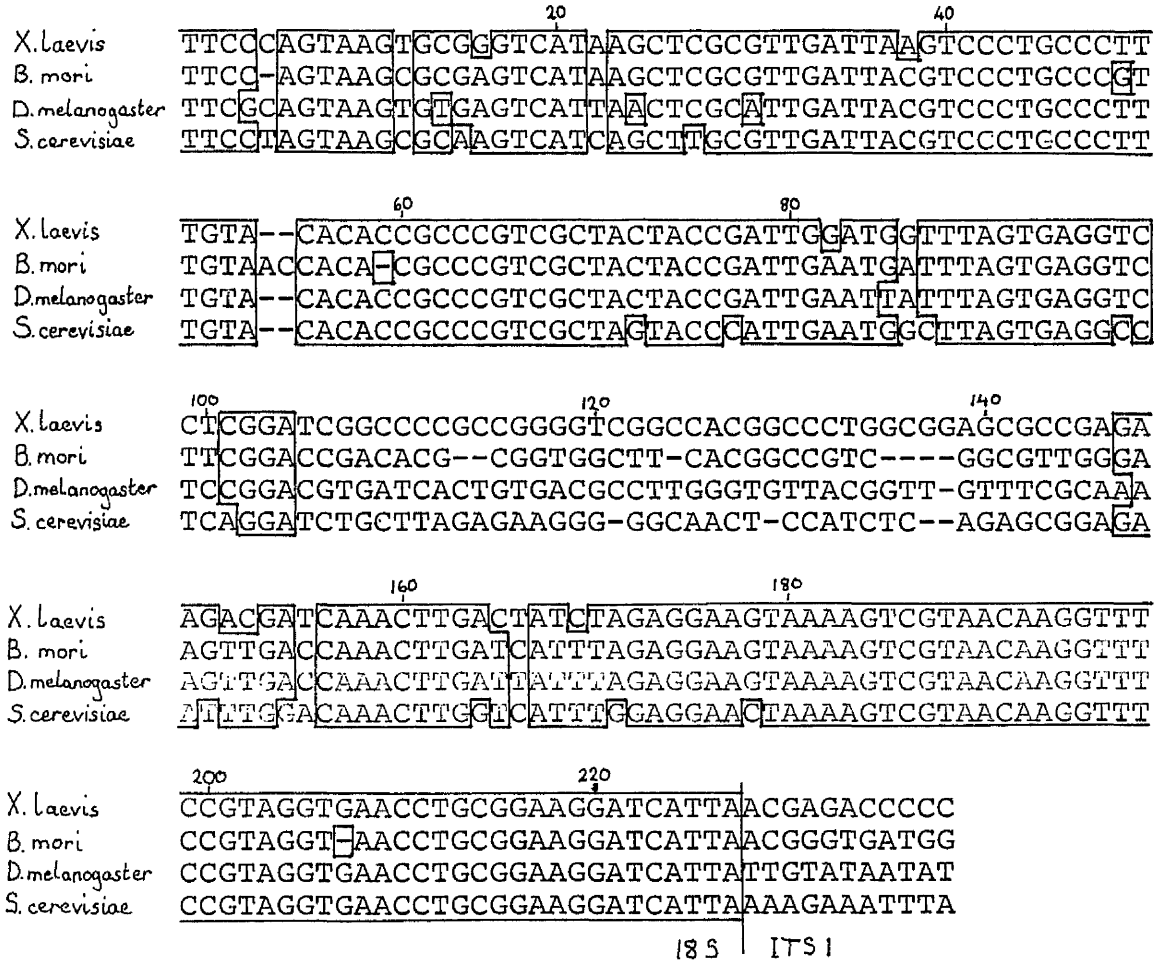


Figure 5.2 Sequence at the 3'end of the 18S gene from four eukaryotes.

Sequences are from: X.laevis, this study; Drosophila melanogaster, Jordan et al, (1980); Bombyx mori, Samols et al, (1979); Saccharomyces cerevisiae, Skryabin et al, (1979).

The X.laevis sequence is numbered from the centre of the EcoRI site.

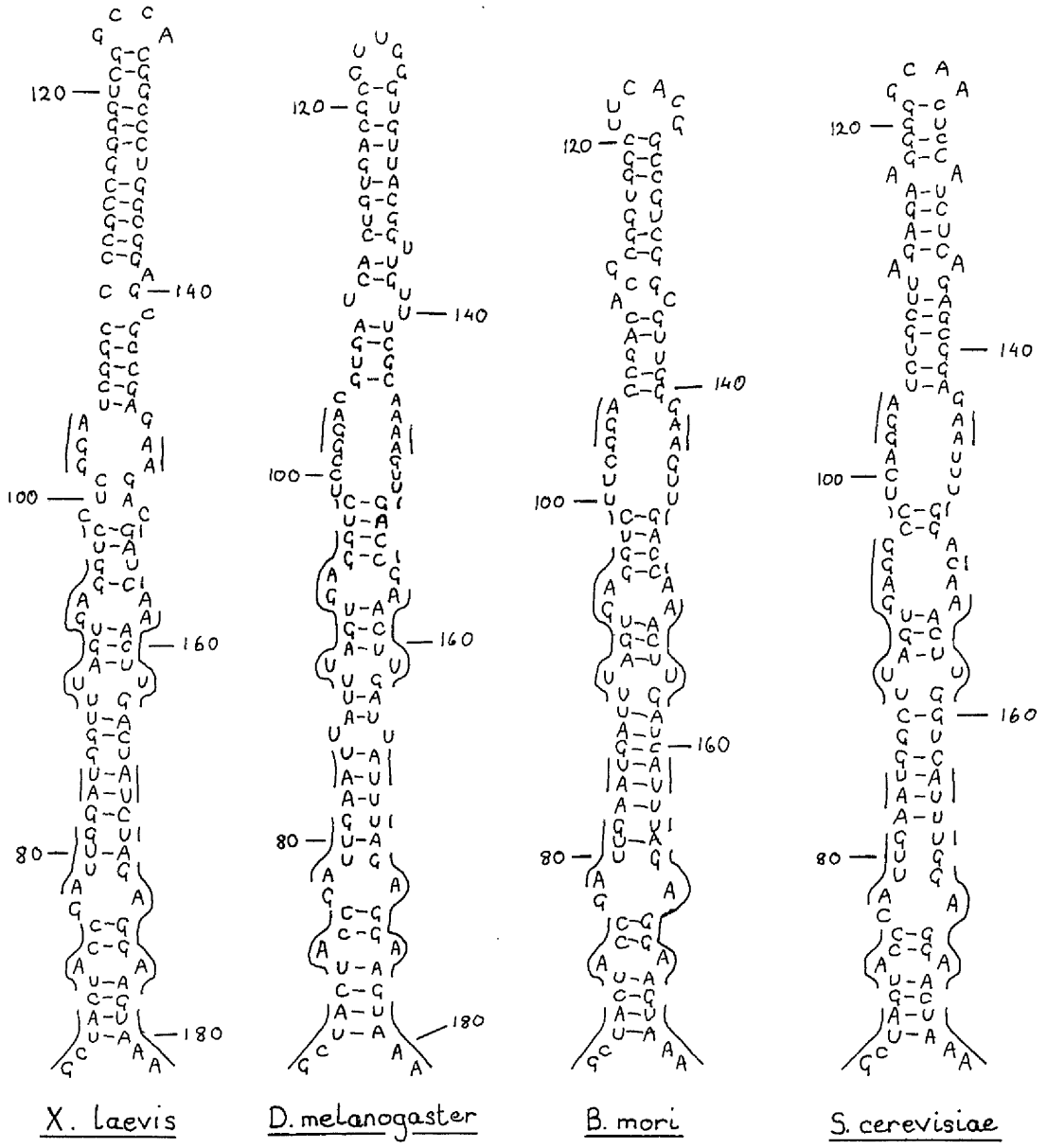
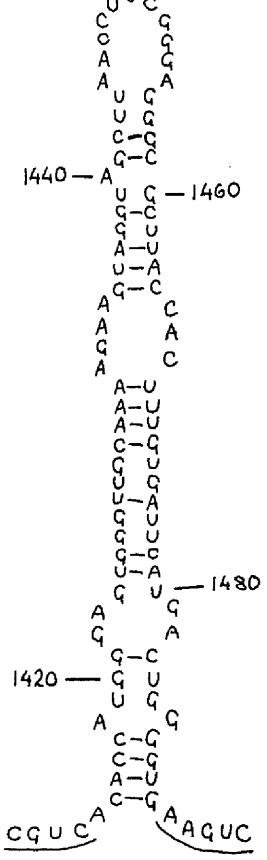


Figure 5.3 Comparison of analogous secondary structures in the 18S rRNA of four eukaryotes.

Each sequence is numbered from the centre of the EcoRI site. Nucleotides conserved in all four species are indicated by a line.



E. coli secondary structure

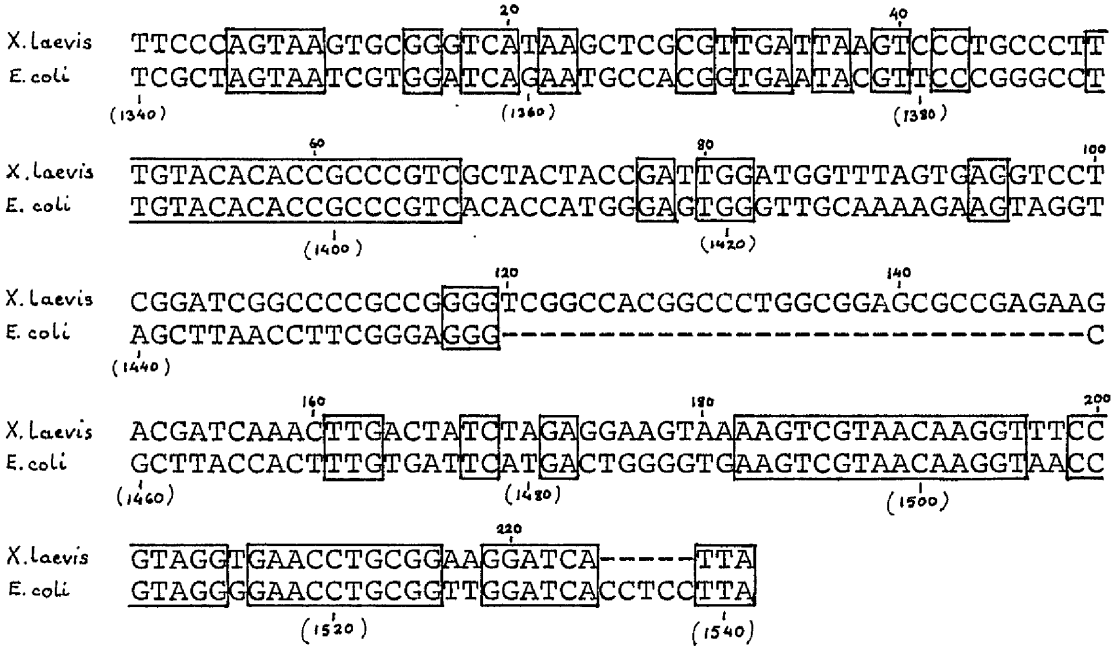


Figure 5.4 The 3'end of E.coli 16S compared with Xenopus 18S.

The sequence of the E.coli 16S gene is from Brosius et al, (1978). Regions of two or more nucleotide homology are indicated. A possible secondary structure is drawn which is comparable to the structures from four eukaryotes shown in figure 5.3.

structure, in the lower part of which both sequence and secondary structure are essentially conserved between the four species.

Comparison of the 18S sequence with E. coli 16S (Brosius et al, 1978) also shows some primary sequence conservation, particularly at the 3' end but also in a region closer to the EcoRI site. The sequences are compared in figure 5.4. Yet again an analogous hairpin structure can be formed from the most variable part of the sequence, as shown in figure 5.4. In this case there is no sequence conservation in any part of the hairpin. The structure is considerably shorter than in eukaryotes, but it occupies the same position relative to conserved sequences.

Since the publication of sequence data considerable progress has been made towards drawing up a tenable secondary structure model of 16S rRNA from E. coli (Woese et al, 1980; Noller & Woese, 1981), and small subunit rRNA from other sources (Zwieb et al, 1981). The structures given here (figures 5.3 and 5.4) agree well with these models.

5.2 First internal transcribed spacer (ITS1)

ITS1 has a remarkable base composition, as shown in table 5.1. Not only does it have a G + C content of 84%, but it also has over three times as many A residues as U residues (in the RNA). Examination of the sequence reveals that many of the A residues are clustered in tracts of purines, as shown in figure 5.5. These occur more frequently, but not exclusively, towards the 5' end of the spacer. Both the extremely high G + C content and the tracts of purines are characteristics very different from most of the 18S sequence. However the external transcribed spacer, which has also been sequenced (Boseley et al, 1979; B.E.H. Maden, unpublished work), shows a similar general pattern and the respective purine tracts are shown for comparison in figure 5.5. The ITS1 also contains a tract of 10 consecutive C residues and one of 10 G residues.

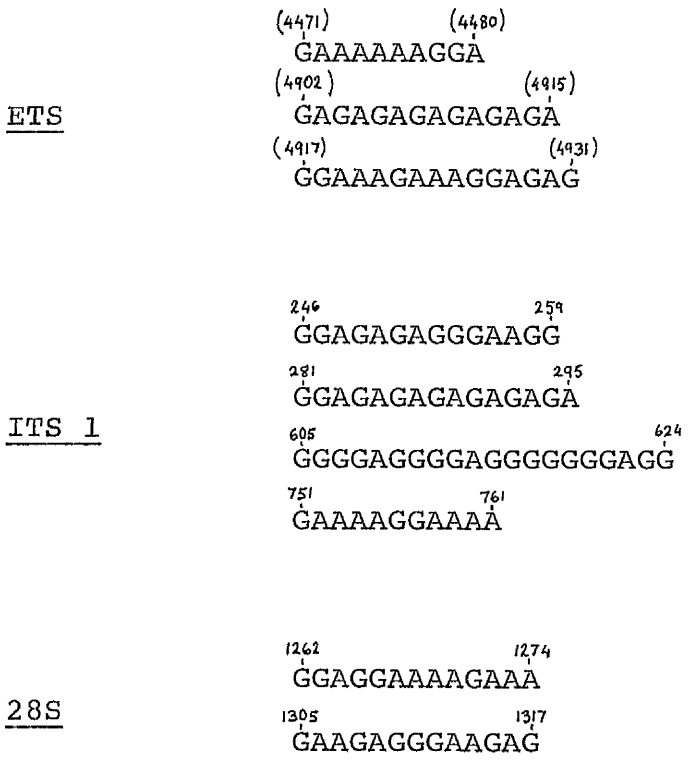


Figure 5.5 'Purine tracts' within the transcription unit.

Tracts of ten or more purines with at least three A residues in the ETS, ITS 1 and near the 5'end of the 28S gene. Tracts from the ETS are taken from the sequence of Moss et al (1980).

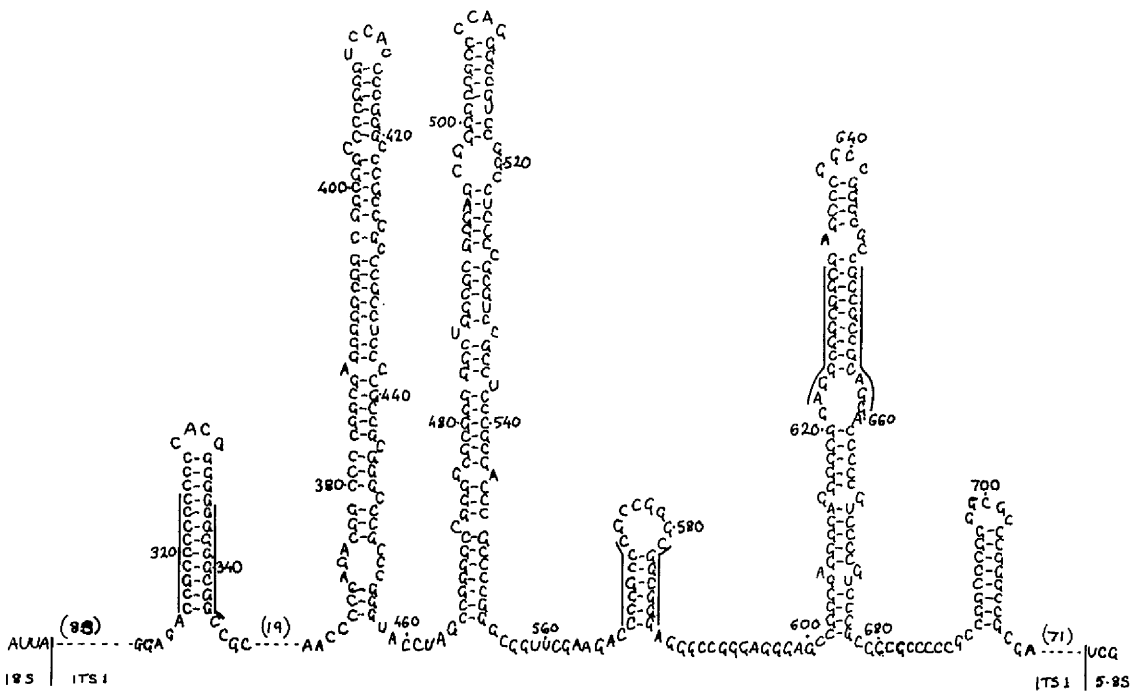


Figure 5.6 Model of possible secondary structures in ITS1.

