NUCLEOTIDE SEQUENCE DETERMINATION OF *XENOPUS BOREALIS* AND HUMAN 18S RIBOSOMAL DNA

Fiona S. McCallum

Thesis submitted to the University of Glasgow for the degree of Doctor of Philosophy.

Department of Biochemistry

July, 1985

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

ACKNOWLEDGEMENTS

I would like to thank Professor Ted Maden for his excellent supervision and much appreciated encouragement throughout this project.

My thanks to Professors R. M. S. Smellie and J. Glover for providing facilities in the Departments of Biochemistry at Glasgow University and Liverpool University respectively. To Mrs. Maureen Wilde, Mr. Tom Fawcett and Mr. Ian Ramsden for photography. To Mrs. Gill Allen for an excellent job of typing the manuscript.

Special thanks to all my friends in C13 and 317.

Most of all I would like to thank my parents and Paul. To Mum and Dad, thank you for all your love and support throughout the years. To Paul, thank you for being there and for giving me something to look forward to.

ABBREVIATIONS

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

rRNA	ribosomal RNA
pre-rRNA	ribosomal precursor RNA
rDNA	genes for ribosomal RNA and associated spacer DNA
ETS	external transcribed spacer
ITS	internal transcribed spacer
NTS	non-transcribed spacer
bp	base pairs
kb	kilobases (1000 bases)
dNTP	deoxynucleotide-5'-triphosphate
ddNTP	dideoxynucleotide-5'-triphosphate
RF	replicative form
IPTG	isopropyl-beta-D-thio-galactopyranoside
X-GAL	5-bromo-4-chloro-3-indoyl-β-galactoside
TEMED	NNN'N'-tetramethylethylenediamine

CONTENTS

FIGURES AND TABLES

PAGE

PAGE

SUMMARY

Prior to this work, the complete sequence of Xenopus laevis 18S ribosomal DNA was known (Salim and Maden, 1981). Limited sequence data on Xenopus borealis included a few hundred nucleotides covering the 5' and 3' ends of the gene (Furlong and Maden, 1983). I have completed the 18S ribosomal DNA sequence for Xenopus borealis from clone pXbr101. As a result of this work, Xenopus laevis 18S ribosomal DNA has been found to contain an extra nucleotide which was formerly undetected in Maxam-Gilbert sequencing gels. The presence of this extra nucleotide was first demonstrated in Xenopus borealis by restriction analysis, and then in both species by dideoxy chain terminator sequencing. The Xenopus borealis sequence differs at only two points from the Xenopus laevis sequence. Both of these differences occur as a result of a base substitution. Both occur in regions of the sequence which are variable in comparisons between more distantly related species. The two differences are readily accommodated in the secondary structure model for Xenopus laevis 18S ribosomal RNA (Atmadja et al., 1984).

The two sites of difference between Xenopus laevis and Xenopus borealis 18S ribosomal DNA were checked in a range of clones from both species. These clones contained both amplifed and chromosomal DNA. The differences were seen to be fixed within each species. Therefore, there is no intraspecies heterogeneity in Xenopus laevis or Xenopus borealis around the two points of difference. This complements the earlier findings of Maden et al., (1982a), that Xenopus laevis 18S gene regions are homogeneous. This is in direct contrast to previous data for the transcribed spacers, which show intraspecies heterogeneity and extensive interspecies divergence. (Stewart et al., 1983, Furlong and Maden, 1983; Furlong et al., 1983).

I have also sequenced the complete 18S gene from human. The human sequence differs from the Xenopus borealis sequence at 123 positions (consisting of 44 insertions and 79 base substitutions). Again, these differences are in regions of

the sequence known to be phylogenetically variable. Comparison of the human 18S sequence with other mammalian data reveals fewer differences.

I have examined two models of 18S ribosomal RNA secondary structure (Xenopus laevis, Atmadja et al., (1984) and rat, Chan et al., 1984). The two models differ from each other in some regions. I have attempted to show (where possible) which arrangement is likely to be more correct. I have tried therefore to indicate the structure of a model which can accommodate both species and so is hopefully more representative of a consensus model for eukaryotic 18S ribosomal RNA.

CHAPTER 1

INTRODUCTION

Ribosomes are the organelles of protein synthesis. They are cytoplasmic structures consisting of RNA molecules and proteins. The RNA components of these structures constitute between 80-90% of the total RNA in the eukaryotic cell and thus comprise the major products of transcription.

Much study has been devoted to ribosomal RNA and to the genetic material that encodes this RNA. Four ribosomal RNA species occur in all eukaryotes designated 18S, 5.8S, 28S and 5S from their approximate sedimentation coefficients. Prokaryotic ribosomes contain 16S, 23S and 5S, but lack 5.8S rRNA as a separate RNA species.

In eukaryotes, the genes coding for 18S, 5.8S and 28S rRNA are located and transcribed together and are commonly referred to as ribosomal DNA or rDNA. 5S rRNA genes are referred to as 5S DNA and are separate from rDNA in most eukaryotes.

Recent studies have culminated in direct nucleotide sequence data on rDNA and rRNA from a wide range of organisms. Comparative analysis and secondary structure models are beginning to give us an insight into which regions of rRNA are conserved and so most likely to be necessary for proper functioning of the mature ribosome. The work described in this thesis concerns the nucleotide sequence analysis of the DNA that encodes 18S rRNA in two species of vertebrates. The main part of this Introduction describes the background which preceded and led to the present studies.

1.1 Structure of Ribosomal DNA

Eukaryotic rDNA consists of repeated units with the 18S, 5.8S and 28S genes

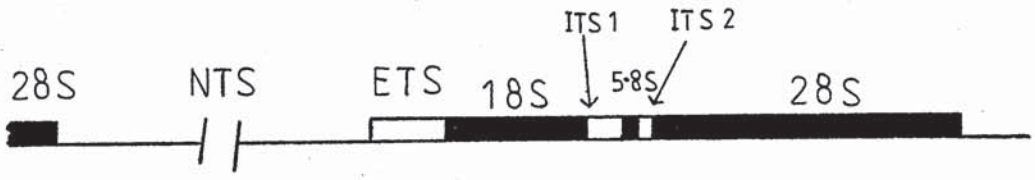


Figure 1.1 General structure of ribosomal DNA repeat

A transcription unit comprises all of the above minus the NTS.

NTS - non-transcribed spacer

ETS - external transcribed spacer

ITS 1, ITS 2, - internal transcribed spacers 1 and 2

arranged consecutively in the genome in the direction 5' ——> 3'. These gene regions, along with three spacer regions, external transcribed spacer (ETS), internal transcribed spacer 1 (ITS 1) and internal transcribed spacer 2 (ITS 2) are collectively termed a transcription unit. Tandemly repeated transcription units are separated by so-called non-transcribed spacer regions (NTS). The general structure is shown in Figure 1.1.

These tandemly repeated units are arranged head to tail at a single or several chromosomal locations termed nucleolar organisers, a nucleolar organiser being the term used to describe one cluster of rRNA genes. In Xenopus laevis, there is one nucleolar organiser per haploid genome for rDNA (Pardue, 1973) and several DNA clusters for 5S DNA (Pardue et al., 1973). In contrast, when considering human ribosomal genes, the arrangement is reversed. rDNA is arranged as tandemly repeated units in the nucleolar organiser regions of the five chromosome pairs Nos. 13, 14, 15, 21 and 22. (Henderson, et al., 1972), whereas 5S RNA genes are only present in one tandem array localised on chromosome No. 1(Steffensen and Duffey, 1974).

The structure of rDNA shown in Figure 1.1 is the arrangement that has most generally been found in eukaryotes. However, major variations occur in lower eukaryotes, including Saccharomyces (Rubin and Sulston, 1973; Valenzuela et al., 1977), Dictyostelium (Maizels, 1976) and Drosophila (Jordan, 1975; Jordan et al., 1976; Glover and Hogness, 1977; Pavlakis et al., 1979). In yeast and Dictyostelium, the repeating unit also contains 5S DNA. In Drosophila, both the 5.8S and 28S gene sequences are interrupted by the presence of additional transcribed spacers; there is also an intervening sequence in some 28S genes.

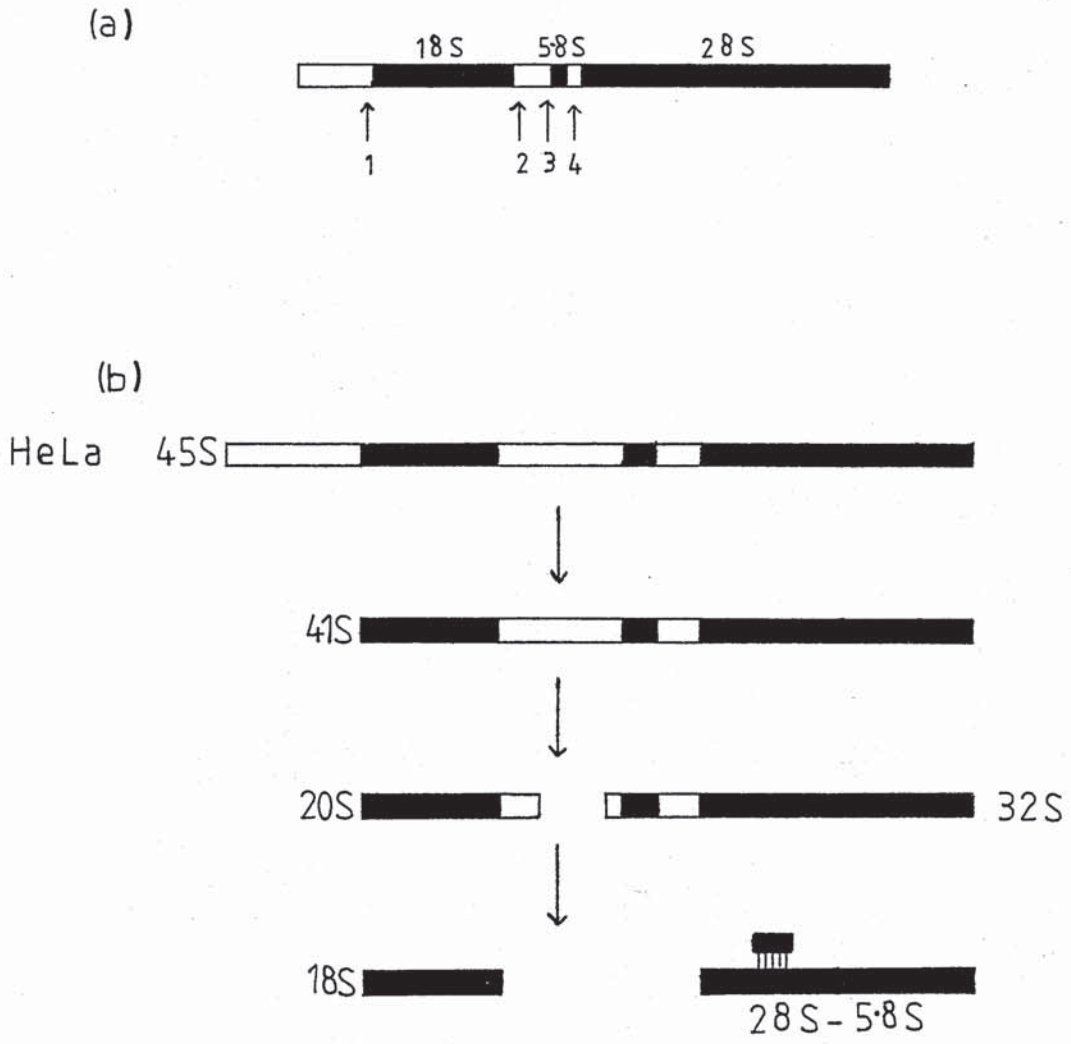
rDNA units demonstrate variation between species in both their overall size and number. The length of a transcription unit is reviewed by Lewin (1980). In many eukaryotes, this unit is approximately 8 kilobases (kb) in length, whereas in mammals such as man, it is 13 kb long. The difference reflects mainly an increase

in size of the transcribed spacers. However, there is also some variation in the size of the 28S gene (Loening, 1968; Schibler et al., 1975). There are also marked differences in the size of a complete rDNA repeat. Mouse, man and calf have total rDNA repeat lengths of between 30-45 kb (Arnheim and Southern, 1977; Wellauer and Dawid, 1979; Meunier-Rotival et al., 1979). In contrast, Xenopus has a complete repeat length of approximately half this size (11-15kb). These large differences are due to variation in size of the NTS. The number of copies of rDNA units varies between species. Dictyostelium discoideum has around 200 copies per haploid genome and Xenopus has around 500 copies (Reviewed by Long and Dawid, 1980). At present, there is disagreement over human rDNA gene dosage with estimates ranging from 50-220 copies per haploid genome (Reviewed by Wilson, 1982).

1.2 Transcription of DNA to RNA

In eukaryotes, each rRNA transcription unit is transcribed as one long precursor molecule which is subsequently processed to release the mature rRNA. The existence of pre-rRNA was first demonstrated by pulse labelling of L-cells and HeLa cells (Perry, 1962; Scherrer and Darnell, 1962). Label was initially incorporated into a 45S molecule which was shown to contain the large and small rRNA's. Subsequently, indirect evidence suggested the presence in pre-rRNA of sequences that are not conserved in ribosome maturation. Base composition studies showed that the precursor molecule has a higher G + C content than the mature 18S and 28S molecules suggesting that the "extra" DNA is rich in G + C (Willems et al., 1968; Amaldi and Attardi, 1968). These extra sequences are now called transcribed spacers and correspond to the similarly designated regions of rDNA as outlined above.

The location of these non-ribosomal sequences was first observed by electron microscopic studies. Comparison of the secondary structure maps of the 45S precursor molecule, intermediate molecules and the mature 18S and 28S rRNA's



L-cells 45S —> 41S —> 18S + 36S —> 32S —> 28S - 5.8S

Xenopus 40S —> 38S —> 18S + 34S —> 30S —> 28S - 5.8S

Figure 1.2 Processing pathways for ribosomal RNA

(a) shows the positions of the cleavage sites on a typical precursor.

(b) shows the cleavage patterns of three different species.

demonstrated the presence of transcribed spacer regions, although the polarity of transcription was wrongly defined (Wellauer and Dawid, 1973). However, conclusive evidence based on sequence data now shows the arrangement to be 18S ——> 28S in the direction 5' ——> 3' along the primary transcript.

Studies by Speirs and Birnstiel (1974) suggested that 5.8S rRNA was located in the spacer region between 18S and 28S rRNA. Recent DNA sequence analysis has shown this to be the case (Hall and Maden, 1980).

1.3 Processing of the Primary Transcript

Processing of the primary transcript involves events of two kinds. One is the modification of nucleotides. The main types of modification are conversion of uridine to pseudouridine or methylation either of the ribose or of the nucleotide base. Most of these modifications occur on the intact precursor molecule. The second processing event is the elimination of the transcribed spacers to give rise to the three mature rRNA species.

For convenience, I will consider these processing events in reverse order. A review is given by Perry (1976).

1.3.a. Cleavage

Most processing pathways can be described in terms of four principle events involving cleavages at or near sites 1 to 4 in Figure 1.2.a. Many research groups have examined this processing sequence in cells from various organisms including HeLa cells (Penman, 1966; Weinberg and Penman, 1970; Maden et al., 1972; Wellauer and Dawid, 1973), mouse L cells (Wellauer et al., 1974a) and Xenopus laevis (Wellauer and Dawid, 1974). The order of cleavage for these three organisms is shown in Figure 1.2.b.

In general, the processing of the primary transcript procedes stepwise along the molecule rather than by randomly timed attack on all cleavage sites. However,

there does appear to be some degree of flexibility in the processing events. For example, the studies of Weinberg and Penman, (1970) and Wellauer and Dawid, (1973), although agreeing with the pathway for HeLa rRNA maturation shown in Figure 1.2.b., identified other intermediates in low abundance which could not be fitted into the general pathway.

1.3.b. Nucleotide modification

As indicated earlier, two major categories of nucleotide modification occur. About one hundred uridines are converted to pseudouridines in the rRNA sequences of several vertebrates which have been examined. This modification occurs within the precursor molecule and locations remain to be established although there are some details of the timing.

Methylations also take place on the precursor molecule during or immediately after transcription, but are confined to regions of the transcript that are processed to mature rRNA species (Greenberg and Penman, 1966). Relatively abundant rRNA methylation appears to be a common feature of eukaryotes. However, the number of RNA methyl groups per ribosome has been found to vary from 70 in yeast to over 100 in vertebrates (Klootwijk and Planta, 1973; Maden and Salim, 1974; Khan et al., 1978).

More recent studies have established the distribution of methyl groups along Xenopus laevis rRNA (Maden and Reeder, 1979; Brand and Gerbi, 1979; Maden, 1980; Salim and Maden, 1981). Methylation patterns are distinctly different in the 18S and 28S rRNA, but both show non-random distribution. In 18S rRNA, more than half the ribose methylations are concentrated in the 5' end of the molecule, whereas base methylations are clustered near the 3' end. Distribution is quite different in the 28S molecule. The 5' end is relatively free of methyl groups, the central region has patches of methylation, but the highest concentration is found in a 1,100 nucleotide stretch, a few hundred nucleotides from the 3' end.