The hairpins would be very stable because of the large number of G-C pairs. Vertical lines indicate strong secondary structure interactions observed as distortions in sequencing gels. A-rich regions at the start and end of ITS1 do not give rise to any obvious secondary structures. Numbers in brackets indicate numbers of nucleotides in single stranded regions (not shown).

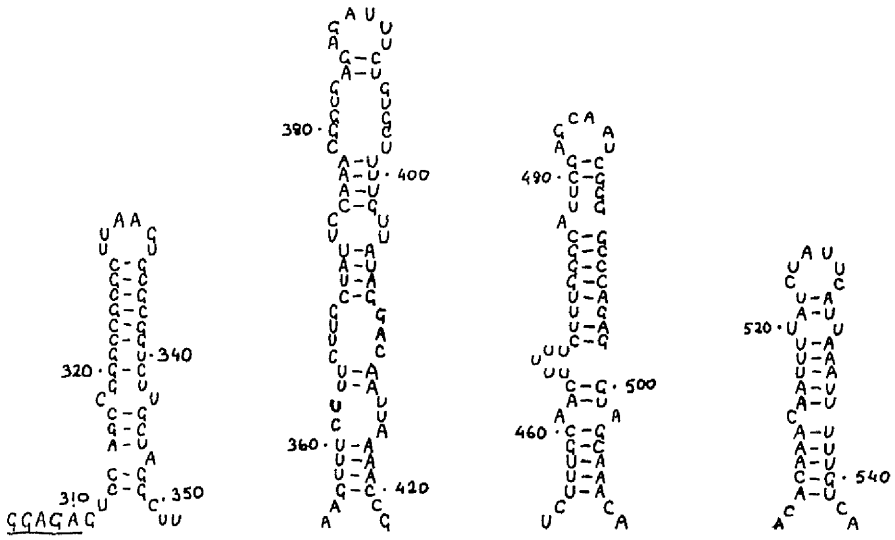


Figure 5.7 Secondary structures in yeast ITS1.

Numbering is from the centre of the EcoRI site. The 18S gene ends at 223 and the 5.8S gene starts at 584. As in the X.laevis sequence (fig 5.6) the regions around the ends of the genes cannot form stable hairpins. The first hairpin is at the same position relative to 18S as the '10C-10G' hairpin of X.laevis, and the sequence GGAGA precedes both.

There is a great potential in the sequence for forming hairpin structures, made especially stable by a large number of G-C pairs. Figure 5.6 shows a model of hairpins which could form within ITS1. It can be seen that the positions at which secondary structure bunching effects were seen on sequencing gels are consistent with the structures suggested by the nucleotide sequence. These hairpins account for a large part of ITS1, but at each end of the spacer there is a region of more than 50 nucleotides which does not appear able to form stable hairpins. This is due in part to the presence of A-rich purine tracts which are not able to participate in secondary structures because there are no U residues with which they can pair.

Again comparisons can be made with sequences available for parts of the ITS from other eukaryotes. Figure 5.2 shows the first few spacer nucleotides from a number of species and it can be seen that there is immediate and complete divergence of sequence on the spacer side of the 18S gene - ITS1 boundary. The only eukaryote other than X. laevis for which there are extensive sequence data for the internal transcribed is yeast (S. cerevisiae - Skryabin et al, 1979; S. carlsbergensis - Veldman et al, 1980). ITS1 from the two yeast species show only a few base differences from each other. Comparison of the yeast and Xenopus spacers reveals no obvious conservation of sequence. Unlike Xenopus, the yeast sequence is generally rich in A+U and has tracts of As and Us comparable to the pattern of Gs and Cs in Xenopus.

However there are two short stretches rich in G + C in yeast, and both of these can form short stable hairpins (figure 5.7). The first is particularly interesting because it could form a hairpin with a stem predominantly of G-C base pairs and in exactly the same position relative to the 18S gene as the '10C - 10G' hairpin of Xenopus ITS1 (figure 5.6).

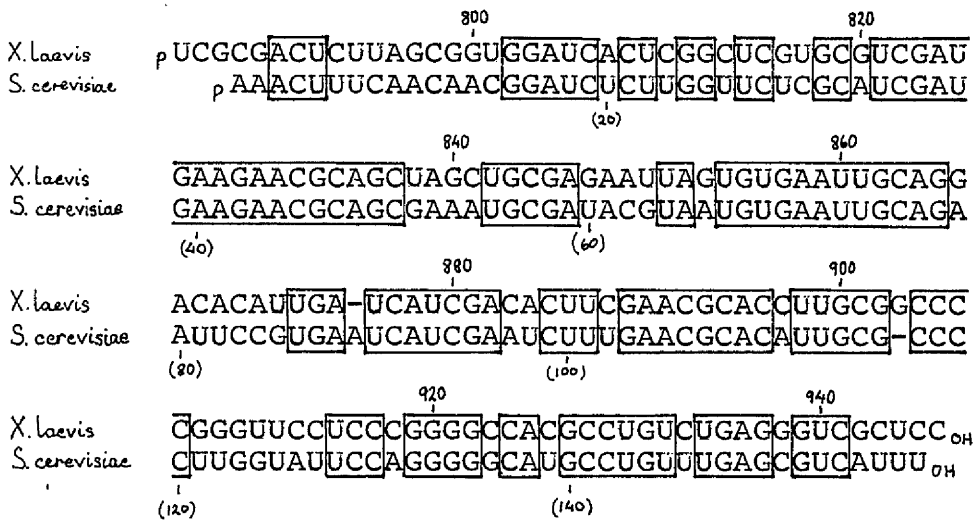


Figure 5.8 Sequence of 5.8S rRNA from Xenopus and yeast.

The yeast (S.cerevisiae) sequence is from Rubin (1973). The X.laevis sequence is deduced from the DNA sequence (this study). Homologies of two or more nucleotides are indicated.

Several other rather weak hairpins are suggested by the yeast sequence, as shown in figure 5.7. Nevertheless it is not generally possible to draw direct correlations between Xenopus and yeast structures, nor between the regions which do not contribute to hairpin formation.

5.3 5.8S gene

The sequence of the 5.8S gene published by Boseley et al (1978) is confirmed by this study, using a different cloned copy of the gene. In both cases a CG dinucleotide at positions 51-52 of the published RNA sequence (Khan & Maden, 1977; Ford & Mathieson, 1978) is absent from the DNA sequence (between 836-837). Since the RNA sequence was derived by comparison of ribonuclease digestion products with 5.8S rRNA from rat hepatoma the position within the sequence of each oligonucleotide was not shown directly. It was mentioned in Chapter 4.4 that the observation of a 5' terminal fragment with the sequence pUCG (Ford & Mathieson, 1978), taken together with the DNA sequence, leads to the conclusion that the start of the gene must be slightly to the left of the previously ascribed position. This new start will incorporate an extra CG dinucleotide in the rRNA sequence.

In summary, the revised 5.8S sequence lacks a CG dinucleotide at the former positions 51-52, but contains an extra CG dinucleotide at the 5' end of the molecule.

The sequence of 5.8S rRNA can be compared between a variety of organisms, as discussed by Pavlakis et al (1979). The sequences of yeast and Xenopus 5.8S genes are compared in figure 5.8. It can be seen that there are regions of considerable homology, interspersed with regions where the sequence has diverged. In particular the sequence at both 5' and 3' ends of the molecule are not homologous, unlike the 3'

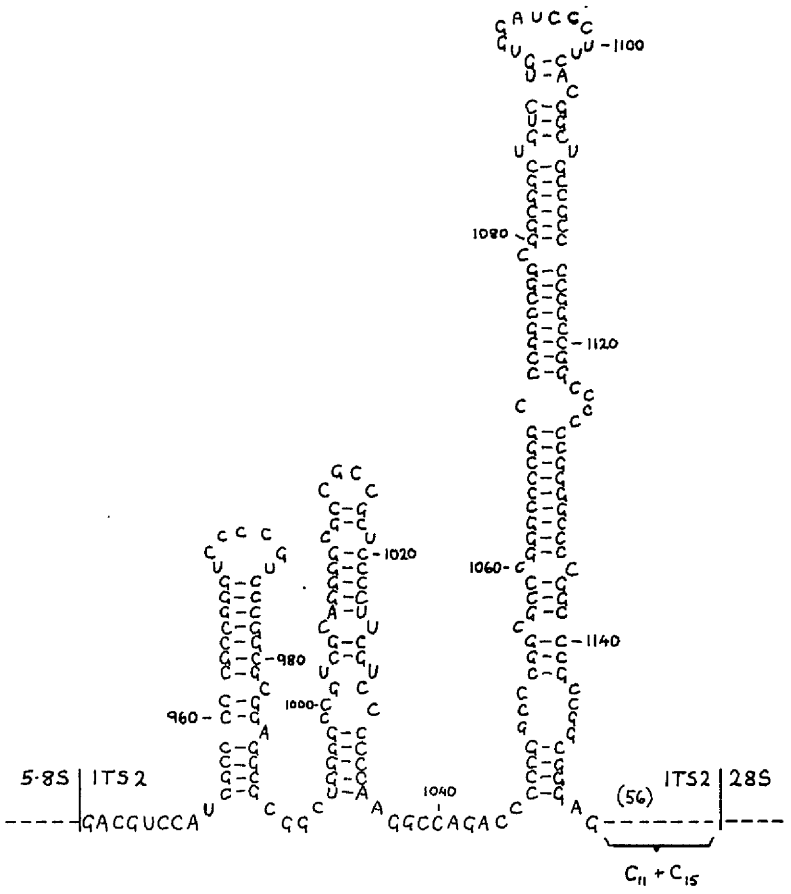


Figure 5.9 Model of possible secondary structures in ITS2.

The region close to the 28S gene cannot form extensive hairpins, largely due to two long C-tracts. Only a short single-stranded region is found next to the 5.8S gene.

end of the 18S gene. Pavlakis et al have pointed out regions which are constant over a range of eukaryotes (underlined in figure 5.8), and also note that the nucleotides between 119 and 143 in the RNA, while divergent in sequence, can always form a stable hairpin structure. This is analogous to the situation in the 3' region of the 18S gene in which certain regions show strong primary sequence conservation, while in other regions secondary structure is conserved in spite of divergence in the sequence.

5.4 Second internal transcribed spacer (ITS2)

The second ITS is again very rich in G + C. Indeed more than 50% of the nucleotides are C (in the RNA), with a tract of 11 Cs and a second of 15 Cs. Unlike ITS1 there are no long purine tracts, and Us are slightly more frequent than As. Again much of the sequence can fold into hairpin secondary structures, as shown in figure 5.9, but the two long tracts of C residues near the start of the 28S gene are likely to remain single-stranded.

Recently the sequence for yeast ITS2 became available (Veldman, G.M., Klootwijk, J. & Planta, R.J., presented at 14th FEBS, 1981). The sequence is similar in character to yeast ITS1; it again has a high A + U content and no apparent homology to X. laevis ITS2.

5.5 5' end of the 28S gene

The short region of the 28S gene which was sequenced again shows characteristics which are typical of gene rather than spacer. The sequence is not extremely rich in G + C, and there are no long tracts of a single nucleotide. There are distinctive purine tracts, as shown in figure 5.5, but unlike such tracts in ITS1 or ETS they are not surrounded by a sequence in which A residues are scarce. The occurrence of the

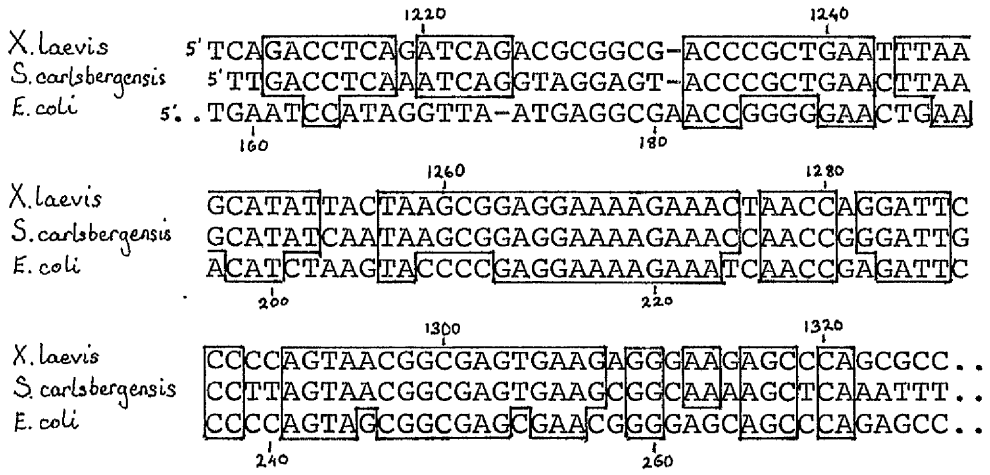


Figure 5.10 The 5'end of the 28S gene from Xenopus and yeast, and comparison to E.coli 23S.

Sequences are from:- X.laevis, this study; S.carlsbergensis, Veldman et al, presented at 14th FEBS (1981); E.coli, Brosius et al (1980).

The X.laevis sequence is numbered from the EcoRI site, so the 5'end of the 28S gene is at 1209. The E.coli sequence is numbered from the 5'end of the 23S gene. Regions with two or more nucleotides homology are indicated.

sequence 'TCAGA' three times at the start of the gene is striking, but such repetition does not occur at the 5' end of either the 5.8S gene or the 18S gene (Salim & Maden, 1980).

Figure 5.10 shows that there is again considerable homology between yeast and Xenopus sequences in the 28S gene. The 5' end of 23S rRNA from E. coli shows no homology with the eukaryotic sequences. However it has been pointed out by Walker (1981) that the sequences can be aligned if the 23S sequence is displaced by 157 nucleotides (figure 5.10). In this position the homology is extensive and includes one of the purine tracts mentioned above. The possible significance of this homology will be discussed in Chapter 7.

5.6 Comparison of different copies of the ITS region

In the preceding sections I have compared the sequence obtained for Xenopus laevis with analogous sequences from other eukaryotes, especially yeast. The conclusions essentially support hybridisation data described in 1.3 and 1.4 in showing that gene sequences are generally highly conserved while transcribed spacer sequences are highly variable between species.

However the genes for ribosomal RNA occur in multiple copies in eukaryotes and it was of considerable interest to assess the extent of variability in the transcribed spacer region within Xenopus laevis.

Having established the nucleotide sequence of the 18S - 28S intergene region from one cloned copy of the transcription unit I proceeded to compare other cloned copies by restriction mapping. pXlrl1, pXlrl02, pXlrl03, and subsequently pXL212 were chosen for comparison to pXlrl01. Region L from each plasmid was subcloned by myself, B.E.H. Maden or M.A. Stewart in this laboratory (Table 3.1).

The five subclones were compared by digestion with restriction

Figure 5.11 Comparison of clones by complete digestion with SmaI

1. Approx. 4 μg pXbr101LM digested with SmaI
2. " " pXL212L " " "
3. " " pX1r103L " " "
4. " " pX1r102L " " "
5. " " pX1r101L " " "
6. " " pX1r11L " " "
7. Approx. 10 μg pX1r11L digested with HinfI (markers).

Products were separated by electrophoresis in a 4% acrylamide gel.
It can be seen that pXL212L and pX1r103L yield identical products, and pX1r101L and pX1r102L yield identical products.