Comparative analyses have demonstrated a high degree of conservation of

methylation patterns over a range of vertebrates. "Fingerprint" analysis of Khan et al., (1978) shows this to be the case for X. laevis, chick, mouse, hamster and human. Even between the two most distantly related species (man and Xenopus) there is 95% homology. Brand and Gerbi, (1979) showed that the distribution of methyl groups in the 28S rRNA correlates with the location of highly conserved sequences within this RNA species over an even broader range of eukaryotes.

1.4 Detailed Study of Ribosomal DNA Structure by Physical and Chemical Methods

1.4.a. Xenopus laevis

Much of our initial knowledge of rDNA structure has come from extensive studies on Xenopus laevis, since this animal has rDNA which can be physically purified. The first work to characterise the arrangement of genes coding for rRNA took advantage of the fact that the rDNA of X.laevis has a higher G + C content than the average for genomic DNA. After fragmenting the chromosomal DNA, an rDNA fraction was isolated directly as a satellite of distinct buoyant density (Brown and Weber, 1968a,b). RNA-DNA hybridisation techniques supported the idea that the 18S and 28S genes alternate, but are interspersed with DNA which is not homologous to either gene. This arrangement was further supported by Birnstiel et al. (1969). Renaturation kinetics determined that the arrangement of rDNA shows a highly reiterated base sequence.

In Xenopus, the ribosomal genes are amplified during oogenesis, (increasing the amount of rDNA greatly over the chromosomal level), thereby providing for the increased needs of the very large oocyte cells for ribosomes. This amplified rDNA remains extrachromosomal. Its buoyant density differs slightly from that of chromosomal rDNA since it lacks methylation on the 5 position of cytosine residues (Dawid, et al. 1970). Comparative analysis showed no difference in base composition between the two classes. Both contain 67% G + C. Both classes demonstrated biphasic melting curves, suggesting that two regions of the DNA

differ markedly in their base composition.

Wellauer et al (1974b) began a more detailed analysis of the structure of the ribosomal repeating unit. Amplified rDNA was cut with restriction endonuclease EcoRI. Using electron microscopic techniques, the secondary structure maps of these single-stranded EcoRI fragments were compared with those of X. laevis rRNA and uncut rDNA. EcoRI restriction generated two classes of fragments, one containing 90% of the 28S gene and with a molecular weight of 3×10^6 and another containing 80% of the 18S gene and all of the NTS with molecular weights ranging from $4.0 - 5.9 \times 10^6$. Thus it was found that the NTS displays length heterogeneity.

The molecular basis of this length heterogeneity was elucidated in several stages. Wellauer et al (1976a) compared four spacer-containing EcoRI fragments by optical melting and by homoduplex and heteroduplex mapping. The NTS was shown to be made up of an internally repeated simple sequence. Analysis of the arrangement of length heterogeneity within the repeating unit of rDNA highlighted a difference between chromosomal and amplified rDNA not detected in earlier studies (Wellauer et al., 1976b). Single-strand rDNA from separate preparations of amplified and chromosomal rDNA were each annealed with a homogeneous cloned spacer-containing EcoRI fragment and the resulting structures viewed by electron microscopy. The arrangement of length heterogeneity was found to be very different in the two classes of rDNA. Between 50-70% of adjacent repeats in a given molecule of chromosomal rDNA differ in length, whereas amplified rDNA repeats show much less variation. Bird (1977) looked at this phenomenon by analysing amplified rDNA from individual oocytes and showed that each oocyte amplifies only a small number (usually one) of the repeats present chromosomally.

Botchan et al (1977) studied the distribution of several restriction sites within the NTS, thus extending the earlier studies of Wellauer et al (1976a,b) by showing that the NTS contains a restriction map characteristic of highly repeated sequences, with a higher order repeating pattern being superimposed on short repeating sequences.

Studies of the NTS culminated in the direct analysis of the nucleotide sequence (Boseley et al., 1979; Moss and Birnstiel, 1979; Moss et al., 1980). In recent years, other regions of the repeating unit of X. laevis rDNA have been sequenced (Salim and Maden, 1980; Hall and Maden, 1980; Salim and Maden, 1981; Maden et al 1982b; Ware et al., 1983). Some of these studies are described later.

1.4.b. Comparison between X. laevis and X. borealis

The studies carried out on X. laevis have given a well documented picture of rDNA structure and of the structural variation among repeats, but no indication of relationships between species or the evolution of such repeated sequences. Early comparisons were made between X. laevis and X. borealis (wrongly named as X. mulleri in some studies).

Cross hybridisation of labelled rDNA from X. laevis and X. borealis respectively, with rDNA from both species showed the gene coding regions to be virtually indistinguishable. In contrast to this, cross hybridisation involving the NTS region showed very little hybridisation (Brown et al, 1972).

Forsheit et. al., (1974) analysed the genes and spacers in the two species by heteroduplex mapping and visualisation of RNA-DNA hybrids. These studies again showed perfect duplex formation corresponding to the presumed 18S and 28S gene regions. These duplex regions were separated by a single loop corresponding to the ITS. Each transcription unit was separated from the next by a large region of partial homology corresponding to the NTS.

Wellauer and Reeder (1975) compared amplified rDNA from the two species by looking at their secondary structure maps, revealed by electron microscopy. Gene regions showed identical structures between the two species, while patterns in the NTS differed. Restriction analysis of X. borealis rDNA with EcoRI also gave rise to two fragments, one homogeneous in length and the other heterogeneous in size due to length variability in the NTS region.

Therefore, it appears that X. laevis and X. borealis have almost identical

rRNA coding regions, but differ quite markedly from each other in both their transcribed and non-transcribed spacers.

1.4.c. Comparisons among a variety of species

Many comparative studies have been carried out on a wide range of organisms and by a variety of methods.

RNA base composition

Amaldi (1969) collated the results of various groups on the base composition of rDNA among a wide range of species, both eukaryotic and prokaryotic. The base composition of rRNA components 16S-18S and 23S-28S is quite variable among different organisms. In most cases, the two rRNA components are significantly different from each other. In addition, the base composition shows an evolutionary pattern. In general, vertebrates tend to have a higher G+C content than invertebrates, other eukaryotes or prokaryotes. For example, E. coli has a G+C content of approximately 50%, Xenopus has between 55-60% and HeLa cells show a G+C content of approximately 60-65%.

RNA:DNA hybridisation

Sinclair and Brown (1971) hybridised ^{3}HrRNA from X. laevis with DNA from over 50 organisms including prokaryotes and eukaryotes. The resulting hybridisation profiles were analysed. It was apparent from these results that X. laevis shows extensive homology with eukaryotes as distantly related as plants and fungi. No measureable homology was detected between X. laevis rRNA and the DNA from prokaryotes.

The important conclusion from these studies was that there is some rRNA homology among all eukaryotes despite considerable differences in total rDNA (including NTS and other spacer regions).

RNA:DNA hybridisation - finer details

Gerbi (1976) went on to look in more detail at the conservation of homologous regions within rRNA's of eukaryotes. Molecular hybridisation studies were carried out with rRNA's from a unicellular organism, three invertebrates, an amphibian (X. laevis) and a mammal (mouse), hybridised to heterologous rDNA. Results determined for the first time that cross-hybridisation is due to conserved regions common to all eukaryotic rRNA's.

Gourse and Gerbi (1980) looked closely at the location of these evolutionarily conserved regions within the rRNA. Restriciton mapping, DNA-DNA and RNA-RNA hybridisation techniques were used to locate conserved sequences. These studies led to the detection of evolutionary conservation only in gene coding regions. However, a feature of hybridisation studies is that stretches of at least 20bp are required to give a positive result, so any shorter region of conservation in the spacers would not be detected. Three regions including one near the 3' end of 18S and two near the 3' end of 28S rRNA are conserved over great evolutionary distance, even between X. laevis and E. coli. Several regions within 18S and 28S rRNA show high conservation between yeast and Xenopus, but others show little conservation.

Therefore, it appears that variation in the structure of ribosomal genes can be a measure of evolutionary distance between a wide range of organisms. These results have led to speculation as to the function of individual regions within rRNA which have been conserved during evolution.

1.4.d. Study of ribosomal DNA structure in Mammals

In recent years, several groups have begun to look at the detailed structure of rDNA from mammals. One disadvantage of the earlier studies on X. laevis is that most analyses were carried out on animals collected from the wild, making it difficult to detect differences in rDNA structure between individuals with a known relationship. However, the study of rDNA from mammals such as rat, mouse and human will give an indication of structural relationships between individuals and to

changes occuring over several generations.

Arnheim and Southern, (1977) characterised the ribosomal genes of mice and human by restriction analysis with EcoR1 and Hind III. Individual mice or humans showed a heterogeneous pattern of restriction fragments resulting from differences in the non-transcribed spacer DNA. Individual mice of the same strain had identical patterns, but as the relationships became more distant, variation was seen in the restriction patterns.

Arnheim (1979) and Grummt et al. (1979) characterised cloned EcoRI fragments of mouse rDNA. Different clones varied in the size of their NTS sequence, but all showed a homogenous transcribed region. Fuke et al (1981) extended this study to rat rDNA. Again different clones demonstrated heterogeneity in the NTS regions. In contrast to this, restriction sites in the 18S region were seen to be conserved between rat and mouse.

Detailed study of human rDNA is also in progress. The major EcoRI fragment of human rRNA genes was visualised by Southern (1975), and subsequently characterised by the restriction analysis of Arnheim and Southern (1977). Wellauer and Dawid (1979) further analysed human rDNA by restriction analysis and electron microscopy. R-loop mapping revealed the positions of EcoRI cleavage sites. One type of repeating unit contained 4 sites, 2 in the gene coding regions and 2 in the NTS, while another type of repeating unit lacked one of the NTS sites. The majority of repeats were seen to be uniform in length with only a few exceptions. So there does appear to be length variation in the human NTS, but it appears to be discrete rather than continuous as seen in Xenopus.

Human rDNA fragments have been successfully cloned into bacteriophage λ. DNA from a fetal lung cell line enriched for rDNA was cloned in λ phage vector Charon 16A. Of 978 clones assayed with labelled 18S and 28S rRNA, 11 contained a 3.8×10^6 dalton fragment of human 18S rDNA. Subsequent restriction analysis demonstrated a variation only in orientation (Wilson et al., 1978).

Erickson et al (1981) carried out comparative analysis on cloned EcoRI-A

fragments. This fragment contains 0.2kb of the 3' end of 18S rDNA, 2.5kb of internal spacer and 5.8S gene and 4.5 kb of 28S rDNA. Comparisons were made of two types of tissue (fetal human liver and placenta) from three individuals. Restriction enzyme digestion and electron microscopic results gave identical results for all 6 cloned fragments. There were no differences in either the gene regions or the internal transcribed spacers. This indicates a remarkable mechanism of conservation of these rDNA species within and between individuals.

Therefore, recombinant DNA technology provides an accurate method of amplifying DNA sequences to allow structural analyses to be undertaken.

1.5 Direct Sequence Analysis

In recent years, several groups have initiated studies to determine the structure of the ribosomal repeating unit at the level of the nucleotide sequence. By looking at the nucleotide sequence directly, we are able to detect the arrangements and organisation of DNA structure that give rise to the characteristic patterns shown previously by physical and chemical methods.

1.5.a. The Non-transcribed spacer

It is a well established phenomenon that the non-transcribed spacer contains regions of highly repeated sequences, with higher order repeating elements being superimposed on these shorter sequences. Study of the NTS of *X. laevis* has culminated in the direct analysis of the nucleotide sequence (Boseley *et al.*, 1979; Moss and Birnstiel, 1979; Moss *et al.*, 1980). One feature of length heterogeneity is a varying number of recognition sites for the restriction endonuclease Bam H1. These "Bam Islands" are closely followed by multiple Alu I recognition sites. There is speculation as to a function for this repetitive structure within the NTS. The Bam Islands closely resemble the region preceding the site of initiation of transcription (Sollner-Webb and Reeder, 1979), giving rise to the suggestion that

they act as sinks for RNA polymerase 1 and thus serve to concentrate polymerase molecules and enhance transcription (Moss, 1983). This phenomenon may not be unique to X. laevis as a similar repeating structure has been demonstrated in Drosophila melanogaster and in X. borealis and X. clivii (Coen and Dover, 1982; Bach et al., 1981).

1.5.b. Gene regions

As of yet, very little is known as to the structural and functional role of rRNA in the mature ribosome. By looking at comparative sequence data we can begin to see parts of the sequence that have remained constant during long periods of evolutionary time and are therefore likely to be critical for ribosomal function.

Small subunit rRNA genes

In recent years, a fairly comprehensive list of sequence data have become available for a wide range of organisms. Salim and Maden (1981) aligned the 18S sequence of Xenopus laevis with that of yeast (Rubstov et al., 1980) and showed a pattern of extensive but interrupted homology between the two species. Addition of the E. coli 16S sequence to this picture (Brosius et al., 1978) shows much lower sequence homology between prokaryotic and eukaryotic rRNA's. Nonetheless, there can be seen many short tracts of clear homology (Zwieb et al., 1981; Stiegler et al., 1981). Recently published mammalian data for rat (Torczynski et al., 1983; Chan et al., 1984) mouse (Raynal et al., 1984) and rabbit (Connaughton et al., 1984) support the picture of extensive sequence homology among eukaryotic 18S rRNA's.

These patterns of primary structure can be interpreted in terms of secondary structure. Secondary structure models have been proposed for prokaryotic, eukaryotic and mitochondrial small subunit rRNA's (Ross and Brimacombe, 1979; Woese et al., 1980; Noller and Woese, 1981; Zwieb et al., 1981; Stiegler et al., 1981; Atmadja et al., 1984; Chan et al., 1984). The most striking feature of all the

models is that they conform to a common plan. All the rRNA's consist of 4 domains, each domain being bound by a short helix. The helices are formed by long range interactions between regions that are distant from each other in the primary structure. The models support the concept that secondary structure of rRNA has been extensively conserved throughout evolution. Therefore, when considering the primary structure of the gene, it is apparent that certain regions must be under intense selective pressure to remain conserved and retain their ability to form a particular secondary structure. Other regions do show quite marked differences in sequence, but obviously, must not be allowed to lay threat to the stability of the secondary structure which must be a pre-requisite for effective ribosome function. However some of the most highly conserved sequences are those which are not base paired in the secondary structure model. It is not yet known whether these regions are truly single-stranded or are in complex tertiary structures.

Large subunit rRNA genes

Study of the large subunit rRNA species is not as advanced as that for the small subunit rRNA's. However sequence data are only now being accumulated including E. coli 23S (Brosius et al., 1980), S. Cerevisiae 26S (Georgiev et al., 1981) Physarum polycephalum 26S (Otsuka et al., 1983). X. laevis 28S (Ware et al., 1983) and rat 28S (Chan et al., 1983). Again, distinct regions of the gene show evolutionary conservation.

1.5.c. Transcribed spacer regions

A different picture of sequence variation emerges upon extending this analysis to the transcribed spacers. These regions have evolved much more rapidly than the gene coding regions. Sequence data are available for the ETS of yeast (Skryabin et al., 1979a), X. laevis (Maden et al., 1982b), and X. borealis (Furlong et al., 1983). Data on ITS sequences are available for yeast (Skryabin et al., 1979b.,

Veldman et al., 1980, 1981), X. laevis (Hall and Maden, 1980), X. borealis (Furlong and Maden, 1983), mouse (Michot et al., 1983) and rat (Subrahmanyam et al., 1982). These data reveal practically little or no homology between phylogenetically distant species. Even comparison between the two closely related species of Xenopus shows extensive differences. However, there are several short sequences which are identical in both species, but these are shown to be displaced in X. borealis with respect to X. laevis, implying that insertions and deletions have played a major role in transcribed spacer divergence in Xenopus. A similar pattern has been shown between the ITS sequences of mouse and rat.

As of yet, the function of these transcribed spacer regions is unknown. However any function is unlikely to be sequence dependent or dependent upon a particular secondary structure. In E. coli, the sequences flanking the small subunit rRNA species interact to form an extensive base paired structure which constitutes a processing site for RNase III (Young and Steitz, 1978). However, neither X. laevis or X. borealis appear to have the potential to form such structures between the 18S flanking regions (Maden et al., 1982b; Furlong et al., 1983). Recent reports suggest a possible role for ITS 2 in the conversion of 32S rRNA to mature 28S and 5.8S rRNA (Crouch, et al., 1983, Bachellerie et al., 1983).

1.5.d. Rate of evolution of gene regions versus spacer regions

Comparison of the nucleotide sequence data across a wide range of organisms demonstrates the greater need to conserve gene sequences as opposed to spacer sequences. Recent work on Xenopus demonstrates quite dramatically the very great difference in the rate of evolution of these two classes of DNA.