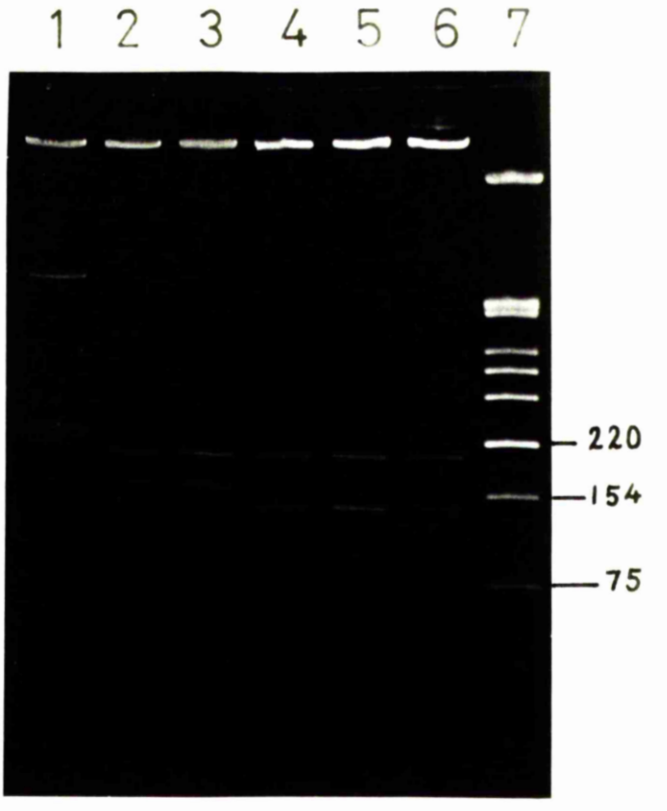


Figure 5.11

Figure 5.12 Comparison of clones by partial digestion with SmaI

Conditions to give partial digestion with SmaI were determined using unlabelled DNA. A small amount of end labelled DNA was added to unlabelled carrier, and digested under the same conditions. The products were separated on a 4% acrylamide gel and detected by autoradiography.

1. region L (labelled at BamHI site) from pX1r11L digested with SmaI
2. " " " " pX1r101 " " "
3. " " " " pX1r102 " " "
4. " " " " pX1r103 " " "
5. region L (labelled at EcoRI site) from pX1r11L digested with SmaI
6. " " " " pX1r101L " " "
7. " " " " pX1r102L " " "
8. " " " " pX1r103L " " "
9. region LM " " " pXbr101LM " " "

In 2, 3 and 4 region L was prepared from a large plasmid and is contaminated, probably with spacer DNA, giving extraneous bands. Nevertheless the arrow points to a region where bands from pX1r11L are not identical in the other plasmids. In 5-8 arrows again point to positions where differences occur between the plasmids. The upper arrow indicates the position of a length heterogeneity. The lower arrow points to an extra band (faint) from pX1r11L. Track 7 is too faint to judge the positions of bands, but other digests (track 3 and figure 5.10) indicate that pX1r102L is identical to pX1r101L.

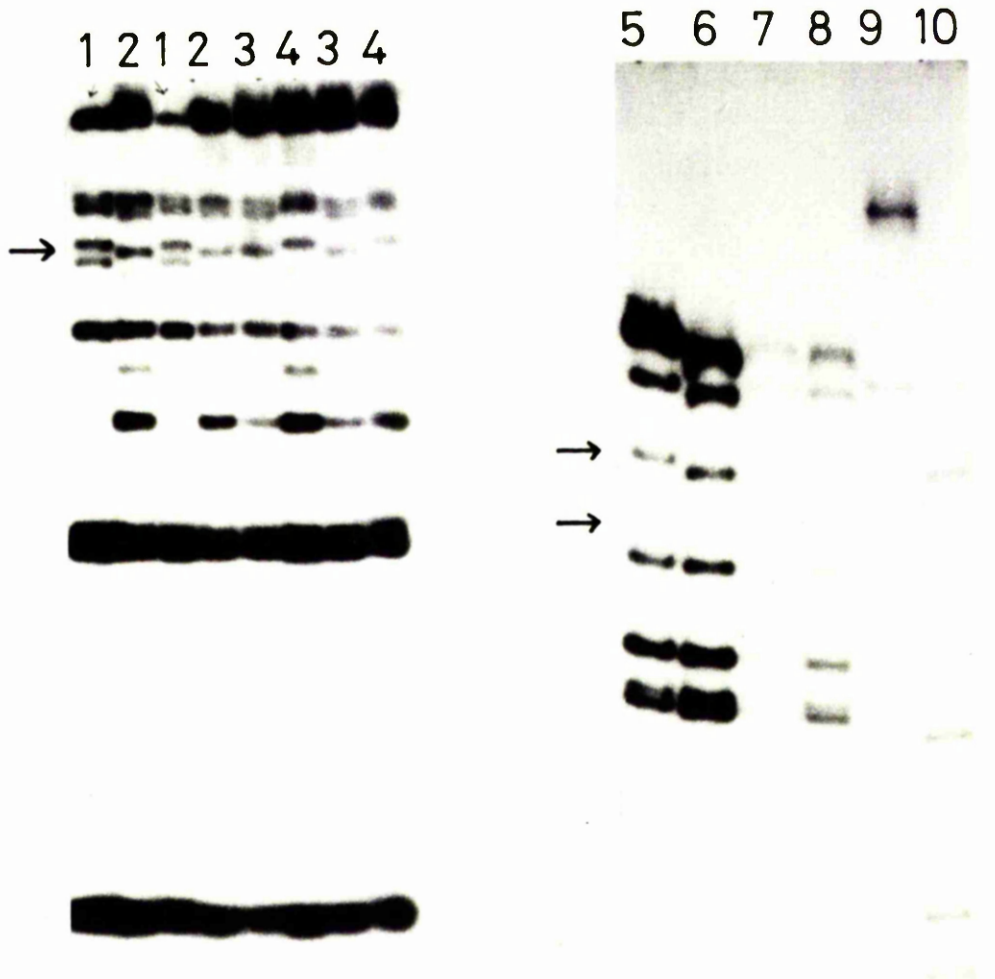


Figure 5.12

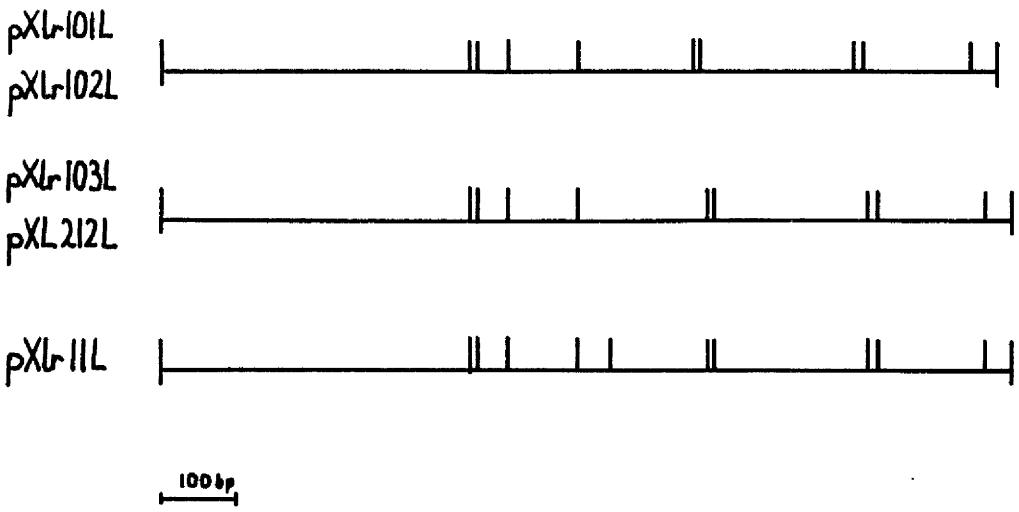


Figure 5.13 Map of SmaI sites in different copies of the transcription unit.

Exact positions of SmaI sites in pXlrl01L are derived from the sequence. Sites in other plasmids are placed by comparison, using data from both partial and complete digestions as shown in figures 5.11 and 5.12.

enzymes. It was found that digestion with SmaI (CCC↓GGG) yields a variable restriction pattern, with three different patterns being represented by the five plasmids (figure 5.11). The difference between the plasmids appears to reside in a difference in length in one major restriction fragment.

To define more closely the differences between the plasmids partial restriction digestion was used, the principle of which was described in 4.2. Again digestion with SmaI gave a variable pattern, as shown in figure 5.12. Partial digestion shows the presence of a length heterogeneity in one region of the plasmid. pXlr103L and pXlr11L both appear to be slightly longer than pXlr101L and 102L. However pXlr11L has an extra restriction site within the fragment which contains the length heterogeneity. This accords with the fragments produced by complete digestion because pXlr103L displays a fragment which is longer than the analogous fragment in pXlr101L while pXlr11L displays a shorter fragment (careful examination does reveal an extra low molecular weight band from pXlr11L, accounting for the other part of the large fragment). Figure 5.13 summarises the probable positions of SmaI sites in the plasmids, using information from both partial and complete digestion. These results show that there is indeed some heterogeneity within ITS1 in different copies of the transcription unit. However partial mapping of pXlr11L and pXlr101L with HhaI, HpaII and HaeIII did not reveal any further differences.

A second opportunity for comparing transcription units arises from the work of Boseley et al (1978) on the 5.8S gene and flanking sequences from pXL212. Firstly I have used the restriction map from pXL212 for much of the sequencing work and the only striking inconsistency which arose was with restriction sites for SmaI/AvaI. This is a

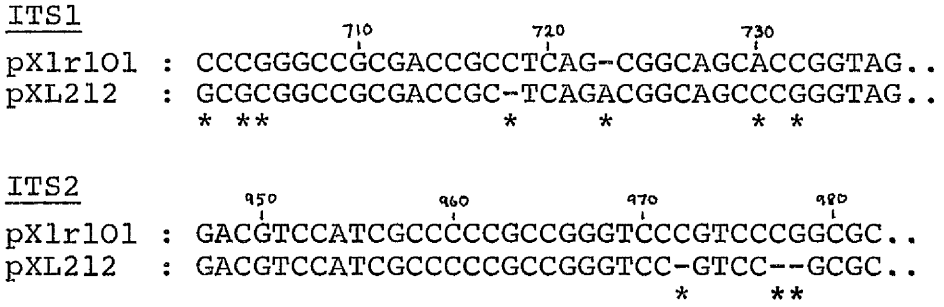


Figure 5.14 Apparent sequence differences between pX1r101 and pXL212.

The sequence of regions flanking the 5.8S gene from this study (pX1r101) is compared with the sequence of Boseley et al (1978), for pXL212.

further indication that sequence heterogeneities cannot be very extensive. Secondly I was able to compare the sequences from pXL212 and pXlr101 and found that while the gene sequences were identical there were some differences in sequences flanking the gene, as shown in figure 5.14.

Thus preliminary work indicated that there were significant but not extensive differences between different cloned copies of the internal transcribed spacer region. It was obvious that to follow up this work it would be necessary to determine the sequence from selected regions of the plasmids which had been compared by restriction mapping. This aspect of the project is currently being pursued by M.A. Stewart in this laboratory. She has not been able to confirm the sequence of Boseley et al (1978) for pXL212; the sequence flanking the 5.8S gene in pXL212 appears to be identical to that for pXlr101. Further sequence work by M.A. Stewart has confirmed the presence of a length heterogeneity and some other interesting results have been obtained.

As an additional subject for comparison region LM from X. borealis was also examined in the preceding experiments (figures 5.11 and 5.12). Restriction mapping by both complete and partial digestion indicates that restriction sites in the X. borealis ITS region are substantially different from those in X. laevis. This is being followed up by J.C. Furlong in this laboratory.

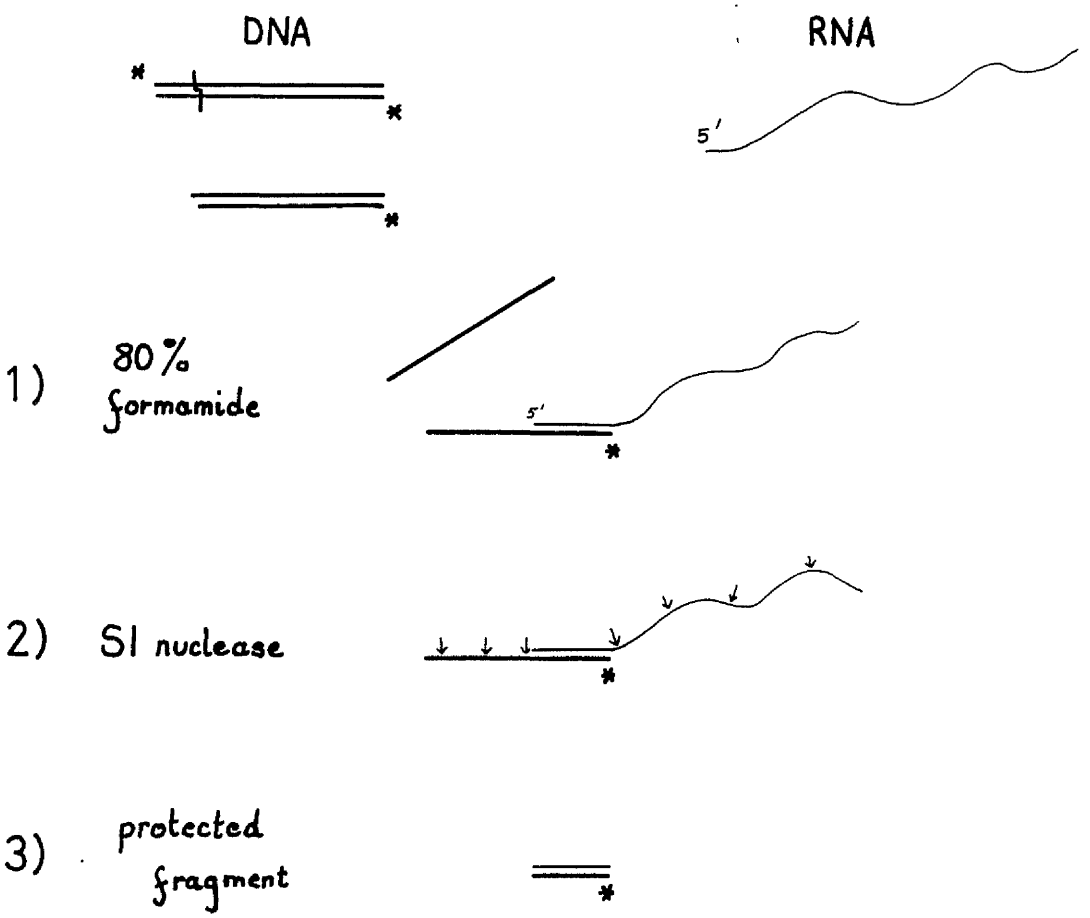


Figure 6.1 Principle of S1 nuclease protection mapping.

1) Mix end-labelled DNA with RNA in the presence of 80% formamide, conditions in which a DNA/RNA hybrid is more stable than a DNA/DNA duplex.

2) Digest with S1 nuclease, specific for single stranded nucleic acid (DNA or RNA).

3) Estimate the size of the end-labelled DNA fragment that has been protected from digestion. (A denaturing gel is used to denature the DNA/RNA hybrid.)

CHAPTER 6. CHARACTERISATION OF rRNA PRECURSORS

Electron microscopy of ribosomal precursor RNA from X. laevis has indicated that there are two classes of precursor molecule generated by cleavage within the internal transcribed spacers (Wellauer & Dawid, 1974). Both of these precursors contain the 28S rRNA sequence, identified by characteristic secondary structures after partial denaturation, together with some material from the internal transcribed spacer region. The smaller precursor, 30S RNA, is thought to be analogous to 32S RNA from mammalian cells. Maden and Robertson (1974) have shown that 32S rRNA from HeLa cells contains the characteristic T1 oligonucleotides from 28S rRNA and 5.8S rRNA, and some other 'non-conserved' products. Thus Xenopus 30S RNA is thought to contain 28S, ITS2, 5.8S, and possibly some part of ITS1. The other class of Xenopus ribosomal precursor, designated 34S, would contain these same elements plus a larger part of ITS1.

Having established the sequence of the 18S-28S intergene region I wished to identify more precisely the points in the transcript at which cleavage of RNA occurs. I employed two different approaches to attempt to solve this problem.

6.1 Mapping by protection from S1 nuclease digestion

To map the 5' end of an RNA molecule a complementary 5'-end labelled DNA fragment is used which overlaps the 5' terminus of the RNA. The region of DNA which hybridises to RNA, which must include the labelled end, is protected from digestion by S1 nuclease (single-strand specific). Thus the size of the protected fragment will demonstrate where the start of the RNA molecule lies in relation to the DNA. The principle is illustrated in figure 6.1.

Figure 6.2 S1 protection mapping of the 5'-ends of mature rRNAs

The 5' end of each mature rRNA was mapped by S1 nuclease protection. The products were separated on a sequencing gel alongside a sequence of the rDNA fragment used for hybridisation.

18S Sequence of fragment Xba/Hinf ($\sim 2 \times 10^6$ cpm/μg) from pX1r14 (see footnote 1)

1. 20 μg 18S rRNA + 70,000 cpm Xba/Hinf hybridised then digested with 1000 u/ml S1[2]
2. " " " " " " 500 u/ml S1
3. " " " " " " 100 u/ml S1
4. 0 rRNA + 70,000 cpm Xba/Hinf " " 1000 u/ml S1.

5.8S Sequence of fragment HhaI700←TaqI 820* ($\sim 3 \times 10^6$ cpm/μg) from pX1r101L

1. 5 μg 5.8S rRNA + 100,000 cpm Hha←Taq* hybridised then digested with 100 u/ml S1
2. " " " " " " 500 u/ml S1.

28S Sequence of fragment BamHI 1095←BstNI 1280* ($\sim 0.5 \times 10^6$ cpm/μg) from pX1r101M

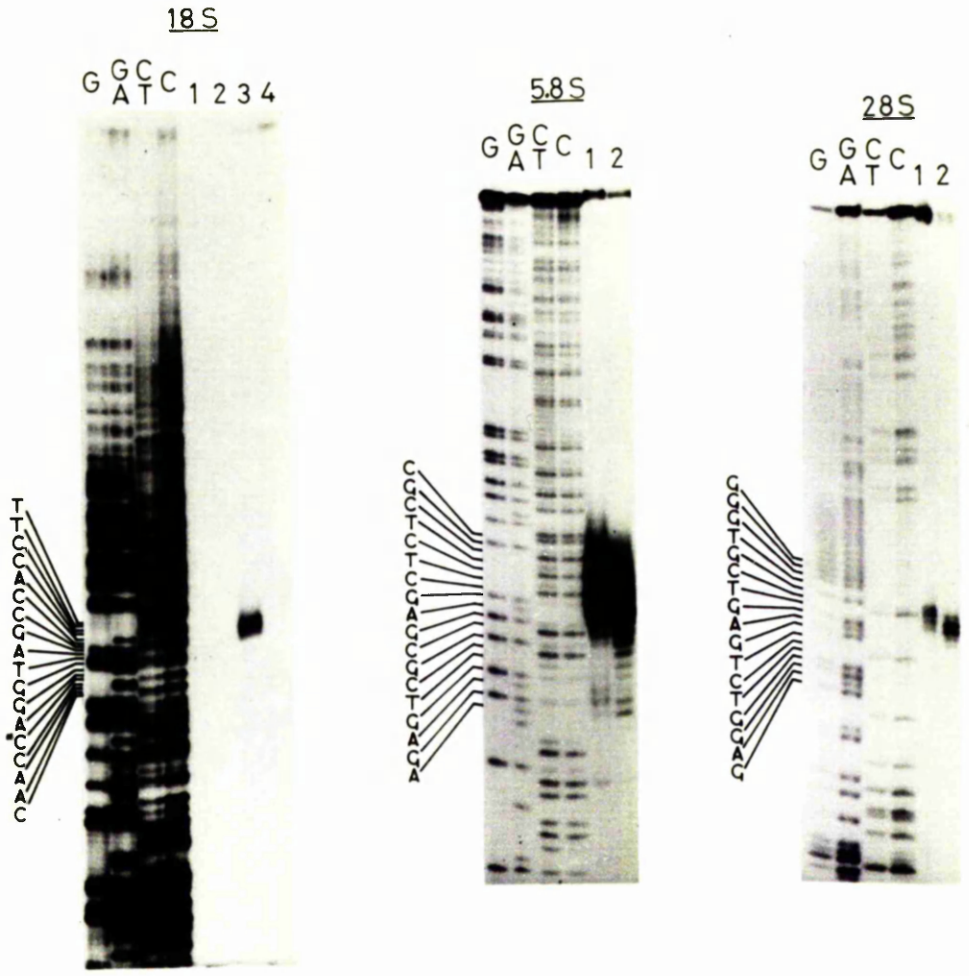
1. 20 μg 28S rRNA + 30,000 cpm Bam←Bst* hybridised then digested with 50 u/ml S1
2. " " " " " " 200 u/ml S1.

DNA fragments are described by the location of the restriction site used to prepare them (* shows site end-labelled).

The diagrams below show the positions of the major S1 bands (above RNA sequence). The known position of the end of the gene (from more direct methods, see 4.4 and footnote 1) is indicated by an arrow under the DNA sequence.

Footnote 1. The sequence around the 5' end of the 18S gene was reported by Salim and Maden (1980). The fragment used here would correspond to 'C' in figure 2 of this reference, although a different plasmid was used.

Footnote 2. S1 nuclease units given here are as defined by Sigma: one unit will cause 1.0 μg of nucleic acid to become perchloric acid per minute at pH 4.6 at 37°C. Other suppliers may give a different definition of a unit.



18S
GGCUACCUGGUUG---3′ RNA
TTCCACCGATGGACCAAC---5′ DNA

5.8S
GAGCUCGCGACUCU---3′ RNA
CGCTCTCGAGCGCTGAGA---5′ DNA
780 790

28S
ACUCAGACCUC---3′ RNA
GGGTGCTGAGTCTGGAG---5′ DNA
1210

Figure 6.2

6.1a Mapping known ends

In preliminary experiments labelled DNA fragments overlapping the 5' ends of the mature RNA species were used in order to establish the viability of the technique and suitable conditions for its use. The DNA fragment that was protected was sized against a sequence determination of the same (undigested) fragment. This allows a direct comparison to be made which will eliminate any anomalies in the mobility of a fragment caused by base composition or secondary structure effects. However it is important to note that a band on a sequencing gel corresponds to a fragment from which the specified nucleotide has been removed.

As an example figure 6.2 shows mapping of the 5' end of 28S rRNA. Location of the start of the 28S gene was described in 4.4. The sequence of DNA complementary to the proposed terminus of the RNA is 5'...GGTCTGAGTCG...3'. The major bands after hybridisation and digestion with S1 nuclease (at the higher concentration) comigrate with the bands for G and T underlined. Thus the two protected DNA fragments must end with A and G respectively, suggesting that the RNA molecule would start 5'CUCAGACC...3' or 5'UCAGACC...3'. Thus the S1 mapping technique confirms the previous conclusion that the 28S gene starts at the first of the three closely linked TCAG sites. However S1 mapping only defines the terminus of the RNA to within a few nucleotides. Similarly mapping the termini of 5.8S and 18S rRNAs yields a DNA fragment that is slightly heterogeneous in length, but the end of the RNA is correctly defined to within a few nucleotides (figure 6.2).

6.1b Mapping protected fragments in the ITS

Having established suitable conditions I proceeded to look for RNA molecules starting within the internal transcribed spacers. A crude

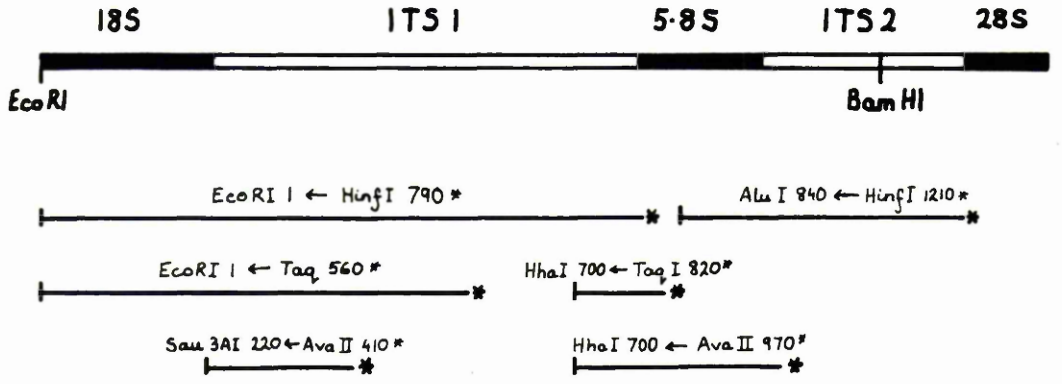


Figure 6.3 Strategy for mapping the 5'ends of precursor RNAs.

End-labelled DNA fragments which were used to search for possible precursors are shown. Fragments are named according to the restriction sites at which they were labelled and cut.

Each fragment was used to look for precursors starting within about 250 nucleotides leftwards from the labelled end. (The large HinfI fragment was also used in a preliminary search over a greater range - see text and figure 6.4)

preparation of 'precursor' RNA was made by taking molecules on the heavy side of the 28S rRNA peak from a sucrose gradient of extracted RNA. The RNA was either from the whole cell (by immediate 'hot phenol extraction') or partly enriched for nuclear RNA (by extracting with phenol at room temperature before proceeding to 'hot phenol extraction' - see 3.8). Both types of preparation will contain a large amount of 28S rRNA as well as possible precursors but by choosing DNA probes which do not contain any of the 28S gene competition between 28S and precursor RNA for hybrid formation is eliminated. The strategy used to search for precursors is outlined in figure 6.3.

In order to make a preliminary search for precursors which start in ITS1 a large HinfI fragment was used and the reaction products were sized in a 4% acrylamide (7M urea) gel. This permits products to be identified over a much wider size range than would be possible on a sequencing-type gel. The resulting gel is shown in figure 6.4, and reveals one protected fragment at approximately 125 nucleotides and a second at about 600 nucleotides.

This result was followed up using labelled DNA fragments starting at different positions within ITS1 and sizing accurately against the sequence of the same fragment. Again a protected fragment was found near the HinfI site, at about 75 nucleotides (figure 6.5). No further bands were seen which could correspond to the 125 nucleotide band on the 4% gel. Evidence from the experiments on mapping known ends suggests that 75 nucleotides will be an accurate determination of the size of the protected fragment. (It is not clear why the 4% gel yields a much larger size estimate, but it is noteworthy that in the work of Brand and Gerbi (1979) mentioned previously (4.4) the size of a DNA fragment protected by the 3' end of 18S rRNA was similarly overestimated

Figure 6.4 Preliminary S1 protection mapping in ITS1

To estimate the position and number of bands generated by S1 protection products were electrophoresed on a 4% acrylamide 7M urea gel.

1. Marker DNA fragments
2. 30 μg 'precursor' RNA + 200,000 cpm RI↤Hinf* hybridised then digested with 100 u/ml S1
3. " " " " " " 500 u/ml S1
4. 0 RNA + 200,000 cpm RI↤Hinf* " " 100 u/ml S1.

Two major protected fragments can be seen of approximately 125 and 600 bp long. These are strongest at the higher nuclease concentrations.

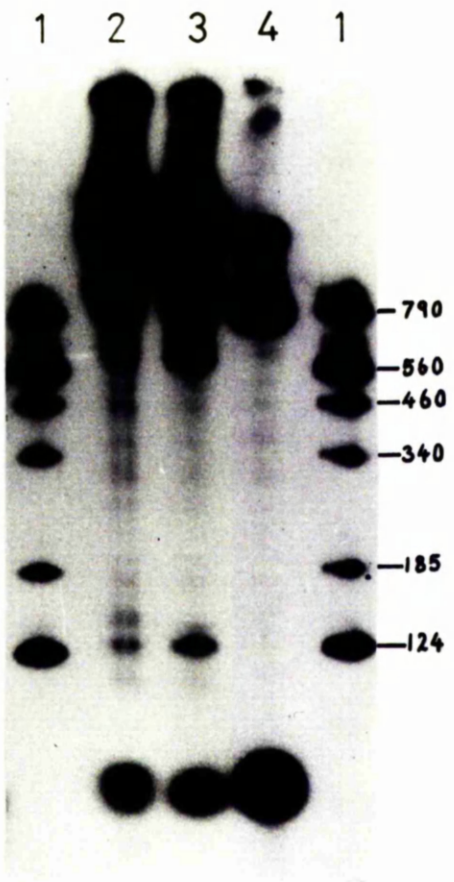


Figure 6.4

Figure 6.5 Further S1 protection mapping : 1

To define any protected fragments more precisely products were separated on a sequencing gel, as in figure 6.2.

The sequence is of the fragment EcoRI 1←HinfI 790* used for hybridisation.

1. 0 RNA + 100,000 cpm RI←Hinf* hybridised then digested with 1000 u/ml S1
2. " " " " " " " 500 u/ml S1
3. 5 μg 18S rRNA + " " " " " " 1000 u/ml S1
4. " " " " " " " 500 u/ml S1
5. 5 μg 'precursor' RNA[a] + " " " " " 1000 u/ml S1
6. " " " " " " " 500 u/ml S1
7. 0 RNA + " " " " " 500 u/ml S1
8. 30 μg 'precursor' RNA[b] +" " " " " 1000 u/ml S1
9. " " " " " " " 500 u/ml S1
10. 3 μg 'precursor' RNA[a] +" " " " " 500 u/ml S1
11. " " " " " " " 1000 u/ml S1
12. 0 RNA + " " " " " 500 u/ml S1.

The diagram below shows the major protected bands which are seen in all samples containing precursor RNA. However the same bands can be seen faintly in track 12 in which no RNA was added.

a) Preparation 'a' is a '30S' fraction of RNA isolated by hot phenol extraction on whole cells, and is thus mainly 28S rRNA.

b) Preparation 'b' was made by hot phenol extraction following cold phenol extraction of cells so should contain a higher proportion of precursor RNA (see 3.8).

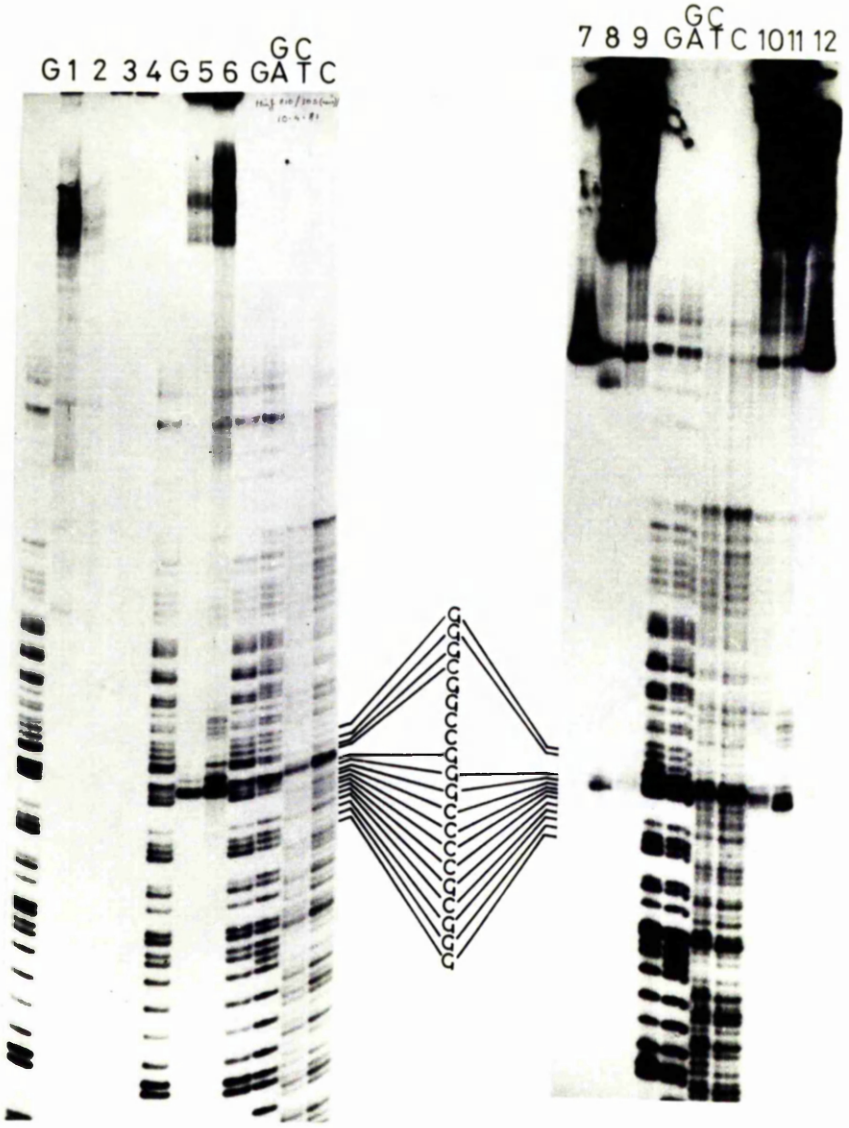

G 1 2 3 4 G 5 6 G A T C
G C
7 8 9 G A T C 10 11 12
G C
↓↓↓
CCCGGGGCGCCC---3' RNA
GGGCGGCCGGGCCCCGCGGG---5' DNA
690
700

Figure 6.5

on an alkaline agarose gel.) The presence of a band at 75 nucleotides from the HinfI site was confirmed using two different 'precursor' RNA preparations, as shown in figure 6.5. However on one gel a faint band was found at the same position in a control S1 digest of the DNA with no RNA present. Faint bands also occur in both experimental and control tracks at other positions.