Maden et al (1982a) searched for heterogeneities in the 18S coding sequences in cloned and uncloned material from X. laevis. The results showed complete homology with the previously established sequence for pXlr101 (Salim and Maden, 1981). In contrast to this, Stewart et al (1983) examined the three transcribed spacer regions of X. laevis in cloned and uncloned amplified rDNA from oocytes

and cloned chromosomal rDNA from erythrocytes. Heterogeneities were found including single base changes and length variants of one to several nucleotides. So even within one species we can see variation in spacer sequence.

Furlong and Maden (1983), in addition to demonstrating the patterns of major divergence between the ITS's of X. laevis and X. borealis, also compared the 5' and 3' ends of the 18S gene, the whole of the 5.8S gene and the 5' end of the 28S gene. In contrast to the spacer sequences, only three differences were found in the gene coding regions. Therefore, even between too such closely related species as X. laevis and X. borealis, we can see the beginnings of spacer divergence.

1.6 Why should more sequence data be added to this picture?

Very little is known as to how rRNA contributes to the proper functioning of the mature ribosome. It is only now that we are beginning to see areas that are obviously important for ribosome formation. By the production of secondary structure models, we can begin to see a common shape. However, there are still some discrepancies between the current models which may be resolved by the addition of more sequence data. Such data should also aid the identification of specific sequences vital for ribosome function. Ultimately, we hope to assign particular functions to specific sequences, as has been achieved for E. coli rRNA.

The work of this project has analysed the primary strucutre of 18S rDNA from X. borealis and from human. X. borealis was an obvious choice to study, as a great deal of comparative analysis between this species and X. laevis was being carried out in our own laboratory, as described above. Human rDNA was chosen because there is the possiblity that studies on human rDNA might lead to knowledge which may be medically applicable. (In this respect, it may be worthwhile noting that human rRNA genes are located on chromosomes associated with common birth defect syndromes).

Research must ultimately be directed at the level of the RNA to obtain a

better understanding of phenomena such as nucleotide modifications and the processing and maturation pathways. For many RNA studies it is advantageous to know the DNA sequence. This is true, for example, for the positioning of rRNA methyl groups. Moreover, prior knowledge of the corresponding rDNA sequence may eliminate the need to carry out full RNA sequencing, a technique which is not as well established as DNA sequencing. The DNA sequencing carried out in this work should be useful for the various objectives outlined above.

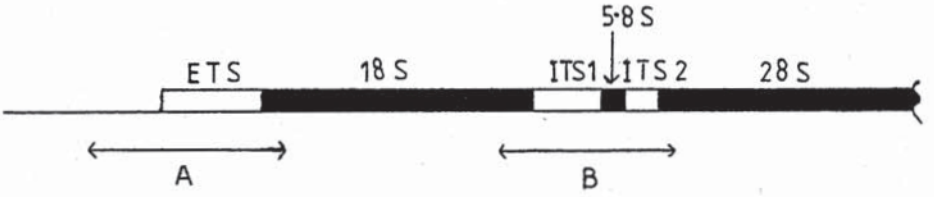


Figure 2.1 Comparative analysis of X. borealis and X. laevis rDNA prior to this present work

Regions A and B as shown above were the areas compared between the two Xenopus species at the outset of this project. The large internal region of the 18S rDNA had been sequenced in X. laevis only. Analysis of this region in X. borealis was undertaken in this present work.

CHAPTER 2

OBJECTIVES AND EXPERIMENTAL APPROACH

2.1. Outline of Studies

As stated in the Introduction, the work of this project has led to the elucidation of the primary structure of both Xenopus borealis and human 18S rDNA.

However, the initial aims were different. It was hoped to analyse the primary structure of the whole of the human rDNA transcription unit. To this end, it was necessary to detect the rDNA sequences within human genomic DNA and to clone and amplify these before carrying out sequence analysis. In particular, I wanted to cut human DNA with a restriction enzyme which would cleave specifically to give the 18S, 5.8S and 28S rRNA genes in a single large fragment. Wilson et al (1978) had shown by Southern blotting experiments that the restriction enzyme Sal I cuts to the 5' side of 18S rDNA and to the 3' side of 28S rDNA, but not in between, thus fulfilling the requirement of the three rRNA genes being present together in one fragment. The initial experiments were directed towards cloning the large rDNA fragment. These experiments were unsuccessful at the time, but a brief outline is given in Chapter 5.

Meanwhile, the comparative sequence analysis between corresponding segments of X. borealis and X. laevis rDNA was being carried out in our laboratory (Furlong and Maden, 1983; Furlong et al., 1983). The regions covered are shown in Figure 2.1. The analysis showed marked divergence between the two species in their transcribed spacers. In contrast to this, the few hundred nucleotides of rRNA coding regions included in these comparative analyses showed only minimal divergence. In particular, only one difference was found towards the 3' end of the

18S gene. Nevertheless, the then available comparative sequence data on eukaryotic 18S rRNA genes had shown that some parts of the 18S sequence were phylogenetically much more variable than other parts.

Given this interesting but incomplete picture of evolutionary rates in different parts of the ribosomal transcription unit, it was desirable to extend the comparative sequence data between X. borealis and X. laevis which were clearly a good model system for such studies. I therefore decided to redefine the initial aims of the project and sequence the large hitherto unsequenced internal region of the X. borealis 18S gene.

Having completed the X. borealis 18S gene sequence, I went on to look at the points of difference between X. borealis and X. laevis in a range of cloned 18S genes from the two species. For a difference to occur between the two species, a nucleotide substitution must have occurred in one or other species since the time of their divergence from a common ancestor. If this respective mutation had occurred relatively recently, then we might expect to see sequence heterogeneity in the respective species, since the mutation will as of yet not be fixed throughout the gene population. I therefore examined several clones from both species to look for any evidence of intraspecies heterogeneity.

Finally, I returned to the human project. My interest was now directed towards the human 18S rDNA sequence, making it less important to have the complete rDNA repeat cloned in one piece. EcoR I clones of human rDNA had been prepared (Wilson et al., 1978; Erickson et al., 1981). Clones were obtained from Wilson's group and the human 18S rDNA sequence was determined in the later part of this work.

2.2. rDNA Clones

The rDNA clones used in the course of this study are described in Figure 2.2 and Table 2.1.

Figure 2.2 Ribosomal DNA fragments contained in the clones used throughout this study

(a) Xenopus fragments: The structure of slightly more than one repeating unit of Xenopus rDNA is shown. Restriction sites relevant to this study are noted. All X. borealis clones contain amplified rDNA, as do X. laevis clones pXlrl01, pXlrl02 and pXlrl03. X. laevis clones pXlcrl, pXlcr2, pXlcr3, pXlcr4 and pXlcr5 contain chromosomal rDNA.

(b) Human fragments: Slightly more than one transcription unit of human rDNA is shown. The restriction sites used for cloning are noted. The rDNA fragments contained in clones pHrA, pHrB and pHrB-SE are indicated. Clone pHrB was not used in this present study. However, I have included it here to show the derivation of clone pHrB-SE.

Note: I have depicted the 5.8S gene of human rDNA within the internal transcribed spacer. As of yet, the precise location of this rDNA species has not been mapped. However, it is contained somewhere within the internal transcribed spacer.

Cloning of all of these rDNA fragments, from Xenopus and Human, is summarised in Table 2.1.

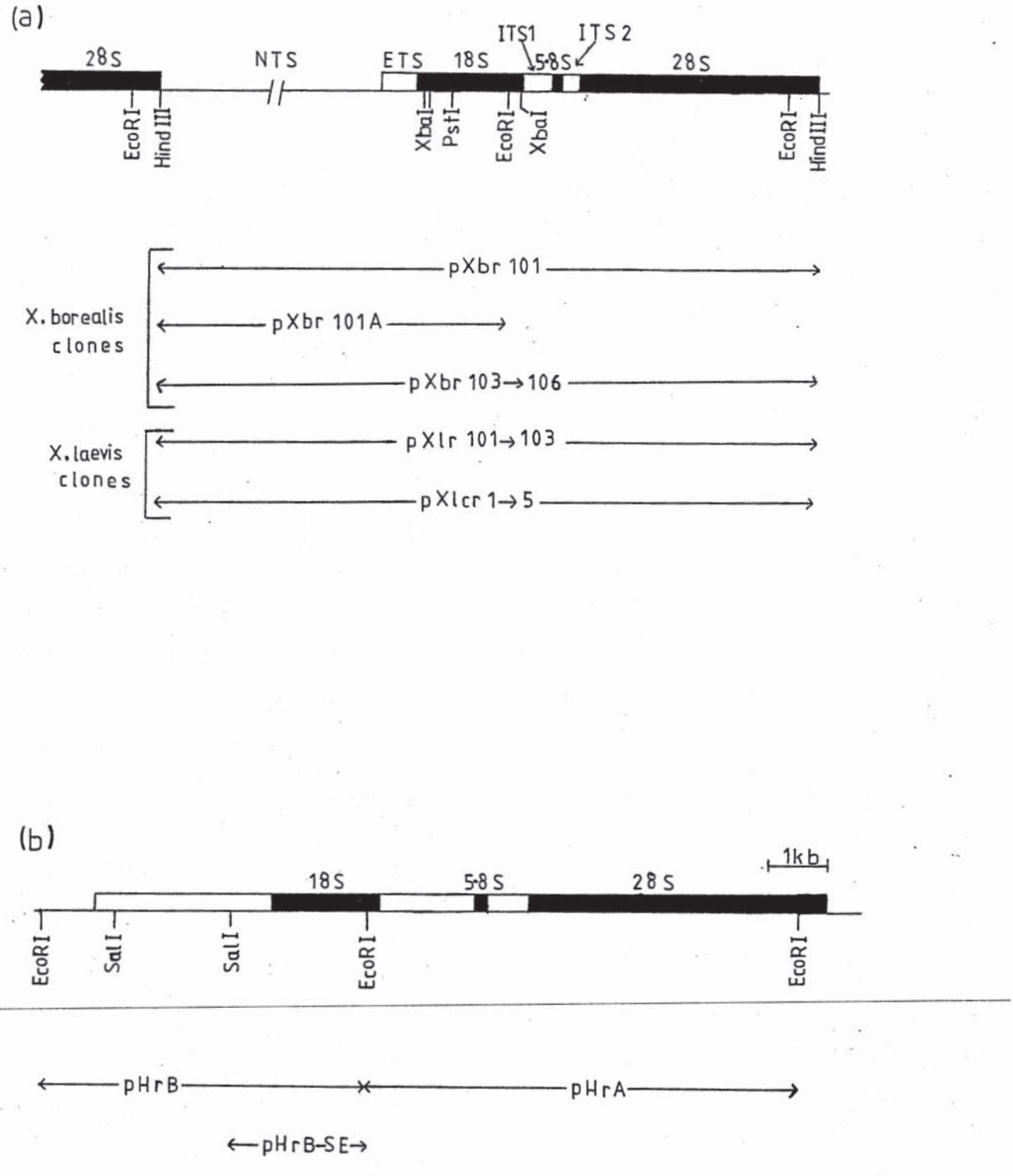

(a)
28S
NTS
ETS
18S
ITS1
ITS2
5·8S
28S
EcoRI
HindIII
XbaI
PstI
EcoRI
XbaI
EcoRI
HindIII
X. borealis clones
pXbr 101
pXbr 101A
pXbr 103→106
X. laevis clones
pXlr 101→103
pXlcr 1→5
(b)
1kb
18S
5·8S
28S
EcoRI
SalI
SalI
EcoRI
EcoRI
pHrB
pHrA
pHrB-SE

FIGURE 2.2

SPECIES	CLONE	VECTOR	ANTIBIOTIC MARKER	
X. borealis	pXbr101	pMB9	Colicin E1 Immunity	Cloned by R.Reeder -unpublished results subcloned from pXbr101 by B. E. H. Maden
"	pXbr101A	pAT153	Amp. resistance	
X. borealis	pXbr103	pMB9	Colicin E1 Immunity	Cloned by R.Reeder - unpublished results
"	pXbr104	"	"	
"	pXbr105	"	"	
"	pXbr106	"	"	
X. laevis	pXlr101	pMB9	Colicin E1 Immunity	Cloned by R.Reeder - unpublished results
"	pXlr102	"	"	
"	pXlr103	"	"	
X. laevis	pXlcr1	pAT153	Amp. resistance	Cloned by M.A. Stewart, (1983), Stewart et al.,(1983)
"	pXlcr2	"	"	
"	pXlcr3	"	"	
"	pXlcr4	"	"	
"	pXlcr5	"	"	
Human	pHrA	pBR322	Amp. and Tet. resistance	(Subcloned from λ by (Wilson's group. Subcloned from pHrB by Wilson's group.
"	pHrB	"	"	
"	pHrB-SE	"	Amp resistance	

Table 2.1 rDNA clones used in this study

The ribosomal DNA fragments contained in the above plasmids are shown in Figure 2.1.

pHrA and pHrB were originally cloned into bacteriophage λ vectors. (Wilson et al., 1978; Erickson et al., 1981).

As already stated in Figure 2.1, clone pHrB was not used in this analysis, but is included in the Table to show the derivation of clone pHrB-SE.

Amp- Ampicillin

Tet- Tetracycline

The X. borealis clones contain extrachromosomal amplified rDNA purified from X. borealis oocytes. X. laevis clones were obtained from two sources. pXlr101, 102 and 103 contain amplified rDNA purified as for X. borealis.pXlcr1-5 contain chromosomal rDNA prepared from X. laevis erythrocytes. Human rDNA clones pHrA and pHrB were intially cloned into bacteriophage λ vectors before being subcloned into pBR 322. Clone pHrB-SE was subcloned from pHrB.

pXbr101 and pXbr101A were used to determine the nucleotide sequence of the 18S gene region of X. borealis. All other Xenopus clones were used in a comparative analysis of the differences found between X. borealis and the established X. laevis sequence (Salim and Maden, 1981).

The human 18S rDNA sequence was determined by analysis of clones pHrB-SE and pHrA.

2.3. DNA Sequencing Methods

2.3.a. Chemical Method of Maxam and Gilbert

This method of DNA sequencing has been well established for a number of years (Maxam and Gilbert, 1977, 1980). A terminally labelled DNA fragment undergoes four different modification reactions, each one specific for a particular nucleotide. Each reaction gives rise to an array of DNA fragments of different lengths. These fragments are resolved by size on a polyacrylamide gel and subsequent autoradiography allows the DNA sequence to be read directly.

2.3.b. Dideoxy Chain Terminator Method of Sanger

This method was first developed during the same period as the chemical method (Sanger et al., 1977). It exploits the ability of DNA polymerase 1 to faithfully synthesise a complementary copy of a single-stranded DNA template. If a short fragment of DNA is annealed to its complementary site on the template, and the four dNTP's are provided, the enzyme will extend this primer 5' —> 3' until

a full complementary strand is synthesised, giving a double-stranded molecule. This fact is utilised in the method as follows. Dideoxynucleotide triphosphates (ddNTP's) lack a 3'-OH group. Therefore, if one of these molecules is added to the growing strand of DNA, chain termination will occur since there is no 3'-OH group available for formation of the next phosphodiester bond. As with the chemical method of sequencing, four separate reactions are carried out. Each reaction mix contains all four dNTP's, one of which is labelled, together with one of the ddNTP'S. By carefully controlling the ratio of dNTP to ddNTP, a whole range of DNA fragments of different lengths are produced. The products of each reaction can be separated, as for the chemical method, on polyacrylamide gels and the sequence read from an autoradiograph. In practice, the large fragment of DNA polymerase 1 (Klenow fragment) is used in the sequencing reactions. This lacks the 5' ——>3' exonuclease activity of the intact enzyme which could otherwise result in degradation of the newly synthesised strand from the 5' primer end.

For two reasons, the chemical method has until recently been more widely used than the dideoxy method. The chemical method can be carried out on double-stranded or single-stranded material, whereas the dideoxy method is specific for single-stranded DNA. Until recently there were no well established procedures for producing this essential single-stranded material. Another disadvantage of this second method was the requirement for a new primer sequence every 200-300 bases. However, these limitations have now been overcome, as described below.

2.3.c. M13 cloning vectors

M13 is a single-stranded filamentous phage whose life-cycle has been exploited to produce single-stranded templates required for dideoxy sequencing. The phage enters a suitable host cell (E. coli F') by way of the F pilus. This single-stranded DNA is converted to the double-stranded replicative form (RF). During subsequent replication, large amounts of newly synthesised single-stranded DNA are produced. These are packaged into viral coat proteins and extruded from the

cell without lysis. Therefore, it can be seen that the insertion of foreign DNA into the RF of the phage would provide a simple way of producing large amounts of the single-stranded template necessary for dideoxy sequencing. Messing et al., (1977) first showed that a stable recombinant could be formed. A Hind II fragment of the lac regulatory gene from E. coli was inserted into wild type M13. This insert contains the DNA sequence for the promoter, operator and the first 145 amino acid residues of the β-galactosidase gene (α-peptide). Gronenborn and Messing, (1978) introduced an EcoRI site into the β-galactosidase gene and produced the first M13 cloning vector, M13 mp 2.