No consistent protected band was seen in the central part of ITS1, using a DNA fragment labelled at a Taq site (at 560). Protected fragments were detected at the extreme 5' end of ITS1 (using Sau3AI 220 ← AvaII 410*) as seen in figure 6.6. Bands are present at two positions, with the relative intensity of the two bands altered according to which of the two RNA preparations was used. In each case a faint band is again seen in the control without RNA.

Hybridisation to a fragment containing the whole of ITS2 did not reveal any protected fragments; ITS2 is only 250 nucleotides long so it is feasible to scan the whole of this region on a sequencing-type gel.

6.1c Mapping '30S' at the 5' end of the 5.8S gene

The previous experiments had been designed to search for precursor molecules starting well inside ITS1, but a '30S' precursor might start at or close to the 5' end of the 5.8S gene. Hybridisation of 'precursor' RNA to a fragment overlapping the 5' end of 5.8S (HhaI 700 ← TaqI 820*) showed that there was indeed a protected fragment in the position of the start of the 5.8S. Although the RNA was prepared by phenol extraction at 50°C it is possible that a small amount of 5.8S rRNA may not have been released and would comigrate with 28S rRNA on a sucrose gradient. To test the possibility that the protected band was due to contaminating 5.8S rRNA a larger DNA fragment was used (HhaI 700 ← AvaII 970*). The labelled end of this fragment lies within ITS2 so

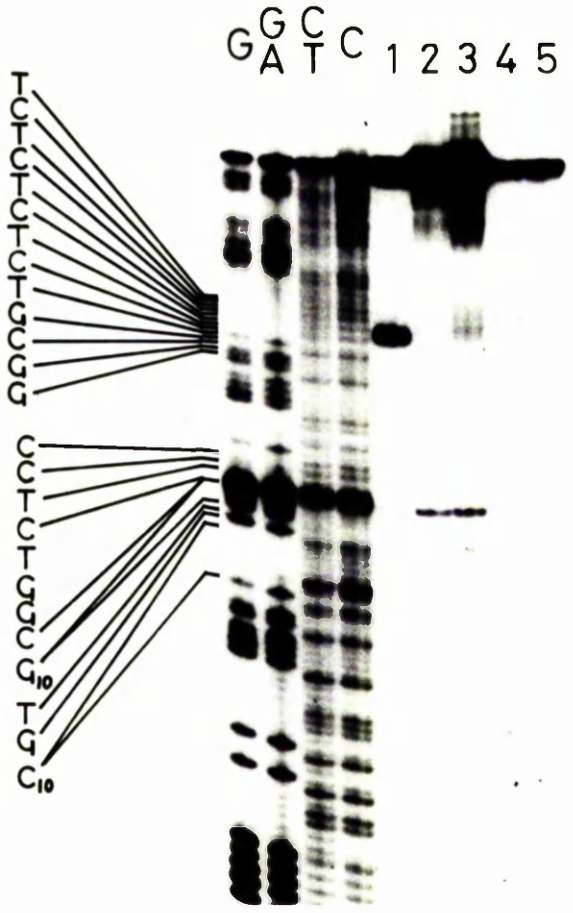
Figure 6.6 Further S1 protection mapping : 2

Protected fragments at the 'left hand' end of ITS1 were analysed using fragment Sau3AI 220←AvaII 410* (see figure 6.3).

1. 3 μg 'precursor' RNA[a] + 100,000 cpm Sau←AvaII* hybridised and digested with 1000 u/ml S1
2. 30 μg 'precursor' RNA[b] " " " " 500 u/ml S1
3. " " " " " " " 1000 u/ml S1
4. 0 RNA + " " " " 500 u/ml S1
5. " + " " " " 1000 u/ml S1.

Two sets of bands are seen, corresponding to the positions shown in the diagrams below. Bands in the same positions are seen in tracks 4 and 5 with no RNA.

a and b are as in figure 6.5.


G G C
G A T C 1 2 3 4 5
T
C T
C T C
C T C T
C T C T
C T G C G
G
C C
C T C
C T G C
C G
10
T G
C
10


AGACGCCC---3′RNA
TCTCTCTCTGCGGG---5′DNA
290
300
CCCCCCCCCCACGGGGGGGGGG---3′RNA
CCTCTGGCGGGGGGGGGGTGCCCCCCCCCC---5′DNA
320
330
340

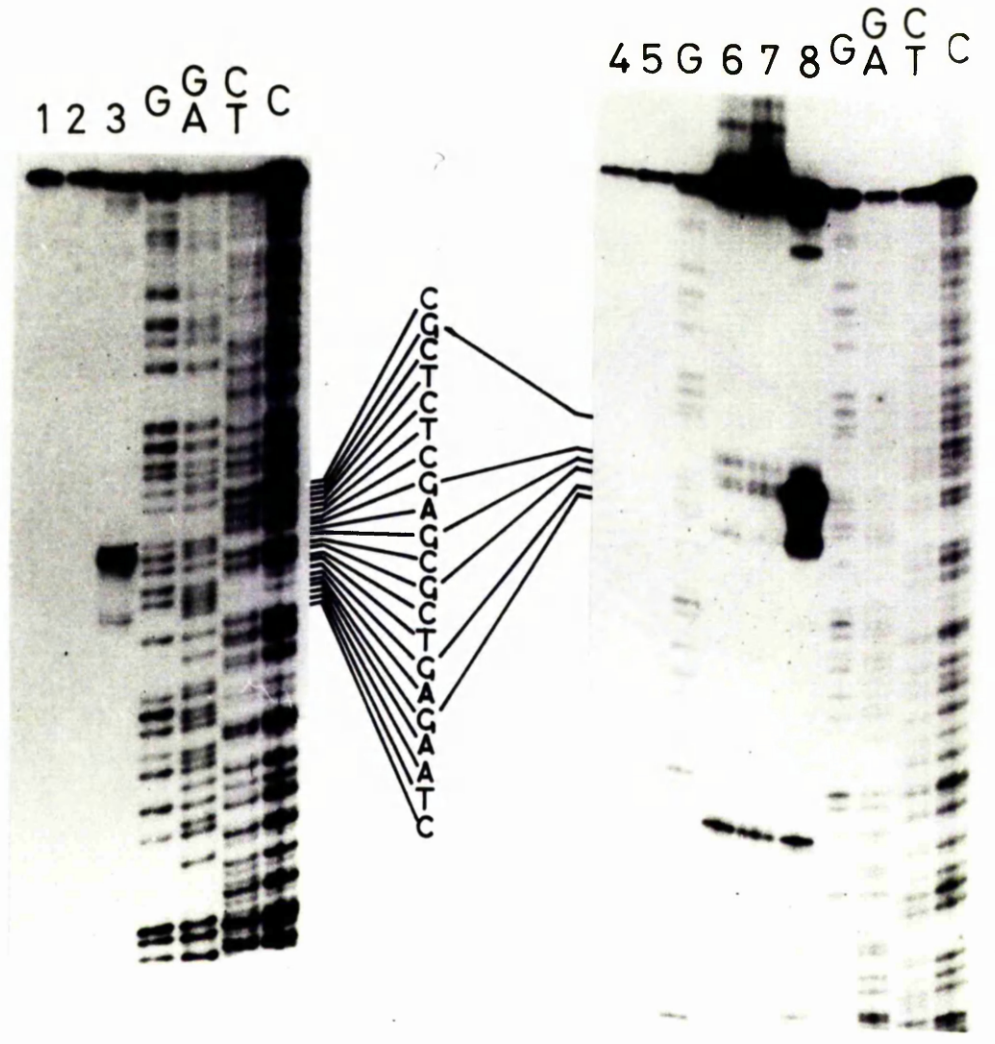
Figure 6.6

Figure 6.7 Further S1 protection mapping : 3

Protected fragments extending to the 5' end of the 5.8S gene were analysed using fragment HhaI 700←AvaII 970* which starts within ITS2.

1. 0 RNA + 100,000 cpm Hha←AvaII* hybridised then digested with 1000 u/ml S1
2. " " " " " " " 500 u/ml S1
3. 3 μg 'precursor' RNA[a] + " " " " " 1000 u/ml S1
4. As 1
5. As 2
6. 30 μg 'precursor' RNA[b] + 100,000 cpm Hha←AvaII* hybridised then digested with 1000 u/ml S1
7. " " " " " " " 500 u/ml S1
8. As 3.

The diagram shows the positions of the major bands which correspond to the start of the 5.8S gene. A further band can be seen in tracks 6, 7 and 8 at a position which would be inside the 5.8S gene.


1 2 3 G G C C
A T
45 G 6 7 8 G G C C
A T
CGCTCTCGAGCGCTGAGAATC

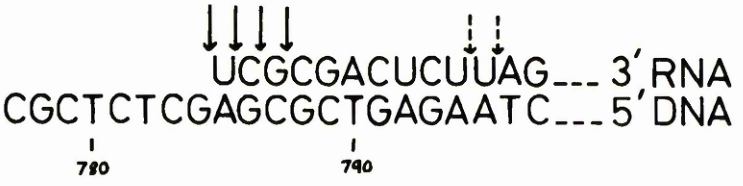

UCGCGACUCUUAG--- 3' RNA
CGCTCTCGAGCGCTGAGAATC--- 5' DNA
780
790

Figure 6.7

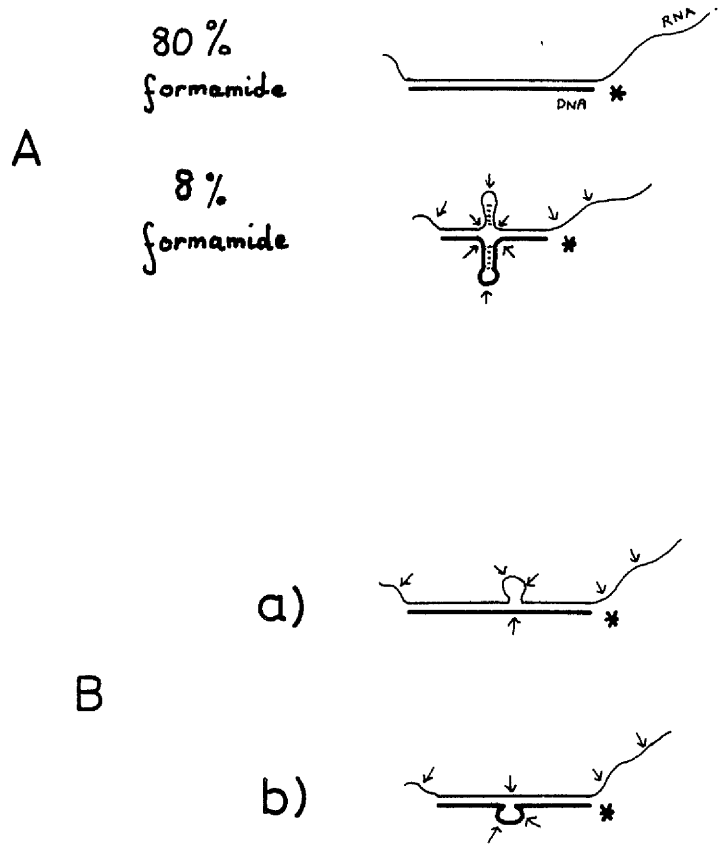


Figure 6.8 Possible causes of artifacts in S1 nuclease mapping.

A. In 80% formamide a DNA/RNA hybrid is more stable than DNA/DNA. However when formamide is diluted for S1 digestion a very stable DNA hairpin could 'loop out' from the DNA/RNA hybrid, giving rise to single stranded regions sensitive to S1 nuclease. Thus the 'protected fragment' observed will not correspond to the 5'end of the RNA.

B. Heterogeneity between DNA and RNA due to either a small deletion (a) or insertion (b) in the DNA could lead to S1 sensitive sites in both DNA and RNA. Again the 'protected fragment' does not correspond to a genuine RNA terminus.

hybridisation to 5.8S rRNA would not protect the DNA from S1 digestion. Figure 6.7 shows that a protected fragment which extends to the 5' end of 5.8S is again present (two RNA preparations were used). This confirms that there is an RNA species in the precursor fraction which has the same 5' end as 5.8S rRNA but extends beyond the 3' end of 5.8S. Since the RNA was isolated as a high molecular weight fraction after denaturation it is probable that this is a large precursor containing the RNA of 5.8S, ITS2 and 28S.

6.1d Analysis of S1 results

The S1 protection experiments suggest that one precursor to 28S has the same 5' end as 5.8S rRNA; this presumably corresponds to the 30S precursor of Wellauer and Dawid (1974).

The three other sets of bands which were seen may not correspond to the termini of genuine precursors because faint bands were also seen in the control track in which no RNA was present. Nevertheless the bands are greatly intensified in the presence of RNA. It was noticed that two of the sets of bands are at positions where the sequencing gel shows a secondary structure effect. One possibility is that a DNA/RNA hybrid is formed in 50% formamide, but that when the formamide is diluted out for S1 digestion the DNA/RNA hybrid is no longer more stable than a DNA/DNA hybrid. Thus stable DNA hairpins may loop out from the duplex forming S1 sensitive sites as shown in figure 6.8. (Such a situation could only occur if RNA molecules spanning the DNA fragment were present, such as 40S rRNA perhaps.) It is curious that at both the positions where this effect could have occurred the position of the predominant bands would apparently correspond to S1 digestion at the base of the hairpin furthest from the labelled end (ie on the 3' side).

A second cause of artifactual S1 results could be lack of homology between RNA and DNA at a particular position in the hybrid. Recent results from M.A. Stewart in this laboratory are very interesting in this respect because they show that there is a short 'insert' in some copies of the transcription unit at exactly the position of the set of 'protected' bands closest to the 5.8S gene. (This insert incidentally accounts for the length heterogeneity observed between different plasmids while restriction mapping with SmaI, see 5.6.) Thus if some or all of the RNA molecules contain the insert the DNA will probably be sensitive to S1 digestion at the position where the RNA loops away (see figure 6.8). I have now found that bands still appear at this position if a DNA fragment from pXlr11 (containing the insert) is used for the hybridisation, which would be expected if the RNA contains a mixture of transcripts with and without the insert. However, though the explanation of heterogeneity in sequence may account in part for the observation of artifactual bands, it does not explain the appearance of the same bands in the control with no RNA.

In summary it was not possible to locate unambiguously any precursors with 5' ends in ITS2 or in ITS1 upstream of the 30S precursor. It was however concluded that discrete protected fragments can be generated for reasons other than hybridisation to the 5' end of an RNA molecule. The presence of strong artifactual bands may make it difficult to recognise genuine but fainter bands caused by hybridisation to rare precursor molecules. These problems should be given due consideration in any application of this technique; in particular it is essential to use a control without RNA.

6.2 'Northern' transfers

In these experiments rRNA from a crude 'nuclear RNA' preparation

Figure 6.9 Electrophoresis of 'nuclear' RNA

'Nuclear' RNA was prepared as described in 3.8, by hot phenol extraction of the interphase layer remaining after cold phenol extraction. The RNA was separated on a sucrose gradient and a '30S' fraction, on the heavy side of the 28S peak, was taken. The 30S fraction was electrophoresed in a 1% agarose gel containing 7.5 mM methyl mercuric hydroxide. Approximately 7 μg of nucleic acid was loaded on each track. It is apparent that the '30S' nuclear fraction was by no means pure, containing a small amount of 18S rRNA (just detectable on the original gel) as well as 28S rRNA and a faint 'precursor' band. There is also some DNA present. Previous experiments had also shown that 28S and 18S rRNA are not fully separated after a single sucrose gradient.