Although M13 does not lyse its host cell, phage growth does retard the growth rate of the cell. On plating out, cells transfored with M13 are identified as turbid plaques on a lawn of uninfected cells. Inclusion of the lac DNA in M13 vectors gives a method for distinguishing recombinants from non-recombinants. Neither the vector nor the host cell can produce a functional β-galactosidase. The host chromosome has a deletion of the lac promoter, which is found instead on the male episome. The episome is also unable to produce a functional β-galactosidase, due to a deletion of the sequence for amino acids 11-41. When the host cell is transformed by an M13 vector, only then is a functional β-galactosidase molecule produced (in the presence of the lac inducer IPTG, isopropyl-beta-D-thio-galactopyranoside) by means of complementation of the two non-functional enzymes (intra cistronic complementation). Such cells will hydrolyse the substrate X-Gal (5-bromo-4-chloro-3-indoyl-β-galactoside) to give a blue dye (bromochloroindole) and this results in blue coloured plaques. In general, if the M13 vector contains a foreign DNA insert within the β-galactosidase coding region (recombinant molecule), no functional β-galactosidase activity is produced and the resultant plaques are white.

Several M13 cloning vectors have been constructed by Messing and co-workers. These vectors have multiple cloning sites inserted into the original EcoRI site of M13 mp 2. The vectors used in this study are shown in Figure 2.3a. The

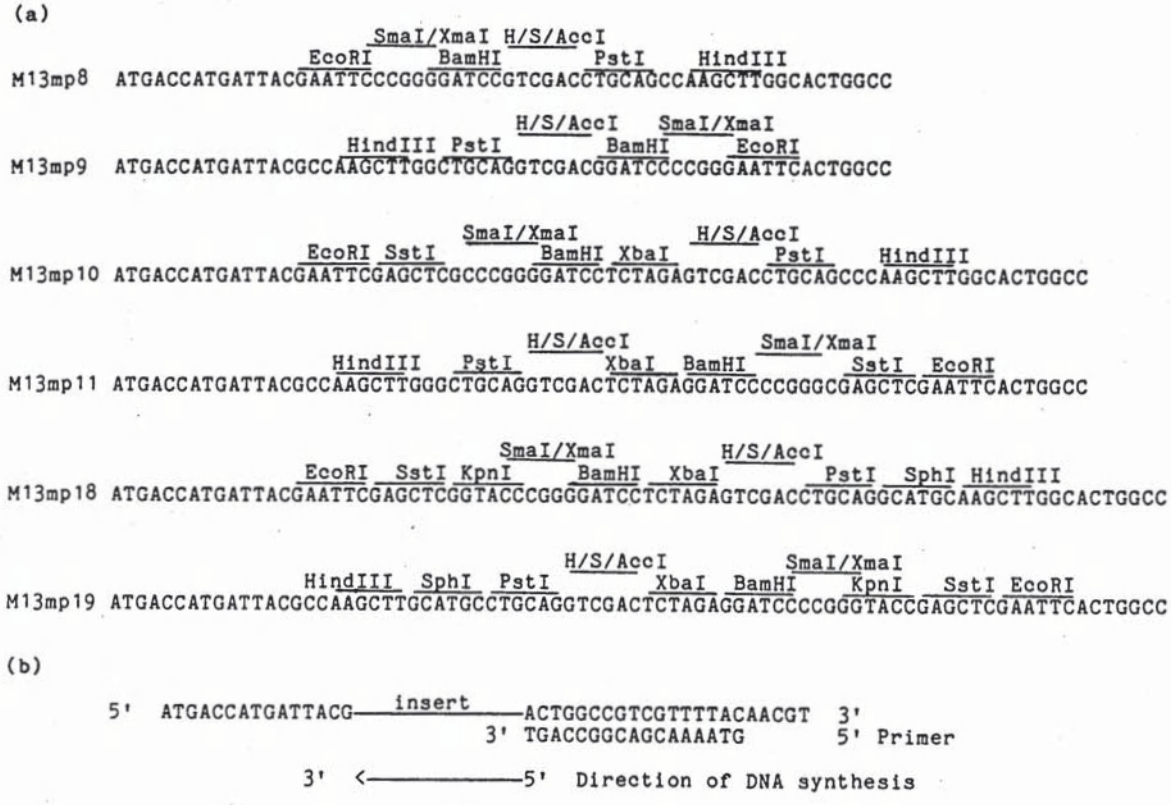


Figure 2.3

(a) M13 cloning vectors used in this study, showing the multiple restriction sites.

H - Hinc II

S - Sal I

(b) 17-mer primer used in this study, and its binding site.

multiple restriction sites can be used individually or in pairs to clone any desired region of the DNA of interest. These vectors comprise pairs, with each pair containing mirror image cloning sites. The advantage of this is that a DNA fragment bound by two different restriction sites can be cloned in opposite orientations. This allows sequencing of both strands from opposite ends of the insert.

2.3.d. Universal primer

Since all fragments are cloned into the same specific region of the M13 genome, it is possible to use a universal primer complementary to the region of M13 flanking the multiple cloning sites. The Klenow fragment will extend the primer 5' ——> 3' passing immediately through the cloning site, and so will give the sequence of any DNA fragment inserted at this position. Such short sequences have been made and are now commercially available (Anderson et al., 1980; Heidecker et al., 1980; Duckworth et al., 1981; Messing et al., 1981). The primer used in this work is shown Figure 2.3.b.

Therefore, the development of M13 cloning vectors and the requirement for only one primer sequence has made the dideoxy sequencing method of Sanger widely applicable (Sanger et al., 1980; Messing et al., 1981). A summary of M13 cloning and dideoxy sequencing is shown in Figure 2.4.

2.3.e. Use of ^{35}S labelled nucleotides in dideoxy sequencing

The use of phosphorous-32 (^{32}P) nucleotides in dideoxy sequencing is well established. However, recently, nucleotides labelled with sulphur-35 (^{35}S) have been produced. Sulphur replaces a non-bridging oxygen in the α-phosphate group of a nucleoside triphosphate. These nucleotides (termed thionucleotides) can be substituted for ^{32}P labelled nucleotides with very little effect on methodology and with several advantages (Biggin et. al., 1983).

Before discussing their use in DNA sequencing, I would like to point out an

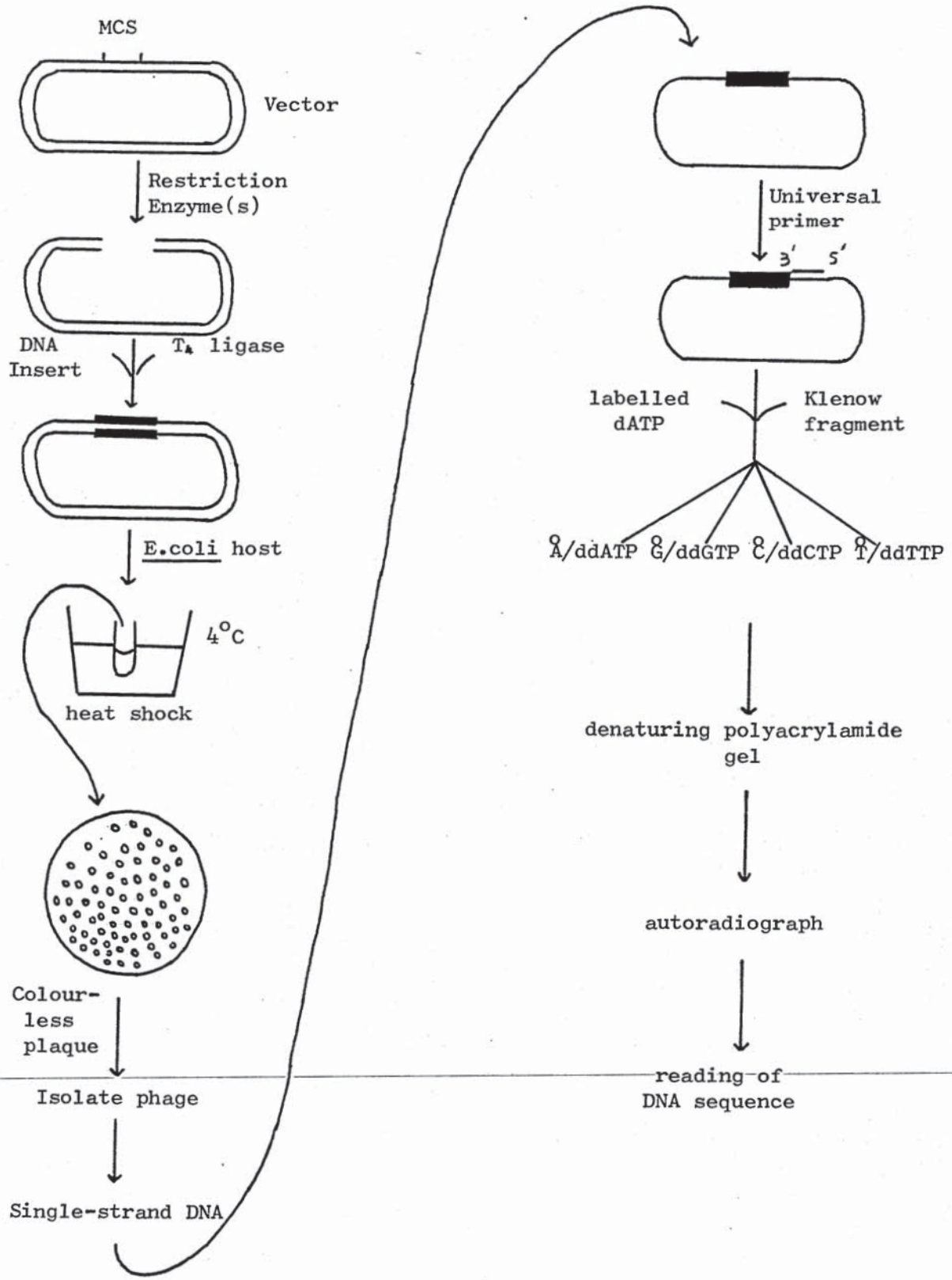


Figure 2.4 M13 Cloning as an Aid to Dideoxy Sequencing

MCS - multiple cloning sites

interesting property of these thionucleotides. The replacement of an oxygen by sulphur introduces chirality into the molecule, with the possibility of two stereoisomers, the R and S forms (Burgers and Eckstein, 1978). These two forms are shown in Figure 2.5. However, it has been shown that synthesis of these thionucleotides by an enzymatic mechanism, results in only the S configuration. The ability to work with a pure isomer has highlighted the stereo-specificity of some interesting enzymes. E. coli DNA and RNA polymerases and T4 DNA ligase all show absolute specificity for the S form. (Eckstein et al., 1982; Putney et al., 1981). In contrast to this, oligonucleotides containing the S form thionucleotides are resistant to several nucleases, including exonuclease III, E. coli and T4 3'——>5' exonuclease activity of DNA polymerase, and snake venom phosphodiesterase (Putney et al., 1981; Kunkel et al., 1981; Burgers and Eckstein, 1979).

Returning to the use of thionucleotides in dideoxy sequencing, thionucleotides are not natural substrates for DNA polymerase 1. The rate of incorporation is slightly lower than for natural substrates and so the sequencing reaction time has to be increased accordingly. Otherwise, no changes need to be made to established protocols. ^{35}S has a half life six times as long as that for ^{32}P and thus allows extra flexibility in the planning of experiments and reduces unnecessary wastage of unused label. The most important advantage of ^{35}S is that it is a soft β emitter. There is less radiation dose, which is especially important for workers processing many clones per day over several days at a time. Secondly, resolution in autoradiography is improved, since the shorter range of β particles results in much sharper bands. Thus sequences can be read much further up into the gel. Since ^{35}S has a very short penetration distance, gels have to be dried before autoradiography, a process which in itself gives increased resolution. Therefore ^{35}S labelled nucleotides are now used for dideoxy sequencing in preference to nucleotides labelled with ^{32}P.

2.4. Use of Both Sequencing Methods in this Study

As outlined earlier, the first objective was to determine the sequence of 18S

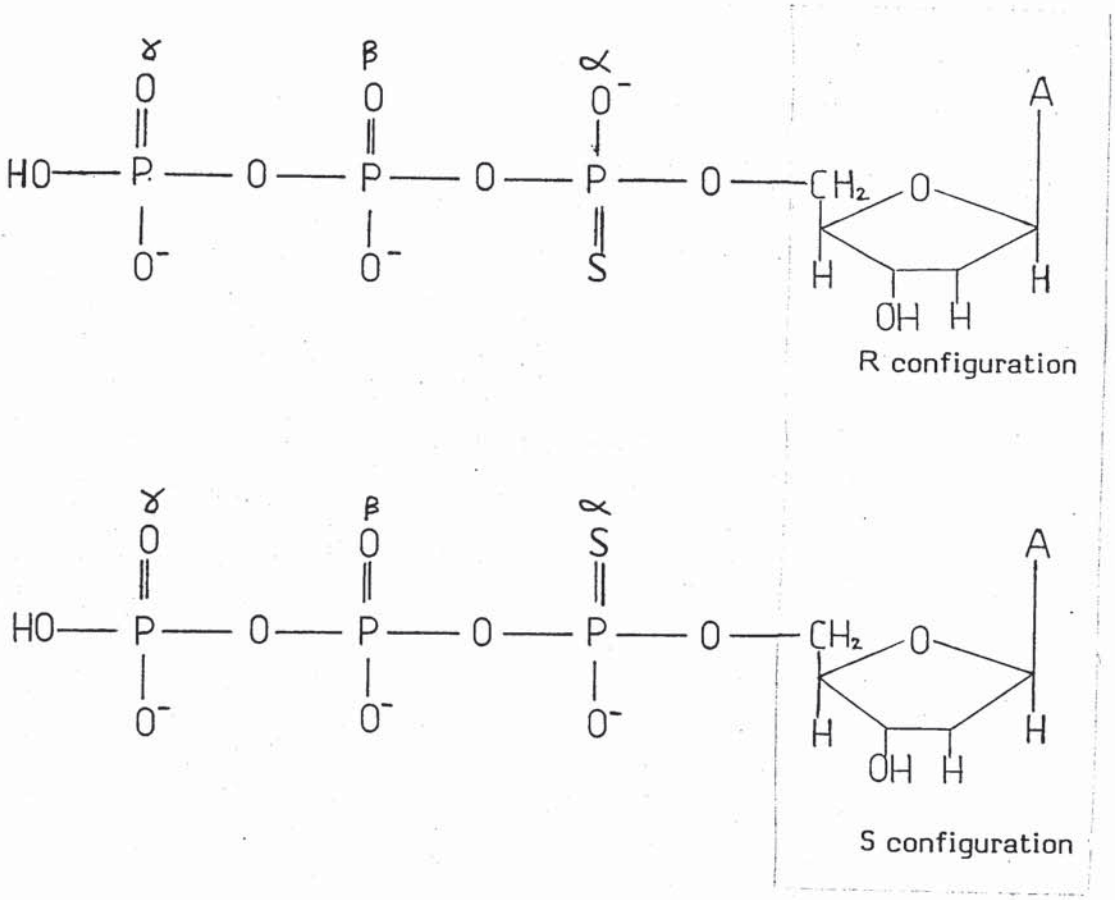


Figure 2.5 Thionucleotides : R and S Forms

The above figure shows the two stereoisomers of dATPαS. The introduction of a sulphur gives a chiral centre at the α phosphate group and results in the possibility of two forms, namely R and S.

rDNA from X. borealis. The work of Furlong and Maden (1983) had shown only one difference near the 3' terminus on comparison with the published sequence for X. laevis (Salim and Maden, 1981). The expectation was that there would probably be relatively few differences from X. laevis 18S rDNA, and so I used features of the X. laevis sequence to establish a sequencing strategy. This strategy utilised many restriction sites in conjunction with the Maxam-Gilbert chemical sequencing method, which was already well established in the laboratory.

It was during the comparative analysis of the differences found between the two Xenopus species that M13 cloning and dideoxy sequencing were introduced. It became apparent that this system was an ideal choice for comparing the differences throughtout a range of X. borealis and X. laevis clones as quickly as possible. Since I was looking at the same point of sequence in all clones, identical fragments were cut out by restriction enzyme digestion. Therefore only one digest of M13 was required. Subsequent ligation and transformation into JM103 gave the required plaque formation. The necessary single-strand template for dideoxy sequencing only takes two days to prepare, with only one more day required to carry out both the sequencing reactions and run the gels. Therefore, actual sequence data accumulate very rapidly.

Initial experiments with the dideoxy sequencing method were carried out with ^{32}P dATP, before subsequently turning to ^{35}S dATP. Comparison of these early gels with the later ones shows clearly the much higher degree of resolution obtained on using ^{35}S labelled nucleotides, discussed in section 2.3.e. (Results, Chapter 4). In setting up this sequencing method, I used individual reagents purchased from various companies and followed protocols from A. Bankier, MRC, Cambridge. However, as the popularity of this system has increased, Amersham now market a complete sequencing kit and I decided to take advantage of this.