Figure 6.9

was separated by electrophoresis and blotted onto activated DBM-paper (see 3.12). Labelled restriction fragments from particular parts of the ITS region were hybridised to the paper to probe for precursor molecules. The hope was to pick out any molecules containing RNA from the ITS and to locate approximately the 5' end by defining which DNA fragments do or do not hybridise to a particular band.

6.2a Probing for precursor bands

Figure 6.9 shows the RNA gel which was blotted onto DBM-paper. After the transfer the DBM-paper was cut in half and each half was probed with a variety of restriction fragments. The fragments used and the results obtained are shown in figure 6.10.

Unfortunately the results are difficult to interpret. A probe from the 18S gene hybridises to 18S rRNA and to a high molecular weight band presumed to be 40S, but this probe also hybridises in the region of 28S rRNA. Similarly a probe for 28S hybridises slightly to 18S rRNA, as well as to a broad '28S containing' band and to 40S rRNA.

A probe containing the whole of ITS1 (Sau3AI 220 ↔ HinfI 790) hybridises to RNA in the 28S region, with a faint indication of hybridisation to higher molecular weight material. However probes from the 5.8S gene (HinfI 790 ↔ BamHI 1095) and from the extreme 5' end of ITS1 (Sau3AI 220 ← AvaII 410* and EcoRI 1 ↔ TaqI 560) all hybridise only to the upper part of the '28S containing' band. These results suggest that the broad 28S band also contains a precursor with RNA from most of ITS1. Such a precursor would be 20% longer than 28S rRNA. Assuming that the mobility is inversely proportional to the log of molecular weight the mobility of the precursor would be 0.93x the mobility of 28S. This difference in mobility could be accommodated

Figure 6.10 Hybridisation to Northern transfers

RNA from the gel shown in figure 6.9 was transferred to DBM-paper. The paper was cut in half and the two tracks independently hybridised to various rDNA probes. Numbers 1-5 show material hybridised to the left hand track and 6-9 show the right hand track.

1. Nick translated probe : pXlr101A (containing 18S gene, see figure 3.1)
2. End-labelled probe : Sau3AI 220←AvaII 410*
3. Nick translated probe : Sau3AI 220↔HinfI 790
4. Nick translated probe : HinfI 790↔BamHI 1095 (+ DNA from pBR322)
5. Nick translated probe : pXlr101M
6. Nick translated probe : EcoRI 1 ↔ XbaI 170 (+ DNA from pBR322)
7. Nick translated probe : EcoRI 1 ↔ TaqI 560 (" ")
8. As 3.
9. As 5.

See text for interpretation of results. Numbers correspond to the position of each restriction fragment within the sequence (figure 4.21).

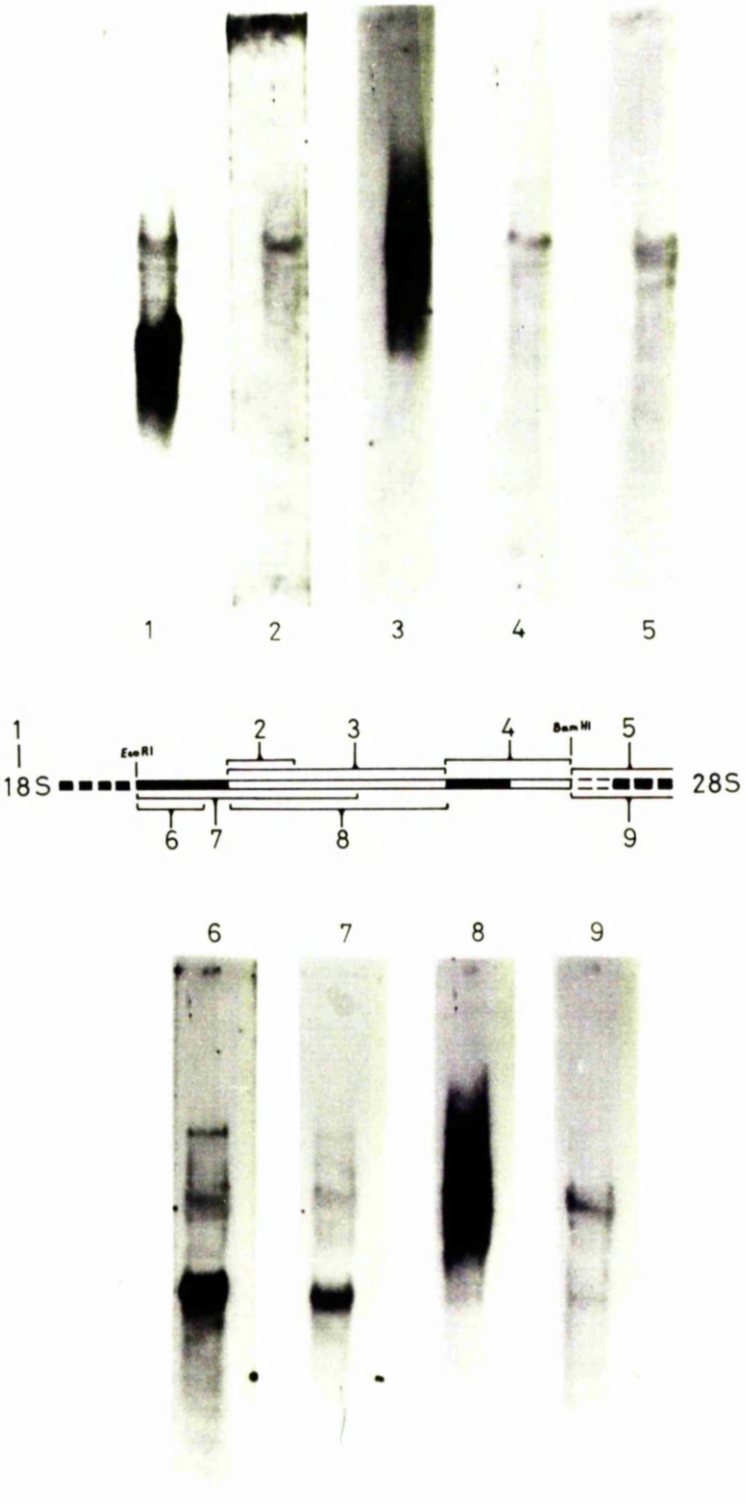


Figure 6.10

plasmid	cpm (two determinations)		average cpm
no DNA	462	406	434
pBR322	1,100	775	938
pX1r101L	162,438	183,391	172,915
pX1r101M	1,815	2,316	2,066
pX1r11R	1,715	1,624	1,670

Table 6.1 Results of cross-hybridisation experiment.

17µg pBR322, 20µg pX1r101L, 22µg pX1r101M and 26µg pX1r11R (see Table 3.1) were each digested with EcoRI and attached to nitrocellulose filters. (The quantities were calculated to be approximately equimolar.) The filters were cut exactly in half, and all filters were hybridised with the nick-translated probe Sau3AI 220 ↔ HinfI 790 from pX1r101L containing the whole of ITS1 (equivalent to 3 in figure 6.10). Filters were washed using the same conditions as for Northern transfers. Filters were dried then counted, and the results are shown above.

within the broad '28S containing' band. It must be concluded that the resolution of the gel is not really adequate to characterise precursors of 28S in this manner, especially when large amounts of mature 28S rRNA are present. The situation is further complicated by apparent cross-hybridisation between 18S and 28S, and between ITS1 and 28S, so it is possible that other bands could also result from cross-hybridisation.

6.2b Analysis of cross-hybridisation

An attempt was made to measure the extent of cross-hybridisation between different regions of the transcription unit using DNA immobilised on nitrocellulose filters. The conditions of hybridisation and washing were the same as for probing the Northern blots above. The results are shown in Table 6.1. The probe from ITS1 (Sau - Hinf) was seen to hybridise at a low level with rDNA from the 28S gene, suggesting cross-hybridisation of less than 1%. Nevertheless even at a low level such cross-hybridisation could show up on an autoradiograph, especially when the amount of 28S rRNA is very much higher than the amount of any precursor species present.

In view of the problems faced in these experiments no firm conclusions could be drawn regarding ribosomal precursor RNA species.

CHAPTER 7. FURTHER DISCUSSION

In this thesis I have reported the determination of the nucleotide sequence of part of a ribosomal transcription unit from Xenopus laevis. The sequence covers the entire 18S-28S intergene region, extending from the EcoRI site near the 3' end of 18S to the first HaeII site in the 28S gene. These results will be discussed in relation to the evolution of rDNA and to mechanisms of rRNA processing in eukaryotes and prokaryotes.

7.1 Evolution of ribosomal DNA

The DNA sequence of the 18S-28S intergene region can be divided into two categories: parts of the sequence which encode mature ribosomal RNA, and parts which encode transcribed spacer RNA. The transcribed spacers have characteristic features, such as an extremely high G + C content and long homopolymeric tracts, which distinguish them from the gene regions (see Chapters 4 and 5). The change in sequence characteristics takes place quite abruptly at the gene-spacer boundaries. Nevertheless there is a region near the 3'-end of the 18S gene which is also very rich in G + C, and other regions within the sequence of the 18S gene have similarly high G + C contents (Salim & Maden, 1981). The sequence of the 28S gene has not yet been published but it is probable that extensive G + C rich regions occur. The presence of many restriction sites for HhaI (GCGC), HpaII (CCGG) and HaeIII (GGCC) in parts of the gene suggests that this may be the case (Boseley et al, 1978). Furthermore Cox et al (1973) detected distinct regions of high G + C content in Xenopus 28S rRNA by an examination of 'melting' properties. The characteristic stable secondary structures observed in 28S rRNA by electron microscopy probably contain these same G + C rich regions (Wellauer & Dawid, 1974).

As discussed in Chapter 5 it is now possible to compare the whole of the 18-28S intergene region with the equivalent sequence from yeast. The respective 18S gene sequences can also be compared (Salim & Maden, 1981; Rubstov et al, 1980). It is noticeable that the regions which have a high G + C content in Xenopus tend to be the regions which have least homology with the yeast sequence. Thus while much of the 18S gene sequence is conserved between these two very distant species the G + C rich regions in Xenopus are very different in yeast. Again the transcribed spacers, which are G + C rich in Xenopus, show no sequence homology. There is also evidence that the G + C rich regions of the 28S gene are not conserved in evolution (Cox & Kelly, 1981, and references therein).

To summarise, there are regions within the genes which share with the transcribed spacers the property of a high G + C content. These same regions are generally not conserved between species, a feature in common with the transcribed spacers. It can be assumed that the regions of DNA which are conserved between Xenopus and yeast are critical to the function of the ribosome, and must be under strong selective pressure to retain a particular sequence. However it is interesting to speculate on the pressures which influence the evolution of the non-conserved regions of the genes and the transcribed spacers, sequences which appear to have some characteristics in common.

I have already noted that a region near the 3' end of 18S which is not conserved and in Xenopus is rich in G + C may participate in formation of a secondary structure which is itself highly conserved in eukaryotes (5.1). In the model of Zwieb et al (1981) the two other major variable regions in the 18S gene (at 660-770 and 1350-1400 in the Xenopus sequence) also form similar secondary structures in Xenopus and yeast in spite of the sequence divergence. Indeed the secondary structure

of the two rRNAs is conserved throughout. It appears that in certain regions of the gene mutations can accumulate gradually but the ability to form a particular secondary structure must be maintained; in other regions the nucleotide sequence is of more specific importance than merely in conserving RNA secondary structure.

Can this pattern be related to the evolution of transcribed spacers? In examining the potential for secondary structure formation in the ITS of Xenopus and yeast I have found one hairpin which is apparently held in common by the two species (5.2). For the rest of the ITS there is no obvious similarity between potential secondary structures. Thus the constraints which influence the evolution of even non-conserved gene regions do not apparently limit the accumulation of base changes in most of the transcribed spacer sequence.

Superimposed on selection to maintain primary or secondary structure there appears to be a general tendency in Xenopus (and other vertebrates) for rDNA to have a high G + C content. The reason for this is not clear.

The non-transcribed spacer of Xenopus is also rich in G + C, but other characteristics are distinctly different from transcribed spacer sequences (Boseley et al, 1979; Moss et al, 1980). The NTS contains regions of repetitive DNA, each having a repeating unit of 100 bp or less. Superimposed on this pattern is a larger repeating unit (of approximately 1 kb) bounded by so-called 'Bam Islands' which contain sequences homologous to the putative 40S promotor sequence (Moss & Birnstiel, 1979). In this context it is important to note that there are no repetitive sequences within the transcribed spacers.

The non-transcribed spacers are heterogeneous in length, due largely to differences in numbers of internal repeats (Botchan et al, 1977; Moss et al, 1980). However there is also a degree of sequence

variation between parts of the NTS in two different cloned rDNA units which have been examined (Moss et al, 1980; Sollner-Webb & Reeder, 1979). Similarly the work reported in this thesis shows that some sequence heterogeneity occurs in the ITS from different copies of the transcription unit. In contrast no heterogeneity has been detected in the gene regions from the same rDNA plasmids (B.E.H. Maden and M.A. Stewart, unpublished work).

Various theories have been put forward to explain the conservation and evolution of multigene families such as ribosomal DNA. Any model proposed must account for the following observations:-

i) All functional copies of the unit must evolve in parallel.

ii) rRNA gene sequences are highly conserved between species.

iii) Transcribed and non-transcribed spacers vary dramatically between species.

iv) Transcribed and non-transcribed spacers show minor sequence variation within the species.

v) Non-transcribed spacers are internally repetitive and may vary in length within a single tandem array (Wellauer et al, 1976b; Buongiorno-Nardelli et al, 1977).

Unequal crossing-over. It has been suggested that unequal crossing-over could explain the evolution of ribosomal DNA (Smith, 1976; see also Fedoroff, 1979). When sister chromatids, each containing a tandem array of repeated genes, are paired, the arrays may be misaligned. Frequent recombination between arrays would thus result in continuous duplication or deletion of individual gene units. A repeat containing a new mutation will either be lost through deletion or will be duplicated and may gradually become the most common gene type. Smith (1976) also predicts that 'non-selected' sequences like the non-transcribed spacer will become internally repetitive following misalignment with regions of

chance homology within the sequence. (Long and Dawid (1979) have reported that the NTS of Drosophila is also repetitive and of variable length.)

Gene conversion. A second mechanism which can reduce divergence between repeated genes is gene conversion (Egel, 1981). It is postulated that during recombination a hybrid duplex DNA is formed, having two strands from different original molecules. If there is any mismatch between strands, due to a mutation in one of the genes, this may be corrected by DNA repair enzymes before a subsequent round of DNA replication takes place. The result would be that one copy of one of the two genes is 'converted' to match the other gene; after subsequent replication there would be three copies of one gene type and only one of the other type.

The effectiveness of either unequal cross-over or gene conversion to maintain homology between many gene copies depends strictly on the relationship between mutation rate and the rate at which divergence is corrected. It is quite possible that different processes may be taking place simultaneously to account for the observed characteristics of ribosomal DNA.

7.2. Evolution of rRNA processing

The region which has been sequenced in this study incorporates four gene-spacer boundaries, each of which must represent a site at which the rRNA precursor is specifically cleaved. Other processing sites may also exist in the transcribed spacers but I have not been able to characterise these.

In an attempt to gain some insight into the mechanisms of rRNA processing it may be useful to consider what is known about rRNA production in E. coli.

7.2.a rRNA processing in E. coli

Processing of bacterial rRNA has recently been reviewed by Apirion and Gegenheimer (1981). The primary rRNA transcript in E. coli contains tRNAs and 5S rRNA in addition to the sequences for 16S and 23S rRNA (prokaryotes have no 5.8S rRNA). The 16S sequence resides near the 5' end of the transcript and is separated from the 23S sequence by an 'internal transcribed spacer' which includes the sequence of one or two tRNA molecules; 5S rRNA and in some cases a further tRNA molecule are encoded at the 3' end of the transcript (Ikemura & Nomura, 1977).