Analysis of the human 18S gene region was carried out completely by M13 cloning and dideoxy sequencing, establishing this as a routine method of sequencing in our laboratory.

CHAPTER 3

MATERIALS AND METHODS

Common chemicals were analar grade supplied by BDH Chemicals, Poole, Dorset, or Fisons Scientific Apparatus, Loughborough, Leics. Exceptions to this are noted in the text.

3.1. Media and Antibiotics

3.1.a. Liquid media

All media were sterilised by autoclaving before use.

L-Broth:
- 1% Tryptone (Difco, West Molesey, Surrey).
- 0.5% Yeast Extract (Difco)
- 0.5% NaCl
- 10mM Tris-HCl, pH 7.4
- 1mM $MgSO_4$

Minimal Medium:
- 1.05% K_2HPO_4
- 0.45% KH_2PO_4
- 0.1% $(NH_4)_2\ SO_4$
- 0.05% Sodium citrate, $2H_2O$
- 0.02% $MgSO_4.7H_2O$
- *5µg/l thiamine-HCl (Vit B1, Sigma, Poole, Dorset)
- *0.2% glucose

* Sterilised separately as a concentrated solution.

YT Medium: 0.8% Tryptone (Difco)

0.5% Yeast Extract (Difco)

0.5% NaCl

2 x YT is double the concentration of this recipe.

3.1.b. Medium containing agar

Make up liquid according to the appropriate formula given in 3.1.a. Just before autoclaving, add Bacto Agar (Difco) to the required concentration

for plates, 15g/l

for top agar, 6g/l

3.1.c. Antibiotics

These were obtained from Sigma and used at the following concentrations.

ampicillin, 100µg/ml

tetracycline, 15µg/ml

These can be added in solid form or from a concentrated stock solution. Solutions are sterilised by filtration and stored in aliquots at -20°C. Tetracycline solutions should also be wrapped in aluminium foil.

3.2 Maintenance of Bacteria and Plasmids

3.2.a. E. coli hosts

E. coli HB101:

$K12$, F^-, pro^-, leu^-, thi^-, $lacY^-$, $hsdR^-$, end A^-, rec A^-, rsp L20, ara-14, gal-K2, xyl-5, mtl-1, SupE 44.

E. coli JM103:

Δlacpro, thi, strA, supE, end A, sbcB15, hsd R4, F'tra D36, pro AB, lac $1^qZ\Delta M15$.

E. coli HB101 was the host for the plasmids used during the course of this project (Bolivar and Backman, 1979). E. coli JM103 was host for the growth of single-stranded bacteriophage M13 and M13 recombinants (Messing et al., 1981).

3.2.b. Long term storage of bacteria

Long term storage of bacteria is achieved by storing in 40% glycerol at -20°C.

To 2.5mls of an exponentially growing culture, add 2.5mls of sterile 80% glycerol. Vortex to ensure good mixing and store at -20°C.

3.2.c. Stock Plates of JM103

Stock JM103 can be streaked onto a glucose minimal agar plate (3.1.b.) using a sterile platinum loop. The plates are then allowed to grow overnight in a 37°C incubator. The plates can be used as a stock of single colonies for at least a month.

3.2.d. Storage of plasmids

Plasmids can be stored in cultures of transformed HB101 as described in 3.2.b.

Native plasmid DNA is stored in TE buffer (10mM Tris-HCl pH 8.0, 0.1mM EDTA). This is stored in a screw-cap tube at 4°C over a drop of chloroform. This should remain pure for several years with regular replenishment of chloroform.

It is also advisable as a safety precaution to store some native DNA at -20°C. This can be used to retransform bacterial host cells should the need arise.

3.2.e. A check of antibiotic resistance

Especially when preparing a new plasmid for the first time, it is advisable to check that transformed bacteria demonstrate the expected antibiotic sensitivity or

resistance. Streak some liquid culture onto the appropriate antibiotic plates (3.1.b.). Incubate plates overnight at 37°C. Antibiotic characteristics of the plasmids used in the course of this project are shown in Table 2.1.

3.3. Large Scale Preparation of Plasmids

Two major differences between E. coli DNA and plasmid DNA have been exploited in the determination of a method to isolate pure plasmid DNA. The E. coli chromosome is much larger than the DNA of commonly used plasmids. The bulk of E. coli DNA extracted from cells is obtained as broken, linear molecules, whereas plasmid DNA is generally extracted in a covalently closed, circular form. All methods devised are based on three basic steps. Growth of bacteria and amplification of the plasmid, harvesting and lysis of the bacteria and lastly purification of the plasmid DNA.

The method shown here is derived from that of Clewell and Helinski (1970).

3.3.a. Solutions

Sucrose solution: 25% sucrose, 0.05M Tris-HCl, pH 8.0

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

The lysozyme solution must be prepared fresh on the day of cleared lysis.

Brij solution: 1% Brij 58 (polyoxyethylene 20, cetyl ether, Sigma)

0.4% sodium deoxycholate

2.5mM EDTA

50mM Tris-HCl, pH 8.0

3.3.b. Growth of bacteria and amplification of the plasmid

The volumes given below are for a 1 litre culture but can be adapted to suit other volumes. A 1 litre culture should yield up to 1mg of plasmid DNA.

Starter cultures are usually inoculated from a 40% glycerol stock. Transfer

0.1ml to 50ml of L-broth (3.1.a.) supplemented with the appropriate antibiotic (as required for plasmid selection) in a 250ml conical flask. Shake at 37°C overnight.

The main culture is best set up around mid-day. Read the O.D.660 of the overnight culture (dilute 1:9 in L-broth). Inoculate 2 x 500ml of L-broth (in 2l flasks), again with the appropriate antibiotic added, with enough cells to give O.D.660 of 0.04. Shake at 37°C until the O.D.660 = 0.5.

Add 100mg of chloramphenicol dry powder per 500ml and incubate at 37°C with shaking for a further 16 hours. During this time, chromosomal DNA synthesis is inhibited while plasmid DNA synthesis continues.

3.3.c. Harvesting and lysis of bacteria

Harvesting:

Harvest the bacterial cells by centrifugation at 5000 rpm for 10 min at 4°C. Discard the supernatant. Resuspend in a total volume of 50ml 10mM Tris-HCl, pH 8.0, 1mM EDTA. This requires vortexing. Centrifuge for a second time and pour off the supernatant, keeping the pellet on ice.

Lysis:

Resuspend the pellet in 8ml of an ice-cold sucrose solution, add 1.8ml of lysozyme solution and swirl on ice for 5 mins. Add 3.3mls of 0.25M EDTA, pH 8.0, and swirl on ice for 5 mins. Add 15ml of Brij solution and swirl on ice for a further 10mins. Centrifuge at 30,000 rpm for 45 mins at 4°C in a Beckman 60Ti rotor. This removes high-molecular weight DNA and bacterial debris. Plasmid DNA is contained in the supernatant. Pour off into a graduated cylinder and note the volume.

3.3.d. Purification of closed circular DNA

Plasmid DNA behaves differently from *E. coli* DNA when the two are centrifuged to equilibrium in caesium chloride (CsCl) gradients containing

saturating concentrations of the intercalating dye, ethidium bromide. Covalently closed, circular DNA binds less ethidium bromide than linear DNA and therefore bands at a higher density in CsCl gradients.

For every 1ml of supernatant add exactly 1g of solid CsCl and mix gently to dissolve all the salt. Add 0.8ml of ethidium bromide (10mg/ml) for every 10ml of CsCl solution. Mix well. The final density of the solution should be 1.55mg/ml. This can be checked by weighing 1ml of the solution.

Transfer the solution to a quick seal polyallomer tube. If required, fill the remainder of the tube with light paraffin oil. Seal tubes. Place in a Beckman Vti50 rotor for 16 hours at 45,000 rpm at 20°C.

View the tube under U.V. illumination. Two bands of DNA can be seen. The upper band consists of linear bacterial DNA and nicked circular plasmid DNA. The lower band consists of closed circular plasmid DNA. An RNA pellet can be seen at the bottom of the tube.

Using a 21 guage needle, puncture the tube and collect the lower band into a glass tube. Alternatively the band can be collected from above using a pasteur pipette. It is first necessary to remove material above the desired band and then to remove the covalently closed, circular DNA using a clean pasteur pipette.

3.3.e. Removal of ethidium bromide

Add an equal volume of isopropanol solution (isopropanol, 0.05M Tris-HCl, pH 8.0, 0.05M NaCl, 0.01M EDTA, pH 8.0, saturated with CsCl). Mix the two phases by inversion. Let the phases separate well (5-10 minutes) to avoid DNA loss. Discard the upper phase. Repeat extraction until the pink colour has completely gone from the aqueous phase and then repeat one more time (usually a total of 4 times).

Add 2 volumes of water and 2.5 volumes of ethanol (calculated on diluted aqueous volume). Chill at -20°C for a few hours to precipitate the DNA. The presence of CsCl enables the DNA to precipitate.

Centrifuge at 10,000 rpm for 10 mins at -10°C to pellet the DNA. Pour off the supernatant. Redissolve the pellet in 3ml of 0.3M sodium acetate. Add an equal volume of phenol saturated with TE buffer, pH 8.0. Vortex for 3 minutes. Separate the phases by low speed centrifugation. Recover the aqueous phase (top phase). Add 2.5 volumes of cold ethanol. Chill at -20°C for 2 hours followed by -70°C for 30 mins. Centrifuge at 10,000 rpm for 10 mins at -10°C. Wash the DNA pellet with 80% ethanol. Re-centrifuge. Dry off the remaining ethanol under vacuum. Dissolve the pellet in 1ml of TE buffer, pH 8.0. Read the O.D.260nm of 10µl of the sample (dilute to 1ml in TE). An O.D.260 of 1.0 is equivalent to a DNA concentration of 50µg/ml. Calculate the total yield of DNA, remembering to take account of the 1/100 dilution. Run 1µl of the sample on a 1% agarose mini-gel (3.5.b.) to check the quality of the preparation.

3.4. Digestion with Specific Restriction Endonucleases

Restriction Enzymes were obtained from Bethesda Research Laboratories (U.K.) Ltd., Cambridge, The Boehringer Corporation (London) Ltd., Lewes, E. Sussex, New England Biolabs., C.P. Labs., Ltd., (U.K. dist.) Bishops Stortford, Herts., and P and S Biochemicals, Liverpool.

3.4.a. Reaction buffers

Each restriction enzyme has optimal reaction conditions as specified by the manufacturers. However, the buffers shown below (determined previously in our laboratory) allow the enzymes to be classified into three groups with respect to preference for the following buffers.

No salt Buffer (1x) 10mM Tris-HCl
6mM $MgCl_2$
0.5mM dithiothreitol
pH 7.6

Low salt Buffer (1x) 10mM Tris-HCl

6mM $MgCl_2$

6mM NaCl

0.5mM dithiothreitol

pH 7.5

High salt Buffer (1x) 10mM Tris-HCl

6mM $MgCl_2$

60mM NaCl

0.5mM dithiothreitol

pH 7.5

Buffers are usually prepared as a 10x concentrated stock and stored at 4°C.

By comparing the above recipes with the manufacturers recommendations it is possible to choose the appropriate buffer for most enzymes.

3.4.b. Setting up restriction digests

Digests are carried out in a 1.5ml Eppendorf Tube in the presence of 1/10 x volume of the appropriate 10x reaction buffer (3.4.a.).

1 unit of enzyme activity is determined as the amount of enzyme required to digest 1µg of DNA to completion in 1 hour. However to ensure complete digestion it is usual to add a several fold excess of enzyme. For a small scale analytical digest (0.2 - 1µg of DNA) add a 5x excess of enzyme per digest. Where a large scale preparative digest is required, to save on the amount of enzyme used, add 1/5 the amount of enzyme and digest to completion overnight. Make sure the contents of the Eppendorf tube are well mixed before incubating at the appropriate temperature.

When DNA is to be cleaved by two or more restriction enzymes, the digestions can be carried out simultaneously if both enzymes work in the same buffer. If this is not the case, the enzyme that works in the buffer of lower ionic strength should be used first. The appropriate amount of salt and the second enzyme can then be added and the incubation continued.

3.5. Separation of DNA Fragments by Gel Electrophoresis

Electrophoresis through gels is the standard procedure used to separate, identify, and purify DNA fragments. Two separate gel types were used here, namely 1% agarose gels and 4% polyacrylamide gels. The choice of gel type depends on the size of the DNA fragments to be separated. In general 4% polyacrylamide gels are used to separate fragments of less than 1000 base pairs, with anything larger than this requiring a 1% agarose gel.

3.5.a. Preparation of 1% agarose gels

Running Buffers

Agarose Electrophoresis Buffer (1x)

40mM Tris-HCl

20mM Sodium acetate

2mM EDTA

pH to 8.0 with conc. acetic acid

TBE Electrophoresis Buffer (1x)

100mM Tris-HCl

100mM Boric Acid

1mM EDTA

pH 8.3

pH should be correct, do not adjust.

1% Agarose Solution

Prepare a 1% (w/v) solution of Agarose (Type II, Sigma) in 1x running buffer.

The mixture is dissolved by heating in a boiling water bath. This solution can be stored molten in a screw cap bottle at 65°C. However, fresh stock should be prepared every few weeks, otherwise resulting gels tend to be brittle and thus difficult to handle.

Agarose Loading Buffer

1/5 Electrophoresis buffer (as above), 2M sucrose and a few grains of bromophenol blue.

Minigel: BRL Model H6

A mini-gel is used as an analytical gel rather than a preparative gel. It is used to check the quality of a DNA preparation or to check that a restriction enzyme digestion has gone to completion. This is especially useful before committing a large-scale digest to a preparative gel.

Seal both ends of the gel platform with waterproof tape (Universal Scientific Ltd., London, England). Pour 12.5ml of 1% agarose solution to cover the platform and position the comb. Allow the gel to set before covering with 1x running buffer. Remove the comb.

Medium Gel: BRL Model H5

This can be used as an analytical gel or as a preparative gel for up to 100µg of DNA. Prepare as for a mini-gel, but use 35ml of 1% agarose solution.

Vertical Gel: BRL Model V16

This type of gel is slightly more difficult to assemble and has tended during the course of this project to be replaced by the medium-gel (above). However, it does give good resolution of fairly large amounts of DNA.

Assemble the apparatus with either 1.5mm or 3mm spacers. Tip back the gel apparatus to about 45° and pour approximately 50ml of 1% agarose solution into the bottom trough. Allow this to set, thus forming a plug. Return the gel system to a vertical position and fill up with gel solution. Insert the comb and allow the gel to set. Fill both top and bottom troughs with 1x running buffer. Remove comb.

Large Slab Gel: BRL Model HO

Large slab gels are used for preparations of 100µg or more of DNA.

First prepare 2% agarose in 1x agarose running buffer to form wicks. Once these have set, place the gel tray in the apparatus and pour in 1% Agarose solution to a depth of about 3mm (200-250ml). Position the comb and allow to set. Again fill both troughs with 1x running buffer.

3.5.b. Running of 1% agarose gels

Mix sample with ½x volume of loading buffer (3.5.b) and apply to the sample well. Electrophoresis is carried out at 100mA until the bromophenol blue (500-800bp) has travelled the desired distance, dependent upon the size of the specific sample being electrophoresed.

Once electrophoresis is complete, remove the gel and stain for 15 minutes in a solution of ethidium bromide (10µg/ml). Place the gel on a U.V. illuminator (Models C-62, TM-36, U.V. Products, Winchester, Hants.) and photograph (Polaroid Cu-5 land camera and 655 positive/negative film).

3.5.c. Recovery of DNA from agarose gels

Various methods of recovering DNA from agarose gels have been tried during the course of this project. However the method shown here has proved to be the most reliable. This method (originally described by McDonnell et al. (1977)) involves electroelution of the DNA band in a dialysis bag.

Preparation of Dialysis Tubing

Always wear gloves when handling dialysis tubing. Cut the tubing into pieces of convenient length. Boil for 10 minutes in a large volume of 2% sodium bicarbonate and 1mM EDTA. Rinse the tubing thoroughly in distilled water. Boil for 10 minutes in distilled water. Allow to cool and store at 4°C.

Electroelution

Having localised the band of interest by U.V. illumination, using a sharp scalpel, cut out the slice of agarose containing the band. Wash a piece of dialysis tubing (see above for preparation) inside and out with distilled water. Seal one end with a suitable clip. Fill the bag to overflowing with 0.5x running buffer (3.5.a.). Holding the neck of the bag, pick up the gel slice with clean forceps and place in the fluid-filled bag. Allow the gel slice to sink to the bottom of the bag. Remove most of the buffer, leaving just enough to keep the gel slice in constant contact with the electrophoresis buffer. Seal the bag with another clip making sure not to trap any air bubbles. Immerse the bag in a shallow layer of 0.5x running buffer in a horizontal gel tank. Pass electric current through the bag for 2-3 hours (100V). Reverse the polarity of the current for 2 minutes, which is just enough time to release the DNA from the inner wall of the dialysis tubing, but not long enough to let it run back into the gel slice.

Remove one clip and recover all the buffer, using a pasteur pipette, into a siliconised 15ml corex tube. Wash out the bag with a small amount of 0.5x electrophoresis buffer. It is a good idea at this point to re-stain the gel piece in ethidium bromide to check that all the DNA has indeed been eluted, then proceed as follows. Measure the elution volume. Add 1/10 volume of 3M sodium acetate pH 5.6 and 3x (total aqueous volume) ice-cold ethanol. Mix thoroughly and place at -20°C overnight. Put at -70°C in dry ice/methylated spirits for 30-60 mins and pellet the DNA by centrifugation at 10,000rpm for 10 mins at -10°C. Pour off the supernatant and dry the pellet under vacuum for 10-15 mins. Redissolve in 0.4ml of 0.3M sodium acetate and transfer to a 1.5ml polypropylene Eppendorf tube. Centrifuge (9950g) for 5 mins to pellet any remaining traces of agarose. Remove the supernatant into a clean eppendorf tube. To this, add 1ml of ice-cold ethanol and precipitate at -70°C for 15-30 mins. Centrifuge (9950g) for 5 mins and discard the supernatant. Re-precipitate the DNA pellet with 100µl of 0.3M sodium acetate and 300µl of ice-cold ethanol. Centrifuge as before. Wash

DNA pellet in 300µl of ice-cold 80% ethanol, chill at -70°C for 5 mins and re-centrifuge. Discard the supernatant and dry the pellet under vacuum. Take up in a suitable volume of TE buffer, pH 8.0 and store at 4°C or -20°C.

3.5.d. Preparation of 4% acrylamide gel

Electrophoresis Buffer

Gels are run in 1x TBE buffer (3.5.a.)

Acrylamide Stock

19% Acrylamide

1% N, N-methylene bisacrylamide

Deionise by stirring over Amberlite MB1 mixed bed resin (5g/100ml of solution) for 30 minutes. Filter to remove the resin and store at 4°C.

4% Acrylamide Gel Mix

10ml acrylamide stock

5ml 10x TBE

5ml glycerol

30ml H_2O

0.4ml 10% ammonium persulphate, freshly prepared.

Acrylamide Loading Buffer

50% glycerol in 1x TBE

0.05% Xylene cyanol

0.05% bromophenol blue

Vertical Gel: BRL Model V16

Assemble the gel apparatus with 1.5mm spacers, including one across the

bottom of the gel plates. To 5ml of gel mix (see above) add 40µl of TEMED and using a pasteur pipette, let the mix run down the side spacers, forming a seal down both sides and across the bottom. Once this has set (5 mins), add 30µl of TEMED to the remaining mix and pour the main gel. Insert comb and allow to set. Remove the bottom spacer and clamp the plates to the gel tank. Fill both troughs with 1x electrophoresis buffer and remove the comb.

3.5.e. Running of 4% acrylamide gel

Pre-run the gel for at least 30 minutes at 200V. Mix sample with loading dye (3µl of dye per 10µl of sample) and apply to gel. Carry out electrophoresis at 200V until the dyes have travelled the required distances (Bromophenol blue travels at 70bp with the slower moving dye xylene cyanol travelling equivalent to 350bp). Once electrophoresis is complete stain and view as for agarose gels (3.5.b.)

3.5.f. Recovery of DNA from acrylamide gels

Acrylamide Elution buffer

0.5M	Ammonium Acetate
0.01M	Magnesium Acetate
0.1%	SDS
1mM	EDTA

Having stained and photographed the gel, cut out the desired band and add to a large blue eppendorf tip (previously sealed at the narrow end and plugged with siliconised glass wool). Grind the gel piece with a clean glass rod. Add 0.6mls of acrylamide elution buffer (see above) and seal top of tip with Nescofilm. Place in a 15ml siliconised Corex tube, seal with Nescofilm and stand at 37°C overnight. Cut off the end of the tip and allow the elution buffer to drain into the corex tube. Rinse tip with 0.2ml of elution buffer. Repeat rinse a further 3 times. To pooled eluate (1.4ml) add 2.5 x volume of ice-cold ethanol (3.5ml). Mix well and stand for 30-60 mins at -70°C in dry ice/methylated spirits. Centrifuge at 10,000 rpm for 10

mins at -10°C to pellet the DNA. Pour off the supernatant and dry pellet under vacuum. Redissolve in 0.4ml of 0.3M sodium acetate and transfer to a clean eppendorf tube. Centrifuge for 5 mins (9950g) to pellet any acrylamide debris. Transfer supernatant to a new eppendorf tube, add 1ml of ice-cold ethanol and stand at -70°C for 15-30 mins. Continue as for agarose gels (3.5.c.).

3.6. Preparation of Fragments for Maxam-Gilbert Sequencing

3.6.a. Phosphatase treatment

The terminal 5' phosphates of DNA are removed by treatment with bacterial alkaline phosphatase (BAP) (Worthington, Flow Labs. Ltd., Irvine, Scotland), further purified according to Hall (1981).

Dissolve the DNA fragment in 90µl of TE buffer, pH 8.0. Add 10µl of BAP. Mix well and incubate for 75 mins at 37°C. Add 100µl of TE-saturated phenol, vortex for 3 mins. Separate the phases by centrifugation for 5 mins (9950g). Transfer the aqueous phase to a clean eppendorf tube. Re-extract the phenol phase with 100µl of TE. Pool the two aqueous phases. Add 20µl of 3M sodium acetate (pH 5.6) and 600µl of ice-cold ethanol. Mix well and precipitate at -70°C in dry ice/methylated spirits for 15 mins. Centrifuge to pellet the DNA. Repeat precipitation with 100µl of 0.3M sodium acetate and 300µl of ethanol. Finally, wash with ice-cold 80% ethanol. Centrifuge to pellet the DNA, remove supernatant and dry pellet under vacuum.

3.6.b. Polynucleotide kinase labelling

This procedure is for labelling 5' protruding ends.

Kinase buffer (1x)

0.05M	Tris-HCl
0.01M	$MgCl_2$
pH 8.0	

Make up a 10x concentrated stock and store at 4°C.

Dissolve the DNA to be labelled in 2µl of H_2O. Add 1µl of 10x kinase buffer (as above), 1µl of 50mM dithiothreitol, 6µl (60µCi) of γ-^{32}PATP (5000 Ci/mmole, Amersham International plc., Amersham, Bucks.) and 1µl (5U) of T4 polynucleotide kinase (PL Biochemicals Inc., Northampton). (If the DNA to be labelled has flush ends, before the addition of enzyme, add 1µl of 10mM spermidine and heat to 90°C for 2 mins). Mix, spin and incubate at 37°C for 30 mins.

Add 40µl of 2.5M ammonium acetate, 160µl of ice-cold ethanol and precipitate at -70°C for 15 mins. Centrifuge (9,950g) and remove supernatant (to radioactive sink). Re-precipitate with 100µl of 0.3M sodium acetate, 300µl ice-cold ethanol. Centrifuge again, remove supernatant and wash with 300µl of 80% ice-cold ethanol. Chill to -70°C for 5 mins. Spin once more, remove supernatant and dry pellet under vacuum.

3.6.c. Secondary digestion and separation of labelled ends.

Chemical sequencing can only be performed on a fragment of DNA labelled at one end, therefore the DNA fragment labelled at both ends (3.6.b.) has to be cleaved.

Choose a restriction enzyme that will cut the DNA fragment asymmetrically. Separate the fragments by electrophoresis on a 4% acrylamide gel (3.5.d.e.). Stain and elute as previously described (3.5.f.). If the amount of DNA is too small to be seen by ethidium bromide staining, then the gel can be autoradiographed for 15 mins (Kodak NS-2T no screen film, Kodak Ltd., Kirby, Liverpool).

Cerenkov count the dry pellet in a scintillation counter (using the tritium channel) to obtain a measure of the radioactivity of the sample.

3.7. Sequencing DNA by the Method of Maxam-Gilbert

This is the chemical method of sequencing. A well documented description is

given in Maxam and Gilbert (1980).

3.7.a. Reagents and solutions

Dimethylsulphate (Gold Label, Aldrich Chemicals, Gillingham, Dorset)

Hydrazine (Kodak Ltd.)

Piperidine (Koch-Light Labs. Ltd., Colnbrook, Bucks.)

Dimethylsulphate buffer (DMS)

50mM	sodium cacodylate
10mM	$MgCl_2$
0.1mM	EDTA
pH 8.0	

Pyridinium Formate Solution

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

DMS Stop Solution

1.5M	Sodium acetate
1M	β-mercaptoethanol (Koch-Light)
100µg/ml	Yeast tRNA

Hydrazine Stop Solution

0.3M	Sodium acetate
0.1mM	EDTA
50µg/ml	Yeast tRNA

3.7.b. Modification reactions and strand scission

Dissolve the dried DNA pellet in H_2O to give at least 100,000 Cerenkov counts in 11µl. Add 4µl of carrier DNA (1mg/ml sonicated calf thymus DNA). Mix

G	G + A	C + T	C
DNA+carrier 3.5µl	DNA+carrier 3.5µl	DNA+carrier 3.5µl	DNA+carrier 3.5µl
DMS buffer 98µl	H_2O 11µl	H_2O 6µl	Sat. NaCl 8µl
DMS 0.5µl	Pyr. formate 2.5µl	Hydrazine 15µl	Hydrazine 15µl
20°C, 5 mins	30°C, 70 mins	20°C, 8 mins	20°C, 10 mins
DMS stop 24µl	freeze, -70°C	Hydrazine stop 60µl	Hydrazine stop 60µl
AR ethanol 400µl	lyophilise, add 10µl H_2O	AR ethanol 250µl	AR ethanol 250µl
-70°C, 15 mins	freeze, lyophilise	-70°C, 15 mins	-70°C, 15 mins

Table 3.1 Volumes and reaction times for base modification reactions

DMS - dimethyl sulphate

Pyr. formate - pyridinium formate

AR ethanol - analar ethanol (James Burrough Ltd., Fine Alcohol Division, London).

thoroughly and aliquot 3.5 µl into each of 4 siliconised 1.5ml Eppendorf tubes. I carried out four reactions, specific for guanine (G), guanine and adenine (G + A), cytosine and thymine (C + T), and cytosine (C). The reactions were carried out as shown in Table 3.1. The reaction conditions can be varied, but those shown in Table 3.1 are optimal for most purposes.

For G, C + T, and C, after precipitating at -70°C for 15 mins, centrifuge (9950g) for 5 mins. Discard the supernatant (to the radioactive sink). Add 60µl of 0.3M sodium acetate and 200µl of analar ethanol and put at -70°C for 15 mins. Recentrifuge to pellet the DNA. Wash in 200µl of 70% ethanol. Chill at -70°C for 5 mins, centrifuge (9950g), discard the supernatant and dry sample under vacuum.

To all four tubes, add 100µl of 1M piperidine (1:9 dilution of concentrated piperidine in H_2O). Heat to 90°C for 30 mins. Centrifuge (9950g) for 30 secs, freeze in dry ice and lyophilise. Take up in 20µl of H_2O, freeze and lyophilise. Repeat water wash once more. Cerenkov count the sample as before (3.6.c.).

3.7.c. DNA sequencing gels

DNA fragments differing in length by only 1 nucleotide were separated by electrophoresis on 6% polyacrylamide gels (40cm x 20cm x 0.4mm thick) (Sanger and Coulson, 1978).

Gel mix (for 2 gels)

30ml of 20% acrylamide stock (3.5.d.)

10ml of 10x TBE (3.5.a.)

27ml H_2O

42g of Urea

0.75ml of 10% ammonium persulphate

filter and de-gas

Loading buffer: formamide plus a few grains of xylene cyanol. Note : put formamide over a few grains of Dowex resin and take the supernatant.

Having prepared the gel plates, make up the gel mix (above), add 40µl of

TEMED and pour the gels. Insert comb (14 x 7mm wells) and allow gels to set.

Pre-run the gels for two hours in 1x TBE buffer (3.5.a.), setting the voltage to 2000 V for 43cm gel plates and current to 25mA per gel (LKB 2103 power supply, LKB Instruments Ltd., LKB House, South Croydon, Surrey, England).

Ideally, dissolve the samples in the appropriate volume of loading buffer to give 10,000 Cerenkov counts (cpm) per 1µl. Boil the samples for 2 mins, chill on ice and load 1µl onto the gel. The length of run is chosen so as to be appropriate for the total length of the DNA fragment. Up to three successive loadings can be made on one gel. Samples have to be reboiled before each loading. Xylene cyanol runs equivalent to 60 bases on a 6% acrylamide gel with 7M urea.

Following autoradiography, it is possible to read up to 180 nucleotides from one gel.

3.7.d. Autoradiography

After electrophoresis, remove one of the glass plates and cover the exposed gel with cling film. Autoradiograph at -70°C using Kodak X-omat H film and an intensifying screen (Cronex Lighting-Plus, Dupont (UK), Huntingdon, Cambs, England). 10,000 counts per loading requires an overnight exposure. If less than 10,000, increase the exposure time accordingly.

3.8. Cloning into M13

The aim of cloning into bacteriophage M13 is to take fragments of double-stranded DNA, and using M13 RF (replicative form) DNA as a vector, produce from the resulting virus, pure single-stranded DNA template suitable for Sanger "dideoxy sequencing".

3.8.a. Preparation of insert and vector

As a rule, it was found better to utilise as few purification steps as possible.

Insert

In most cases, it is necessary to initially separate the insert DNA from the parent plasmid by restriction enzyme digestion and agarose gel electrophoresis. This DNA fragment can then be further restricted for particular cloning experiments. After digestion, heat inactivate at 70°C for 15 minutes. If a thermo-stable enzyme has been used, inactivate by phenol extraction.

Vector

As with the preparation of insert, the simplest procedure for the preparation of vector is to carry out the required enzyme digest and heat inactivate or phenol extract. However, any uncut circular molecules will subsequently transform with a high efficiency to give a high background of blue plaques. If one wishes to reduce this background, purify the linear form of the vector by agarose gel electrophoresis. If the vector has been linearised by digestion with a single enzyme, then subsequent ligation and transformation will again result in a high background of blue plaques due to religation of the vector. This can be reduced by removing the 5' phosphates with bacterial alkaline phosphatase and phenol extraction. However, when deciding on any purification step, it must be borne in mind that while reducing the background, it is also inevitable that there is a reduction in recombinant plaques.

3.8.b. Ligation

Ligase buffer (1x)

50mM	Tris-HCl
8mM	$MgCl_2$
pH 7.6	

This is prepared as a 10x stock and stored at 4°C.

The ligation is carried out in a 1.5ml Eppendorf tube. Use 200ng of vector and a 3x molar excess of insert. Carry out the reaction in a final volume of 10µl

containing 1µl of 10x ligase buffer (see above), 0.1mM ATP and 10mM dithiothreitol. If the insert or vector preparations have not been gel purified or phenol extracted, the presence of restriction enzyme buffer cancels the need to include ligase buffer. Add 0.5 units of T4 DNA ligase (Bethesda Research Laboratories (U.K.) Ltd., Cambridge). Incubate at 16°C for 2 hrs. Dilute the volume 10 fold with a mixture containing 1x ligase buffer, 0.1mM ATP and 10mM DTT. Add 1 unit of T4 DNA ligase and incubate overnight at 16°C. By carrying out a two stage ligation reaction you favour two opposing events. At high DNA concentrations, intermolecular reactions are favoured. On dilution, the conditions then favour intramolecular collisions (Dugaiczyk et. al., 1975). By carrying out this two stage procedure, we hopefully produce hybrid molecules of one vector containing one insert. However, during the course of this project, I have found that transformation with the concentrated ligation mix gives a high efficiency of recombinant formation.

3.8.c. Transformation

Preparation of Competent Cells

Inoculate 50ml of 2x YT (3.1.a.) with 0.5ml of an overnight culture of JM103. Shake at 37°C, until O.D.660 is 0.3-0.4. During this time, inoculate 5ml of 2x YT with 50µl of the primary culture and shake at 37°C during making competent cells to obtain exponentially growing cells for plating.

Collect the cells by centrifugation at 2500 rpm for 10 mins at 4°C. Pour off the supernatant and resuspend the pellet in 25ml of ice-cold 50mM $CaCl_2$. Keep on ice for 20 mins. Centrifuge at 4000 rpm for 10 mins at 4°C and resuspend the cells in 5ml of 50mM $CaCl_2$. Keep on ice until ready to use.