Processing of the rRNA precursor in E. coli appears to take place in two stages. Primary processing divides the transcript to yield the immediate precursor of each rRNA molecule while secondary processing trims these precursors to yield the mature rRNAs. Secondary processing (but not primary processing) has an obligatory requirement for a ribonucleoprotein substrate (Apirion & Gegenheimer, 1981).

The key enzyme in primary processing of rRNA is RNaseIII, which cleaves double-stranded RNA with some degree of sequence specificity. It has been found that the sequences flanking both 16S rRNA (Young & Steitz, 1978) and 23S rRNA (Bram et al, 1980) can base-pair, forming a long duplex stem with a 'loop' containing the entire 16S or 23S sequence. The duplex stems contain preferred recognition sites for RNaseIII. After RNaseIII cleavage, maturation enzymes complete the removal of precursor sequences to yield mature rRNA. Mutants have been found which lack any RNaseIII function. In these mutants ribosomal RNA processing is slower but mature rRNA is still produced, with secondary processing apparently taking place correctly but less efficiently on somewhat larger precursor molecules.

Less is known about the maturation enzymes, particularly those which are involved in production of 23S rRNA. One enzyme has been

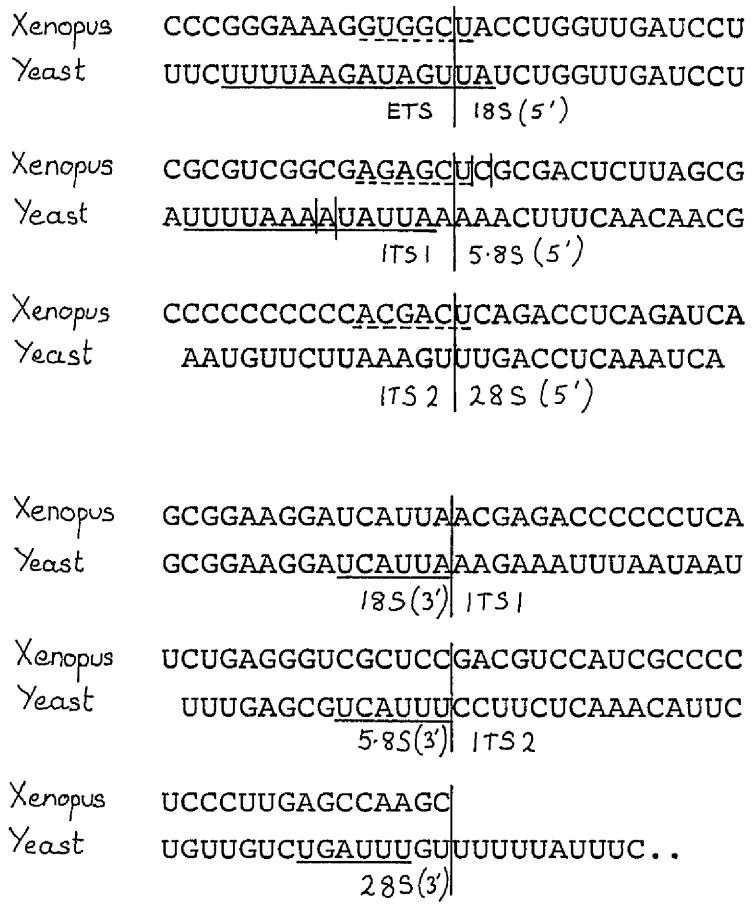


Figure 7.1 Processing sites in X.laevis and yeast.

Sequences around the gene-spacer boundaries of X.laevis (this study) and S.carlsbergensis (Veldman et al, 1980a and presented at 14th FEBS, 1981) are shown.

Sequences underlined with a solid line are partially conserved sequences in yeast noted by Veldman et al (1980a). Sequences underlined with a dashed line are partially conserved in Xenopus.

isolated which effects the final endonucleolytic cleavage at the 5' end of 16S rRNA (Dahlberg et al, 1978). A second enzyme apparently cleaves at the 3' end of 16S rRNA (Hayes & Vasseur, 1976). Both of these enzymes require a ribonucleoprotein substrate. Separate enzymes are presumably involved in processing 23S rRNA.

7.2.b rRNA processing in eukaryotes

Having considered the features of rRNA processing in prokaryotes, does this help the interpretation of information on eukaryotic processing? In both types of organism the mature rRNAs are transcribed as part of a large precursor molecule, and a series of processing events is required to produce mature rRNA.

The sequence at processing sites. In Chapter 2 I discussed a strategy of looking for critical characteristics of processing sites by comparing analogous sites from the same or different eukaryotes. Data from this study and other published work enable a comparison of nucleotide sequences at processing sites to be made. The sequences around known processing sites in X. laevis and Saccharomyces carlsbergensis are shown in figure 7.1. Sequences that have been suggested by Veldman et al (1980a) as 'conserved sites' possibly relevant to processing in yeast are underlined.

A sequence similar to the partially conserved sequence at the 3' end of yeast rRNA species is found at the 3' end of Xenopus 18S rRNA and 5.8S rRNA, but not at the 3' end of 28S rRNA. In Xenopus the 3' end of 28S rRNA is thought to be coincident with the 3' end of the primary transcript so processing does not occur at this point (Sollner-Webb & Reeder, 1979). (In yeast a few nucleotides are removed from the 3' end of the transcript, Veldman et al, 1980b.) However since the 'conserved sequence' lies within the mature rRNA the similarity between yeast and Xenopus could be due to factors other than conservation of a site for RNA cleavage.

The partially conserved sequence near the 5' end of yeast rRNAs is not seen at the 5' end of any of the X. laevis rRNAs. In its place a shorter partially conserved sequence is found, PuNPuPuC↓U (figure 7.1), although in 5.8S rRNA the molecule may also be trimmed by a further two nucleotides.

It will be of interest to see whether other eukaryotes have comparable conserved sequences at processing sites. On the other hand several distinct enzymes are almost certainly involved in processing in E. coli, so we should not necessarily expect different cleavage sites within a eukaryote to be completely analogous. More direct evidence such as characterisation of processing enzymes is essential before the significance of these partially conserved sequences can be confirmed.

Secondary structure at processing sites. Tentative secondary structure models of transcribed spacers described in Chapter 5 suggest that the regions close to gene-spacer boundaries tend to have sequences which cannot fold into hairpin structures. In E. coli the regions flanking the mature rRNAs are complementary and form a large duplex stem containing sites for primary processing by RNaseIII. I have not been able to find any regions of extensive complementarity between the sequences flanking 18S or 5.8S rRNA, and there is apparently no transcribed spacer beyond the 3' end of 28S rRNA (Sollner-Webb & Reeder, 1979). Veldman et al (1980a) also reported a lack of complementarity between flanking sequences in yeast precursor rRNA.

Cleavage by an RNaseIII-like activity as in E. coli implies simultaneous removal of flanking sequences from both sides of an rRNA molecule; in eukaryotes processing is sequential, for instance the ETS is cleaved from the 5' end of 18S well before cleavage occurs at the 3' end (see also Chapter 1.5). Thus the data suggest that eukaryotes do not have a mechanism which can be compared with primary processing in E. coli.

Genes for tRNA. In prokaryotes the 'internal transcribed spacer' contains genes for tRNA. I have made a computer search of the entire 18S-28S intergene region to see if any tRNA-like sequences are present (TRNA, Staden, 1980). No such sequences could be found.

Origin of 5.8S rRNA. One difference which does distinguish eukaryotic rRNA processing from processing in E. coli is the production of 5.8S rRNA. It has recently been suggested by Nazar (1980) that the 5' end of 23S rRNA in E. coli is homologous to 5.8S rRNA in eukaryotes. This was based on a comparison of the sequence of trout 5.8S rRNA with the 23S sequence. Jacq (1981) has noted that the 5' end of 23S rRNA is approximately 50% homologous to four different eukaryotic 5.8S rRNAs; while this is not a very high percentage it is comparable to the level of homology between E. coli 5S RNA and 5S RNA from the same four eukaryotes (X. laevis, S. cerevisiae, D. melanogaster and HeLa).

Reference has already been made to the sequence conservation which can be seen between Xenopus and E. coli rRNAs (5.1 and 5.5). This work reports the first published sequence of the 5' region of a eukaryotic 28S gene so the observed similarity with E. coli is novel and interesting. Figure 5.10 shows that the two sequences have considerable homology when the 5' end of 28S is lined up with positions 158 onwards of the 23S sequence. Since eukaryotic 5.8S rRNAs are approximately 160 nucleotides long this observed region of homology is consistent with the theory that 5.8S rRNA corresponds to the 5' region of prokaryotic 23S rRNA (Walker, 1981).

A second criterion which can be used to compare two sequences is to see whether equivalent secondary structures can be formed. Kelly and Cox (1981) have reported a region of interaction between 5.8S and the 3' end of 25S rRNA from N. crassa. A secondary structure model for the region of this interaction has been proposed, and an analogous structure

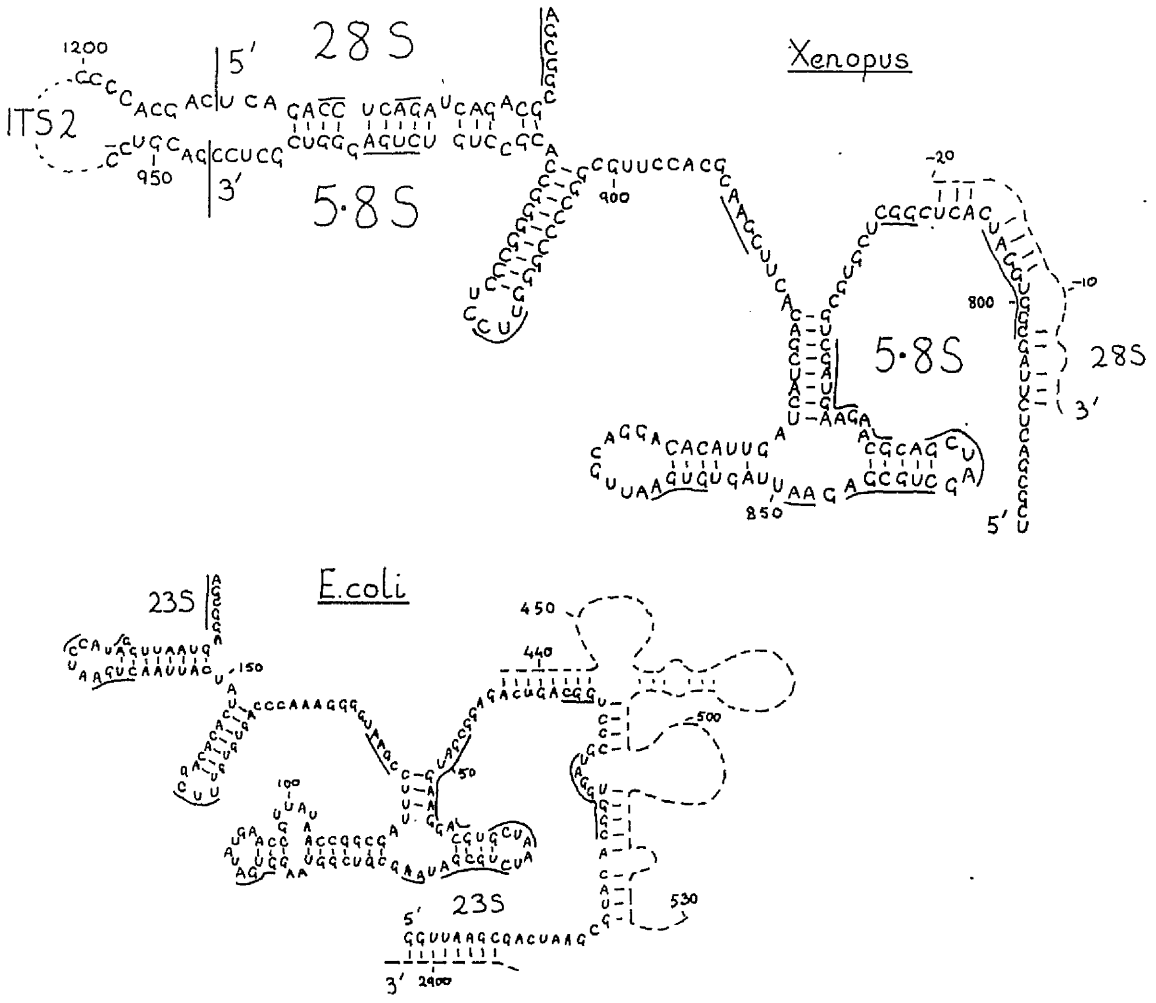


Figure 7.2 Possible interaction between 5.8S and 28S sequences.

A model of Xenopus 5.8S rRNA and its possible interaction with 28S rRNA is shown, constructed by comparison with the model of Glotz et al (1981) for the 5' end of E.coli 23S rRNA, also shown.

Solid lines indicate sequence homology between Xenopus and E.coli. Dashed lines show possible interactions with other parts of the molecule (it is not yet possible to look for an interaction with the central part of 28S in Xenopus).

For Xenopus numbering is from the EcoRI site as before, except at the 3'end of 28S where numbers are from the 3'end. For E.coli numbering is from the 5'end of 23S rRNA.

can be drawn showing interaction between the 5' and 3' ends of E. coli 23S rRNA (Cox & Kelly, 1981).

A complete secondary structure model has now been proposed for E. coli 23S rRNA by Glotz et al (1981). Again the structure of the 5' end of 23S is directly comparable to the secondary structure of 5.8S in an 'open' conformation (figure 7.2). However the interactions within 23S are somewhat different to those suggested by Cox and Kelly (1981). It is proposed that the '5.8S-like' sequence at the 5' end of 23S interacts with positions 165-175, 436-445, 515-527 and 2895-2902 of the 2904 nucleotide long 23S sequence. Interactions with the central region of the molecule are particularly well supported by experimental evidence. In this context the interaction with positions 165-175 is particularly interesting since this is the region which is correlated with the 5' end of 28S rRNA. By analogy I have examined the sequences at the 3' end of 5.8S and the 5' end of 28S in Xenopus, looking for complementarity. This reveals that extensive interaction is possible, as shown in figure 7.2. Base pairing is also possible between the 5' end of 5.8S and the 3' end of 28S (Sollner-Webb & Reeder, 1979), but it is not clear if this is comparable to the interaction in the E. coli model.

The data described above strongly suggest that 5.8S rRNA is structurally, and perhaps functionally, related to the 5' end of prokaryotic 23S rRNA. 5.8S rRNA is separated from 28S rRNA by a transcribed spacer; thus an extra stage is involved in eukaryotic rRNA processing for which there is no equivalent in E. coli.

7.2.c Further speculation on processing in eukaryotes

To summarise, there appears to be no eukaryotic mechanism directly analogous to primary processing of prokaryotic rRNA. However there is some similarity between secondary processing in E. coli and rRNA

processing in eukaryotes. For example eukaryotic processing may share with secondary processing the requirement for the rRNA to be complexed with particular proteins (Maden et al, 1969; Willems et al, 1969). It is not known what type of characteristics are recognised by the secondary processing enzymes, but Young and Steitz (1978) point out that in the proposed structure of the 16S precursor (with flanking sequences base paired) the sites of secondary processing are in single-stranded regions. This again coincides with the tentative structure of Xenopus processing sites. However eukaryotic processing also requires the cleavage of ITS2, for which there is no analogous cleavage in E. coli.

I would like to suggest that eukaryotic rRNA processing may proceed in two distinct stages, perhaps involving different mechanisms. The first stage, possibly analogous to secondary processing in E. coli, involves cleavages in ETS and ITS1 which yield mature 18S rRNA and 30-32S precursor RNA. The 30-32S molecule contains 5.8S, ITS2 and 28S rRNA, and is structurally comparable to E. coli 23S rRNA with an inserted sequence. The second stage is then to remove ITS2 (the interaction between 5.8S and 28S sequences preumably exists in the precursor molecule). If the model shown in figure 7.2 is correct the two necessary cleavage sites are presented close together in the molecular structure and could perhaps be cleaved at the same time. However an earlier cleavage also takes place, at least in yeast and Drosophila, since precursors to 5.8S are found with extra sequences at the 3' end (Veldman et al, presented at 14th FEBS, 1981; Long & Dawid, 1980b). In Drosophila the 'hidden breaks' in 5.8S and 28S rRNA mentioned in 1.1 could be produced by a mechanism comparable to the removal of ITS2 (Pavlakis et al, 1979).