Transformation and Plating Out

Aliquot 0.3ml of competent cells to 5ml sterile snap cap tubes. Add the ligation mixes. Also, as a check of transformation efficiency, add 1ng of uncut

vector to 1 tube of cells. Keep all tubes on ice for 40 minutes. During this time melt some soft top agar (3.1.b.) and keep in a 42°C water bath. After the 40 minute incubation on ice, heat shock the transformation mixes at 42°C for 2 mins. Place at room temperature while adding other ingredients to the soft top agar. To 3ml of agar add 25µl of 2% X-Gal (5-bromo-4-chloro-3-indolyl-β-galactoside, (Sigma): 2% in dimethyl-formamide), 25µl of 100mM IPTG (Isopropyl-β-D-thio-galactopyranoside, Sigma), and 0.2ml of exponentially growing cells. Mix and add to the transformation mixes and pour onto YT plates (3.1.b.). Allow the plates to set, invert and incubate at 37°C overnight. Transformed cells show up as plaques on a lawn of uninfected cells. White plaques indicate transformation by recombinant molecules and thus provide the starting material for single-strand template preparation.

3.9. Preparation of Single-Stranded Template

It is important to ensure that all equipment and reagents are nuclease free to prevent random nicking of the DNA. Glassware should be washed, rinsed thoroughly in double-distilled water and heat sterilised. Plastic ware such as eppendorf tubes and disposable pipette tips should also be sterilised. Wear gloves throughout the procedure.

Up to 24 single-strand preps can be handled comfortably in one day. Pipette sufficient 2x YT for the number of plaques being picked into a sterile glass tube (1.5ml per plaque). To this, add 1/100x volume of an overnight culture of JM103. Mix well and aliquot 1.5ml into each of the required number of 10ml sterile snap cap tubes. Inoculate each tube with a single white plaque using sterile wooden cocktail sticks. It is important to have good aeration of the bacteria during subsequent growth phase, so shake the tubes hard (350 rpm) at 37°C for 5 hours. Transfer to 1.5ml Eppendorf tubes and centrifuge for 5 mins (9950g). Tip off the supernatant into a new eppendorf tube. Store the bacterial pellet at 4°C (this

provides a double-stranded stock of recombinant molecules). It is important to remove all remaining traces of cells from the supernatant, as contamination by chromosomal DNA on the subsequent sequencing gels gives rise to a high background or smearing effect down the length of the gel. To eliminate this, centrifuge the supernatant once more and carefully pipette 1-1.2ml into a new eppendorf tube. Add 200µl of 20% polyethylene glycol (PEG) 2.5M sodium chloride, vortex well and leave at room temperature for 30 minutes. Centrifuge for 5 mins (9950g). At this stage a white viral pellet is visible at the bottom of the tube. Draw off the supernatant using a drawn-out pasteur pipette connected to a water pump. Centrifuge for 3 mins to gather any residual PEG off the walls of the tube. Draw off as before. It is very important to ensure that no traces of PEG remain, as subsequent sequencing gels would show spurious extra banding if all PEG were not removed. The viral coat proteins are removed by extraction with phenol. Add 200µl of TE, 100µl of TE saturated phenol and vortex for 10 secs. Stand for 5 mins and then vortex again. Centrifuge (9950g) for 5 mins and transfer the aqueous phase to a clean eppendorf tube. Add 20µl of 3M sodium acetate (pH 5.6) and 500µl of ice-cold ethanol. Mix well and precipitate at -20°C overnight. Transfer to -70°C for 30 mins. Centrifuge (9950g) for 5 mins to pellet the DNA. Discard the supernatant. Wash the pellet with 300µl of 80% ethanol. Chill at -70°C for 5 mins. Centrifuge and dry the DNA pellet under vacuum. Take up in 20µl of TE, pH 8.0. It is a good idea to check that the virus has grown and that template material has been successfully recovered. Therefore run 2.5µl of the sample on a 1% agarose gel. Store the rest of the template at -20°C, ready for sequencing.

3.10. Sequencing by the Sanger "dideoxy method"

Several variations were attempted (discussed in Chapter 2) before finally deciding to carry out sequencing using the Amersham sequencing kit.

Provided in the Kit

The kit contains 10mM stocks of each of the 4 nucleotide triphosphates and the 4 dideoxy nucleotide triphosphates, chase solution (0.5mM mixture of each of the 4 dNTP's), DNA polymerase 1 (Klenow fragment), klenow reaction buffer (annealing/reaction buffer) and a 17-mer M13 primer.

Therefore the only reagent not contained in the kit is the radioactive nucleotide. In this work, I have used [α-^{35}S]dATPαS at >600 Ci/mmole (Amersham).

3.10.a. Working solutions

Use only sterile double-distilled water for all dilutions. Store all nucleotide stocks and working solutions at -20°C.

deoxy NTP working solutions: Dilute the 10mM stocks supplied in water (1 in 20), to give 0.5mM working solutions. 0.5mM dATP is not required when sequencing with [α^{35}S]dATPαS.

deoxy NTP mixes (A°, C°, G°, T°): make up as follows:

	°A	°C	°G	°T
0.5mM dCTP	20µl	1µl	20µl	20µl
0.5mM dGTP	20µl	20µl	1µl	20µl
0.5mM dTTP	20µl	20µl	20µl	1µl
1x TE, pH 8.0	20µl	20µl	20µl	20µl

dideoxy NTP working solutions: It is possible to "fine-tune" the chain termination reaction by altering the ddNTP concentrations. Increasing the dideoxynucleotide concentrations results in shorter fragments, while lower concentrations result in less frequent chain termination and hence longer fragments. Thus the dideoxy concentrations can be altered to suit the particular DNA being sequenced. For 18SrDNA I have found the following concentrations to be ideal. 25µMddATP, 33µMddCTP, 75µMddGTP and 500µMddTTP. Dilutions of the 10mM stocks were made in water.

dNTP/ddNTP mixes (A, C, G, T mix)

To each dNTP mix, add an equal volume of the corresponding ddNTP working solution.

3.10.b. Annealing reaction

The first stage of the sequencing reaction is to anneal the primer to the single-stranded template. Carry out the reaction in a 1.5ml Eppendorf tube. Add 5µl of single-stranded DNA template, 1µl of primer, 1.5µl of Klenow reaction buffer and 2.5µl of double-distilled water. Mix and incubate in a 65°C incubator for 1-2 hrs. The annealing reaction can be carried out the day before the sequencing reaction. If so, store the annealed primer/template mix at -20°C.

3.10.c. Sequencing; polymerisation reactions

Thaw the dNTP/ddNTP mixes, the chase mixture and the label and store on ice. If the annealing reactions have been carried out on the previous day, allow the primer-template mixes to come up to room temperaure. The sequencing reactions are carried out in 1.5ml eppendorf tubes. For each clone being sequenced, label 4 tubes A, G, C and T. For ease of manipulations, it is best to remove the caps. Arrange the tubes in an eppendorf rack tilted from the horizontal. The sequencing reactions are carried out at ambient temperature. To the annealed primer-template, add 15µCi of labelled nucleotide and 1 unit of Klenow fragment. Mix well and aliquot 2.5µl just inside the rim of each of the labelled tubes. To each tube, spot 2µl of the relevant dNTP/ddNTP mix, being careful not to touch the previous addition. Spin briefly (9950g) to mix and thus start the polymerisation reactions. Allow the reactions to continue for 20 minutes. During this time, there may be some "pausing" of primer extension due to the low concentration of the labelled nucleotide dATP. Spot 2µl of chase mix just inside the rim of each tube and again mix with a brief spin at 20 minutes. Chasing with cold dATP will convert into high molecular weight material, any "paused" chains which have not

incorporated a dideoxynucleotide. Stop the reactions after a further 20 minutes by the addition of 4µl of formamide dye (prepared as previously described (3.7.c.) with the addition of bromophenol blue). The samples are now ready to be run on a sequencing gel.

3.10.d. Selective-Screening (T-tracking)

T-tracking is most useful for primary screening in a random cloning experiment. By carrying out only one reaction on a large number of single-strand templates, it is possible to pick out the maximum number of unique sequences for subsequent full sequence analysis.

Annealing Reaction

For each 10 clones, make up 4µl of primer, 6µl of Klenow reaction buffer and 12µl of double-distilled water. Add 2µl of this priming mix to the required number of 500µl microcentrifuge tubes. Add 2µl of DNA template to each tube. Mix and incubate for 1-2 hours in a 65°C incubator. Briefly spin contents to the bottom of the tube.

Sequencing Reaction

To 16µl of $\overset{o}{T}$ /ddTTP mix, add 30µCi of labelled nucleotide and 2 units of Klenow fragment. Mix well and spot 2µl inside each tube of annealed primer-template mix. Start the reaction with a brief spin. Carry on exactly as for full sequencing protocol,using only 1µl of chase mix and 1µl of formamide dye.

3.10.e. Running the gels

Prepare 6% acrylamide gels as previously described (3.7.c.), this time using 26 x 4mm combs. Boil the samples for 3.5 minutes and load 2µl of each. Run the gels at 30mA and 40 watts. To maximise the length of sequence that can be

determined from each clone, two separate loadings are necessary. Run one loading for 4.5 hours and the other for 1.5 hours, or until the bromophenol blue reaches the bottom of the gel. Under these conditions, the bromophenol blue runs at the same position as the unlabelled primer and first few bases, equivalent to approximately 20 nucleotides in total. It is most convenient to load these samples at the same time on two separate gels. However, it is possible to run them both on the same gel if necessary. If this is so, re-boil the samples before loading the shorter run. It is possible to read 250-300 nucleotides from the two loadings.

3.10.f. Autoradiography

After gel electrophoresis, remove one of the glass plates and fix the gel in 2 litres of 10% methanol, 10% acetic acid for 30 minutes. Drain for a few minutes and then layer a wet piece of 3MM filter paper on top of the gel, being careful not to trap any air bubbles. Invert, and lift off the glass plate leaving the gel attached to the paper. Cover with cling film and dry for approximately 20 minutes on a gel drier at 80°C (Bio Rad, Slab Gel Drier Model 1125B, Bio Rad Chemical Division, Richmond, California). The advantages of drying the gel are two fold. Firstly, drying gives improved resolution of the bands. Secondly, drying is especially important when using ^{35}S label, as this radiation has a penetration of only a few µm. After drying, remove the cling film and expose overnight at room temperature using Fuji RX film (Hanimax (UK) Ltd., Hanimax House, Swindon) and a side-lock cassette (Kenex (E-M) Ltd., Langley House, Harlow). A longer exposure may subsequently be required.

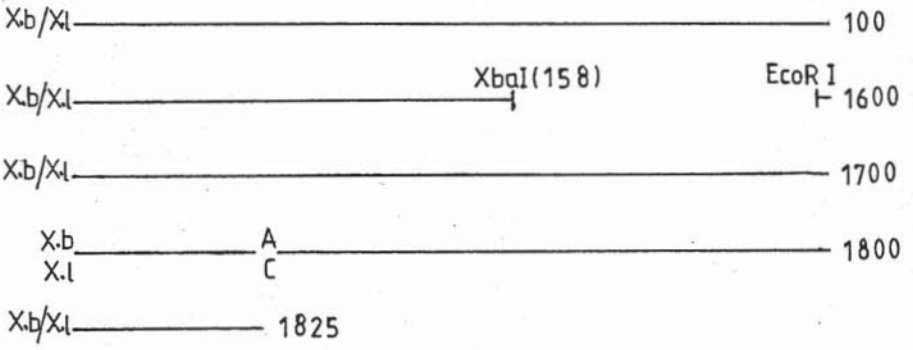


Figure 4.1 Comparative data on the 18S gene region of X. borealis and X. laevis prior to this study

The complete X. laevis sequence was determined by Salim and Maden (1981).

Limited data for Xenopus borealis covered the 5' 163 nucleotides and the 3' 235 nucleotides (Furlong and Maden, 1983).

Comparison of the two sequences demonstrates a single base substitution (C —> A) at nucleotide 1723, numbering according to the Xenopus laevis data.

Region to be covered is this study:

XbaI ——————————————————— EcoRI

CHAPTER 4

ANALYSIS OF THE 18S rDNA NUCLEOTIDE SEQUENCE OF XENOPUS BOREALIS AND COMPARISON WITH XENOPUS LAEVIS

Prior to this present study, only a few hundred nucleotides of Xenopus borealis 18S rDNA sequence have been available for comparison with Xenopus laevis (Furlong and Maden, 1983). A single difference was found to occur about 100 nucleotides from the 3' end of the gene (Figure 4.1.). This difference is a C ——> A substitution which lies in a region already known to be variable in comparisons between the 18S genes of more distantly related eukaryotes. Other such phylogenetically variable regions exist within the 18S gene. The initial objective in this present study was to extend the comparison of the two Xenopus species by completing the analysis of the 18S nucleotide sequence for Xenopus borealis.

4.1. Preparation of the internal region of 18S rDNA from X. borealis clone pXbr101

Plasmid pXbr101 and its subclone pXbr101A were used as sources of DNA for the main sequence analysis (see Figure 4.2). The parent clone contains a complete rDNA transcription unit bound by Hind III restriction sites cloned into the Hind III site of pMB9. The subclone pXbr101A contains the indicated Hind III/EcoRI fragment cloned into pAT 153 (Figure 4.2. (b)). The region to be sequenced in this study is bound by an Xba I site at 158 nucleotides and an EcoRI site at 1597 nucleotides (Figure 4.1). To simplify the subsequent preparation of restriction fragments for sequence analysis, I decided to excise this region from the plasmids. pXbr101 or pXbr101A was digested with XbaI and EcoRI. This results in the region of interest being divided into two fragments, due to the presence of a second Xba I

Figure 4.2. Preparation of the 18S Gene Region from pXbr101 and pXbr101A

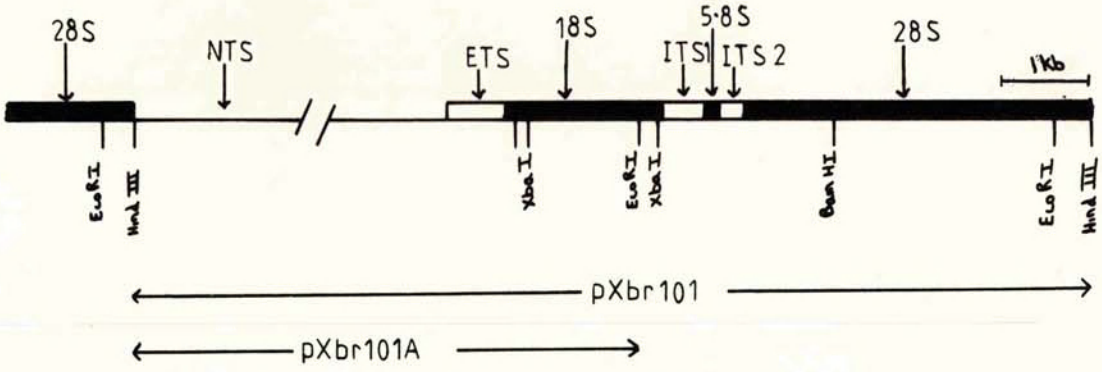
Figure (a) shows the regions of ribosomal DNA contained in plasmids pXbr101 and pXbr101A. (b) shows the orientations of these fragments in pMB9 and pAT 153 respectively.

100µg of each of the two plasmids were digested with EcoRI and XbaI. The resulting fragments were separated by electrophoresis through a 1% agarose slab gel. The bands are numbered in order of decreasing size (c).

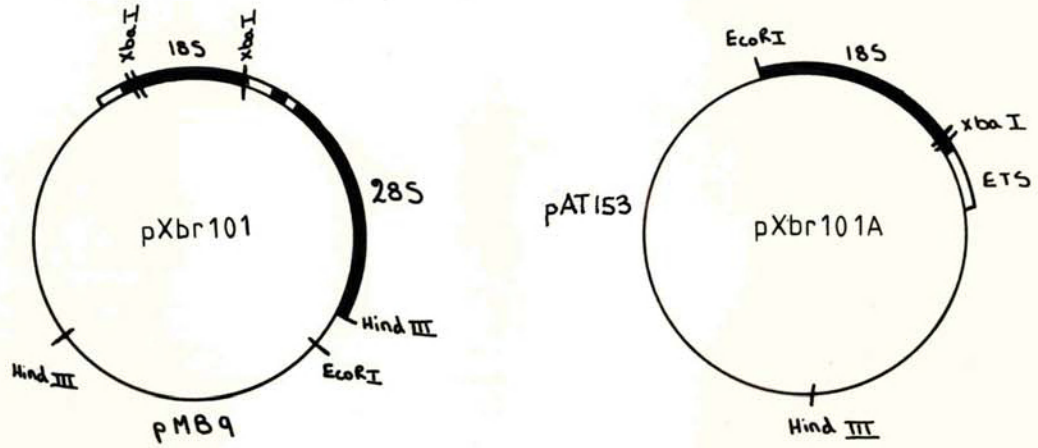
Band 3 from pXbr101 and band 2 from pXbr101A contain the fragment XbaI(260)/EcoRI (1597).

Band 6 from pXbr101 and band 3 from pXbr101A contain the small Xba I fragment (158-260).