There is some evidence that there is a genuine distinction between these two stages in rRNA processing. The two stages are certainly separated in time: if HeLa cells are labelled with radioactive

RNA precursors mature 18S rRNA appears in the cytoplasm in 30 minutes, while mature 28S rRNA does not appear until 70 minutes after labelling (Penman et al, 1966). During this time 32S RNA accumulates in the nucleolus at much higher levels than any other intermediate precursor species. Secondly two distinct types of precursor particle have been isolated from mammalian cells. The 80S precursor particle contains 45S RNA, many ribosomal proteins and at least seven non-ribosomal proteins; the 60S precursor particle contains 32S RNA, many ribosomal proteins, and six of the same non-ribosomal proteins together with three other non-ribosomal proteins not seen in the 80S particle (Auger-Buendia & Longuet, 1978). Thus the non-ribosomal protein complement of the smaller precursor particle is not just a subset of the proteins in the 80S particle, again indicating a distinction between the two stages of rRNA processing.

Finally Toniolo et al (1973) have isolated a line of hamster cells containing a temperature sensitive mutation involving 32S processing. At the non-permissive temperature cleavage of 32S RNA is inhibited and mature 28S rRNA is not produced. The mutation is apparently in a single gene since normal revertants can be obtained with the RNA processing activity restored. This again strongly suggests that a separate mechanism is involved in the second stage of rRNA processing.

In conclusion I hope I have shown that the determination of a nucleotide sequence can yield results which are relevant to a wide variety of topics. Thus the sequence reported here has given some insight into the evolution of different parts of the rDNA unit, into the relationship between sequence conservation and secondary structure conservation, and into some aspects of ribosomal RNA processing.

REFERENCES

Aarstad, K. & Oyen, T.B. (1975) FEBS Letters 51, 227-231

Alberty, H., Raba, M. & Gross, H.J. (1978) Nucleic Acids Research 5, 425-434

Alwine, J.C., Kemp, D.J. & Stark, G.R. (1977) Proc.Nat.Acad.Sci.USA 74, 5350-5354

Anderson, S., Gait, M.J., Mayol, L. & Young, I.G. (1980) Nucleic Acids Research 8, 1731-1743

Apirion, D. & Gegenheimer, P. (1981) FEBS Letters 125, 1-9

Arnheim, N. & Southern, E.M. (1977) Cell 11, 363-370

Auger-Buendia, M.A. & Longuet, M. (1978) Eur.J.Biochem. 85, 105-114

Bauer, W. & Vinograd, J. (1968) J.Mol.Biol. 33, 141-171

Berk, A.J. & Sharp, P.A. (1978) Proc.Nat.Acad.Sci.USA 75, 1274-1278

Bird, A.P. (1977) Cold Spring Harbor Symp. Quant. Biol. 42, 1179-1183

Birnboim, H.C. & Doly, J. (1979) Nucleic Acids Research 7, 1513-1523

Birnstiel, M., Grunstein, M., Spiers, J. & Hennig, W. (1969) Nature 223, 1265-1267

Bisbee, C.A., Baker, M.A., Hadji-Azimi, I. & Fischberg, M. (1977) Science 195, 785-787

Bolivar, F. & Backman, K. (1979) Methods in Enzymology 68, 245-267

Bolivar, F., Rodriguez, R.L., Betlach, M.C. & Boyer, H.W. (1977) Gene 2, 75-93

Boseley, P.G., Tuyns, A. & Birnstiel, M.L. (1978) Nucleic Acids Research 5, 1121-1137

Boseley, P., Moss, T., Machler, M., Portmann, R. & Birnstiel, M. (1979) Cell 17, 19-31

Botchan, P., Reeder, R.H. & Dawid, I.B. (1977) Cell 11, 599-607

Bram, R.J., Young, R.A. & Steitz, J.A. (1980) Cell 19, 393-401

Brand, R.C. & Gerbi, S.A. (1979) Nucleic Acids Research 7, 1497-1511

Brosius, J., Palmer, M., Poindexter, J.K. & Noller, H.F. (1978) Proc.Nat.Acad.Sci.USA 75, 4801-4805

Brosius, J., Dull, T.J. & Noller, H.F. (1980) Proc.Nat.Acad.Sci.USA 77, 201-204

Brown, D.D., Wensink, P.C. & Jordan, E. (1972) J.Mol.Biol. 63, 57-73

Brownlee, G.G. (1972) Determination of Sequences in RNA. North Holland/American Elsevier, Amsterdam.

Buongiorno-Nardelli, M., Amaldi, F., Beccari, E. & Junakovic, N. (1977) J.Mol.Biol. 110, 105-117

Chambon, P. (1975) Ann.Rev.Biochem. 44, 613-638

Chambon, P. (1981) Scientific American 244, 48-59

Clewell, D.B. & Helinski, D.R. (1970) Biochemistry 9, 4428-4440

Cockburn, A., Taylor, W. & Firtel, R. (1978) Chromosome 70, 19-29

Cohen, S.N., Chang, A.C.Y. & Hsu, L. (1972) Proc.Nat.Acad.Sci.USA 69, 2110-2114

Cox, R.A. (1977) Prog.Biophys.Molec.Biol. 32, 193-231

Cox, R.A. & Kelly, J.M. (1981) FEBS Letters 130, 1-6

Cox, R.A. & Thompson, R.D. (1980) Biochem. J. 187, 75-90

Cox, R.A., Huvos, P. & Godwin, E.A. (1973) Israel J. Chem. 11, 407-422

Dahlberg, A.E., Dahlberg, J.E., Lund, E., Tokimatsu, H., Rabson, A.B., Calvert, P.C., Reynolds, F. & Zahalak, M. (1978) Proc.Nat.Acad.Sci. USA 75, 3598-3602

Dawid, I.B. & Wellauer, P.K. (1976) Cell 8, 443-448

Dawid, I.B., Brown, D.D. & Reeder, R.H. (1970) J.Mol.Biol. 51, 341-360

Dudov, K.P., Dabeva, M.D., Hadjiolov, A.A. & Todorov, B.N. (1978) Biochem. J. 171, 375-383

Dugaiczyk, A., Boyer, H.W. & Goodman, H.M. (1975) J.Mol.Biol. 96, 171-184

Egel, R. (1981) Nature 290, 191-192

Ehresmann, C., Fellner, P. & Ebel, J.P. (1971) FEBS Letters 13, 325-328

Eladari, M.E. & Galibert, F. (1976) Nucleic Acids Research 3, 2749-2755

Eperon, I.C., Anderson, S. & Nierlich, D.P. (1980) Nature 286, 460-467

Erdmann, V.A. (1981) Nucleic Acids Research 9, Nucleotide Sequence Supplement r25-r42

Fedoroff, N.V. (1979) Cell 16, 697-710

Ford, P.J. & Mathieson, T. (1978) Eur.J.Biochem. 87, 199-214

Forsheit, A.B., Davidson, N. & Brown, D.D. (1974) J.Mol.Biol. 90, 301-314

Friedrich, H., Hemleben, V., Meagher, R.B. & Key, J.L. (1979) Planta 146, 467-473

Gerbi, S.A. (1976) J.Mol.Biol. 106, 791-816

Gerlach, W.L. & Bedbrook, J.R. (1979) Nucleic Acids Research 7, 1869-1885

Glotz, C., Zwieb, C., Brimacombe, R., Edwards, K. & Kössel, H. (1981) Nucleic Acids Research 9, 3287-3306

Glover, D.M. & Hogness, D.S. (1977) Cell 10, 167-176

Goldsbrough, P.B. & Cullis, C.A. (1981) Nucleic Acids Research 9, 1301-1309

Gourse, R.L. & Gerbi, S.A. (1980) J.Mol.Biol. 140, 321-339

Hadjiolov, A.A. & Nikolaev, N. (1976) Prog.Biophys.Molec.Biol. 31, 95-144

Hall, L.M.C. & Maden, B.E.H. (1980) Nucleic Acids Research 8, 5993-6005

Hayes, F. & Vasseur, M. (1976) Eur.J.Biochem. 61, 433-442

Ikemura, T. & Nomura, M. (1977) Cell 11, 779-793

Jacq, B. (1981) Nucleic Acids Research 9, 2913-2932

Jordan, B.R. (1975) J.Mol.Biol. 98, 277-280

Jordan, B.J., Jourdan, R. & Jacq, B. (1976) J.Mol.Biol. 101, 85-105

Jordan, B.R., Latil-Damotte, M. & Jourdan, R. (1980) FEBS Letters 117, 227-231

Karrer, K.M. & Gall, J.G. (1976) J.Mol.Biol. 104, 421-453

Kelly, J.M. & Cox, R.A. (1981) Nucleic Acids Research 9, 1111-1121

Khan, M.S.N. & Maden, B.E.H. (1976) J.Mol.Biol. 101, 235-254

Khan, M.S.N. & Maden, B.E.H. (1977) Nucleic Acids Research 4, 2495-2505

Khan, M.S.N., Salim, M. & Maden, B.E.H. (1978) Biochem. J. 169, 531-542

Loening, U.E., Jones, K.W. & Birnstiel, M.L. (1969) J.Mol.Biol. 45, 353-366

Long, E.O. & Dawid, I.B. (1979) Nucleic Acids Research 7, 205-215

Long, E.O. & Dawid, I.B. (1980a) Ann.Rev.Biochem. 49, 727-764

Long, E.O. & Dawid, I.B. (1980b) J.Mol.Biol. 138, 873-878

Maden, B.E.H. (1980) Nature 288, 293-296

Maden, B.E.H. & Robertson, J.S. (1974) J.Mol.Biol. 87, 227-235

Maden, B.E.H. & Salim, M. (1974) J.Mol.Biol. 88, 133-164

Maden, B.E.H., Vaughan, M.H., Warner, J.R. & Darnell, J.E. (1969) J.Mol.Biol. 45, 265-275

Maden, B.E.H., Khan, M.S.N., Hughes, D.G. & Goddard, J.P. (1977) Biochem.Soc.Symp. 42, 165-179

Markov, G.G. & Arion, V.J. (1973) Eur.J.Biochem. 35, 186-200

Maxam, A.M. & Gilbert, W. (1977) Proc.Nat.Acad.Sci.USA 74, 560-564

Maxam, A.M. & Gilbert, W. (1980) Methods in Enzymology 65, 499-560

Meunier-Rotival, M., Cortadas, J., Macaya, G. & Bernardi, G. (1979) Nucleic Acids Research 6, 2109-2123

Morrow, J.F., Cohen, S.N., Chang, A.C.Y., Boyer, H.W., Goodman, H.M. & Helling, R.B. (1974) Proc.Nat.Acad.Sci. USA 71, 1743-1747

Moss, T. & Birnstiel, M.L. (1979) Nucleic Acids Research 6, 3733-3743

Moss, T., Boseley, P.G. & Birnstiel, M.L. (1980) Nucleic Acids Research 8, 467-485

Nazar, R.N. (1980) FEBS Letters 119, 212-214

Noller, H.F. & Woese, C.R. (1981) Science 212, 403-411

Pardue, M.L. (1973) Cold Spring Harbor Symp. Quant. Biol. 38, 475-482

Pardue, M.L., Brown, D.D. & Birnstiel, M.L. (1973) Chromosoma 42, 191-203

Pavlakis, G.N., Jordan, B.J., Wurst, R.M. & Vournakis, J.N. (1979) Nucleic Acids Research 7, 2213-2238

Penman, S. (1966) J.Mol.Biol. 17, 117-130

Penman, S., Smith, I. & Holtzman, E. (1966) Science 154, 786-789

Perry, R.P. (1967) Progress in Nucleic Acid Research and Molecular Biology 6, 219-257

Perry, R.P. (1976) Ann.Rev.Biochem. 45, 605-629

Rigby, P.W.J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) J.Mol.Biol. 113, 237-251

Rubin, G.M. (1973) J.Biol.Chem. 248, 3860-3875

Rubin, G.M. & Sulston, J.E. (1973) J.Mol.Biol. 79, 521-530

Rubstov, P.M., Musakhanov, M.M., Zakharyov, V.M., Krayev, A.S., Skryabin, K.G. & Bayev, A.A. (1980) Nucleic Acids Research 8, 5779-5794

Salim, M. (1972) Ph.D. Thesis, University of Glasgow

Salim, M. & Maden, B.E.H. (1980) Nucleic Acids Research 8, 2871-2884

Salim, M. & Maden, B.E.H. (1981) Nature 291, 205-208

Samols, D.R., Hagenbüchle, O. & Gage, L.P. (1979) Nucleic Acids Research 7, 1109-1119

Sanger, F. & Coulson, A.R. (1975) J.Mol.Biol. 94, 441-448

Sanger, F. & Coulson, A.R. (1978) FEBS Letters 87, 107-110

Sanger, F., Nicklen, S. & Coulson, A.R. (1977) Proc.Nat.Acad.Sci.USA 74, 5463-5467

Schibler, U., Wyler, T. & Hagenbüchle, O. (1975) J.Mol.Biol. 94, 503-517

Shine, J. & Dalgarno, L. (1974) Biochem.J. 141, 609-615

Skryabin, K.G., Krayev, A.S., Rubstov, P.M. & Bayev, A.A. (1979) Dokl.Akad.Nauk.USSR 247, 761-765

Slack, J.M.W. & Loening, U.E. (1974) Eur.J.Biochem. 43, 59-67

Smith, A.J.H. (1979) Nucleic Acids Research 6, 831-848

Smith, G.P. (1976) Science 191, 528-535

Smith, H.O. & Birnstiel, M.L. (1976) Nucleic Acids Research 3, 2387-2398

Sollner-Webb, B. & Reeder, R.H. (1979) Cell 18, 485-499

Staden, R. (1978) Nucleic Acids Research 5, 1013-1015

Staden, R. (1979) Nucleic Acids Research 6, 2601-2610

Staden, R. (1980) Nucleic Acids Research 8, 817-825

Sutcliffe, J.G. (1978) Nucleic Acids Research 5, 2721-2728

Tinoco, I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. & Gralla, J. (1973) Nature N.B. 246, 40-41

Tobler, H. (1975) in Biochemistry of Animal Development Vol III, ed H. Weber, pp 91-143. Academic Press

Toniolo, D., Meiss, H.K. & Basilico, C. (1973) Proc.Nat.Acad.Sci.USA 70, 1273-1277

Vass, J.K. & Maden, B.E.H. (1978) Eur.J.Biochem. 85, 241-247

Veldman, G.M., Brand, R.C., Klootwijk, J. & Planta, R.J. (1980a) Nucleic Acids Research 8, 2907-2920

Veldman, G.M., Klootwijk, J., de Jonge, P., Leer, R.J. & Planta, R.J. (1980b) Nucleic Acids Research 8, 5179-5192

Walker, W.F. (1981) FEBS Letters 126, 150-151

Weinberg, R.A. & Penman, S. (1970) J.Mol.Biol. 47, 169-178

Wellauer, P.K. & Dawid, I.B. (1974) J.Mol.Biol. 89, 379-395

Wellauer, P.K., Reeder, R.H., Carroll, D., Brown, D.D., Deutch, A., Higashinakagawa, T. & Dawid, I.B. (1974a) Proc.Nat.Acad.Sci.USA 71, 2823-2827

Wellauer, P.K., Dawid, I.B., Kelley, D.E. & Perry, R.P. (1974b) J.Mol.Biol. 89, 397-407

Wellauer, P.K., Dawid, I.B., Brown, D.D. & Reeder, R.H. (1976a) J.Mol.Biol. 105, 461-486

Wellauer, P.K., Reeder, R.H., Dawid, I.B. & Brown, D.D. (1976b) J.Mol.Biol. 105, 487-505

Wensink, P.C. & Brown, D.D. (1971) J.Mol.Biol. 60, 235-247

Willems, M., Penman, M. & Penman, S. (1969) J.Cell.Biol. 41, 177-187

Winicov, I. (1976) J.Mol.Biol. 100, 141-155

Woese, C.R., Magrum, L.J., Gupta, R., Seigel, R.B., Stahl, D., Kop, J., Crawford, N., Brosius, J., Gutell, R., Hogan, J.J. & Noller, H.F. (1980) Nucleic Acids Research 8, 2275-2293

Young, R.A. & Steitz, J.A. (1978) Proc.Nat.Acad.Sci.USA 75, 3593-3597

Zwieb, C., Glotz, C. & Brimacombe, R. (1981) Nucleic Acids Research in the press (9 , 3621 – 3640)