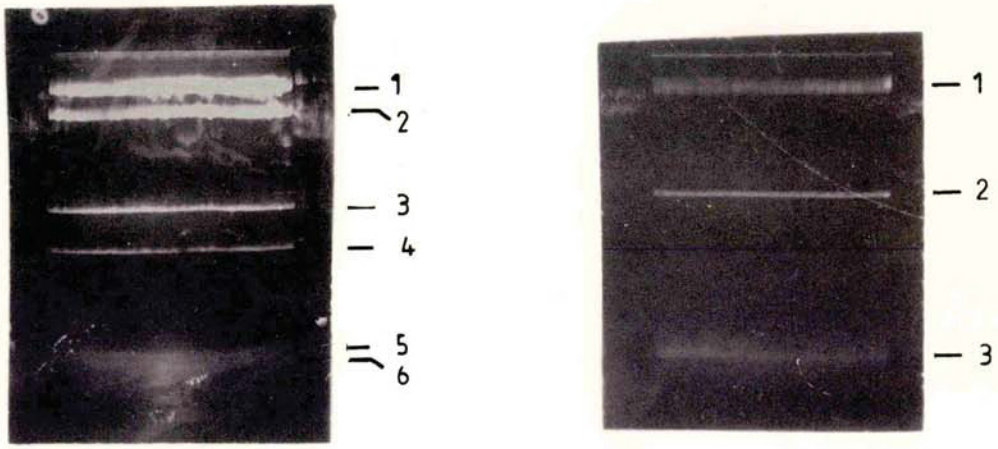
These bands were carefully excised from the gels and eluted in preparation for subsequent sequence analysis.



(a)



(b)



(c)

FIGURE 4.2

site about 100 nucleotides downstream from the first (Figure 4.2(a)). Restriction fragments were separated by gel electrophoresis (Figure 4.2(c)) and the appropriate bands eluted and worked up ready for subsequent analysis (Methods, 3.5(a), (b), (c)).

4.2. Sequencing Strategy

The region of the sequence determined here extends from the Xba I site at 158 nucleotides to the EcoRI site at 1597 nucleotides. This region has already been sequenced in X. laevis and so a detailed restriction map was available (Salim and Maden, 1981). It seemed likely that the same restriction sites would be present in X. borealis 18S rDNA, and this was found to be the case. The strategy followed is shown in Figure 4.3. The complete sequence was built up from a large number of overlapping fragments. Restriction fragments were 5' end-labelled and subjected to secondary digestion to produce fragments labelled at one end as required for Maxam-Gilbert sequencing. Most of the sequence was determined from both strands as indicated in Figure 4.3. Of the few areas where only one strand was covered, the appropriate gels gave a clear unambiguous reading. Any doubtful readings, due to compression effects caused by the formation of local secondary structures, were checked on the opposite strand. The two Xba I sites and the EcoRI site were used as starting points for sequencing runs, but no sequencing runs were carried out through these sites. However, they were previously read through in X. laevis.

4.3. Comparison of the internal 18S gene region of X. borealis clone pXbr101 with the corresponding region in X. laevis

Comparison of the X. borealis sequence determined in this study with the reported X. laevis data reveals almost complete homology within the XbaI (158)/EcoRI (1597) intergene region. On initial inspection, I found two points of difference from the reported X. laevis sequence. Both are discussed in turn below.

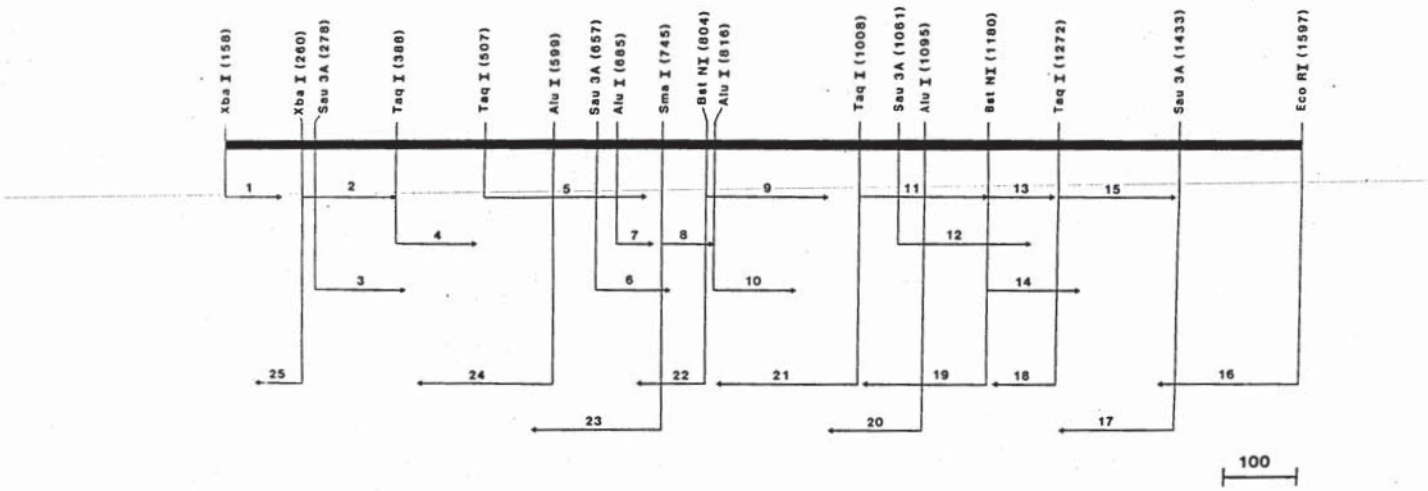


Figure 4.3 Sequencing strategy for the 18S gene region of X. borealis clone pXbr101

Sequencing strategy for the region from the first XbaI site at the 5' end of the 18S gene to the 3' EcoRI site. Only those sites used for 5' end labelling are noted. The upper (rightwards arrows) denote the sequencing runs on the "S" strand (synonomous to RNA). Lower (leftwards) arrows denote sequencing of the "C" strand (complementary to RNA). Fragments are numbered sequentially along the "S" and "C" strands for ease of reference. The arrow tips indicate the maximum length of sequence read from a particular restriction site.

4.3.a. Identification of an "extra" nucleotide following G684

The first point of difference from the published X. laevis 18S sequence was the presence of an A residue following G 684 in the sequence. However, this A residue is not readily detected in Maxam-Gilbert sequencing gels. Rightwards gels contained a compression artefact due to localised secondary structure formation, which made it impossible to define the nucleotide sequence around this area (Figure 4.3., fragments 5 and 6). On first inspection, sequencing of the leftwards strand appeared to provide the sequence unambiguously (Figure 4.3, fragment 23). However, reinspection did suggest the possibility of a slight compression, just enough to mask the presence of a T residue at this point.

This A residue was revealed initially by restriction. An A following G684 on the rightwards strand would give an AluI recognition site (AGCT). Several AluI sites are present around this region (see top of Figure 4.4.). However, Figure 4.4 demonstrates how it was possible to confirm the presence of an AluI site at position 685. The XbaI/SmaI fragment was restricted with AluI and the products separated on a 4% acrylamide gel, alongside a HaeIII digest of the SmaI/EcoRI fragment which provided appropriate size markers. The length of the fragment extending leftwards from the SmaI site was shown to be less than 65 base pairs, indicative of the presence of an AluI site around position 685. In the absence of this site, a fragment of about 80 base pairs would have been found, extending through to the next AluI site at position 662. This was strong evidence to support the presence of an extra A residue at nucleotide 685.

The existence of the extra AluI site at this point was confirmed by sequencing as follows. An AluI digest was performed on the XbaI/EcoRI fragment and the restriction products again separated by polyacrylamide gel electrophoresis. (Figure 4.5). An extra AluI site would give a fragment of 130 base pairs extending from AluI at nucleotide 816 to the extra AluI site at position 685, rather than a fragment of 150 base pairs (bounded by Alu 662). Inspection of the relative spacings on the gel does suggest that bands 3 and 4 (118 base pairs) are more

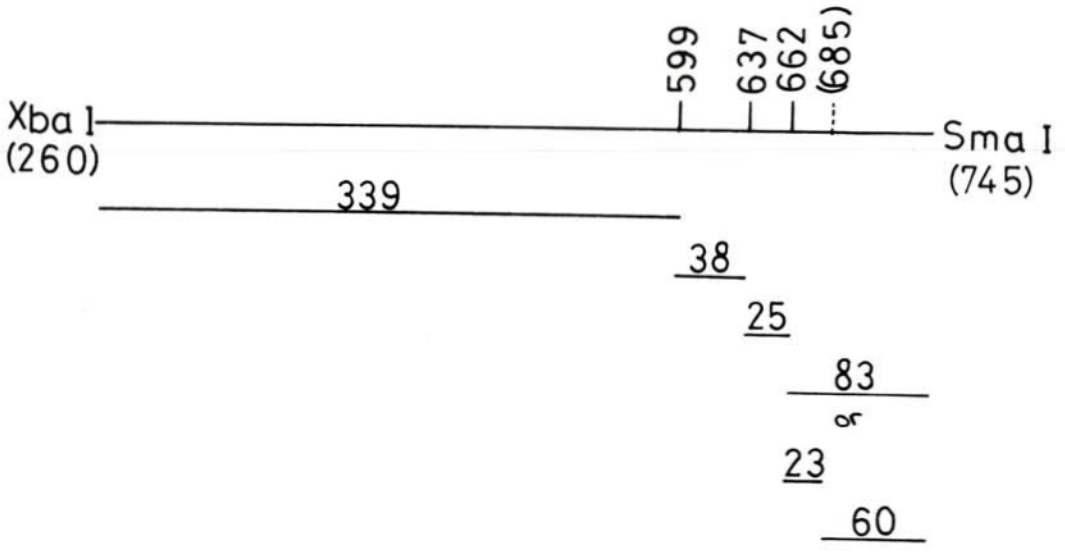
Figure 4.4 Identification of an extra AluI site by restriction analysis

3µg of the indicated XbaI/SmaI fragment was digested with restriction enzyme AluI. The two possible restriction patterns are shown in the top figure.

A HaeIII size marker digest was carried out on the SmaI/EcoRI (1597) fragment.

The products of digestion were separated on a 4% acrylamide gel and the bands viewed by U.V. illumination. From the gel it can be seen that no band is present in the AluI track at ~ 80 nucleotides. However, a band is evident running just slightly ahead of the marker at 65 nucleotides. This is indicative of an extra AluI site at position 685. Band 4 in the AluI track will now be a doublet containing the small 25 and 23 base pair fragments.

Note: The 65 base pair band in the HaeIII track and the 60 base pair band in the AluI track have lower intensities than expected. These bands run coincident with the xylene cyanol marker dye which interferes with the U.V. illumination.


599
637
662
(685)
Xba I
(260)
Sma I
(745)
339
38
25
83
&
23
60

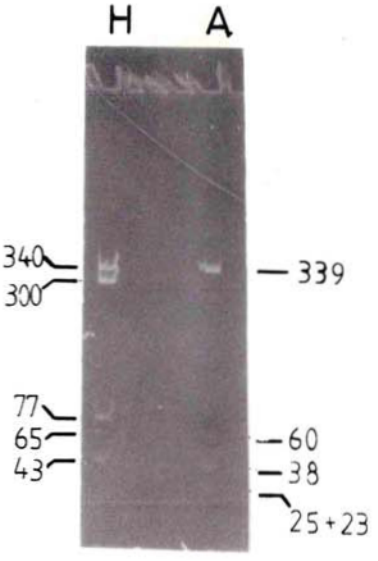

H
A
340
300
339
77
65
43
60
38
25+23

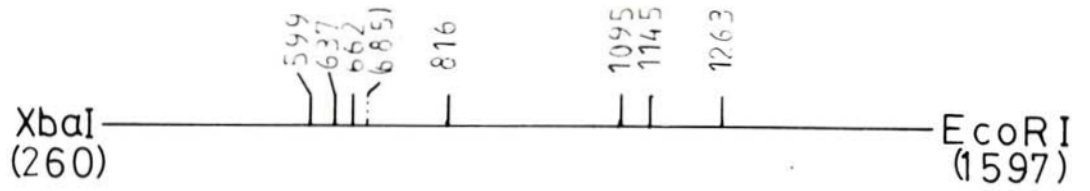
FIGURE 4.4

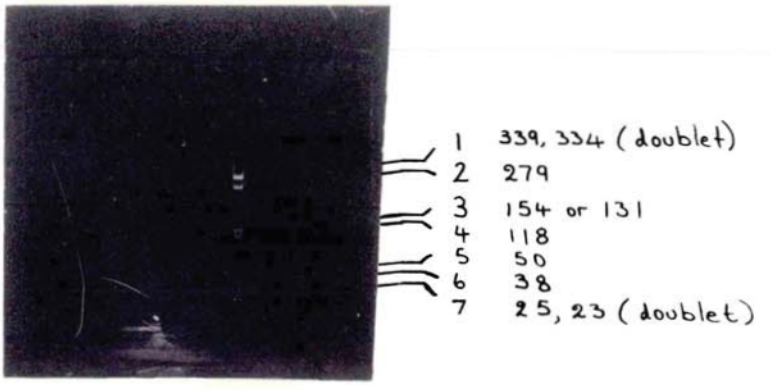
Figure 4.5 Attempt to sequence rightwards from the extra Alu I site

5µg of the XbaI/EcoRI fragment was digested with AluI and the products separated on a 4% acrylamide gel. If the extra AluI site is real, then we expect to obtain a fragment of ~130 nucleotides rather than ~150 nucleotides. Viewing of the gel shows bands 3 and 4 to be very close together suggesting that band 3 is in fact ~130 nucleotides rather than ~150. This band was carefully excised, labelled at the 5' end and secondary restricted with BstNI.

Sequencing was carrried out according to Maxam-Gilbert and the gel is shown at the bottom of the figure. Reading of the sequence can be seen to start at nucleotide 690 indicating that labelling has occured at nucleotide 685 rather than at 662. Nucleotides are numbered according to the complete sequence (Figure 4.7).

Note: Bands 5, 6 and 7 are not visible on the photograph but could be seen on the original gel.





Fragment 3: 662(685) ⟹ BstNI ↓ 816

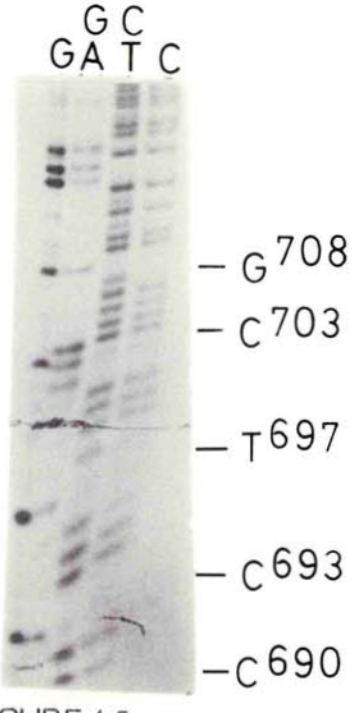


FIGURE 4.5

closely spaced than would be expected if band 3 were 150 base pairs long. Fragment 3 was carefully excised from the gel, 5' end-labelled and subjected to secondary digestion with BstNI. Full Maxam-Gilbert sequencing was carried out from the left-hand end (Methods 3.7). The resulting sequencing gel is shown at the bottom of Figure 4.5. Reading of the sequence commenced at nucleotide 690 indicating the left hand AluI site to be at position 685 rather than 662.

In conclusion, although Maxam-Gilbert sequencing gels through this region failed to identify the A residue at position 685, the combined results of Alu I restriction and sequencing of the relevant Alu I digestion product establishes the presence of the A residue.

I should like to mention at this point that this extra nucleotide is not a real difference between X. borealis and X. laevis. For similar reasons to those described above, Maxam-Gilbert sequencing failed to identify an A residue at this point in the X. laevis sequence. Direct sequencing evidence for the presence of this A residue in both X. borealis and X. laevis will be discussed later in this Chapter.

4.3.b. Identification of a base substitution at nucleotide 679

The internal 1440 nucleotides of the X. borealis clone pXbr101 were found to contain only one real site of variation from the X. laevis 18S rDNA sequence. There is a G —> A substitution at nucleotide 679. This region was sequenced on both strands (fragments 5, 6 and 23 in Figure 4.3). The resulting autoradiographs are shown in Figure 4.6. Sequence of the "S" strand through this region is contained in two gels starting from the Taq 1 site at 507 and the Sau 3A site at 657. Both of these gels demonstrate an A residue at nucleotide 679 (Figure 4.6(i) and (ii)). Similarly, sequencing leftwards from the Sma I site at 745 shows a corresponding T residue on the "C" strand (Figure 4.6 (iii)).

This point in the sequence is only a few nucleotides before the "extra" Alu I site, discussed in 4.3.a. The "compression artefact" that conceals the presence of

Figure 4.6 Identification of a G ——> A base substitution at nucleotide 679

Fragments 5, 6 an 23 all demonstrate a base substitution at nucleotide 679. * denotes this position on each gel.

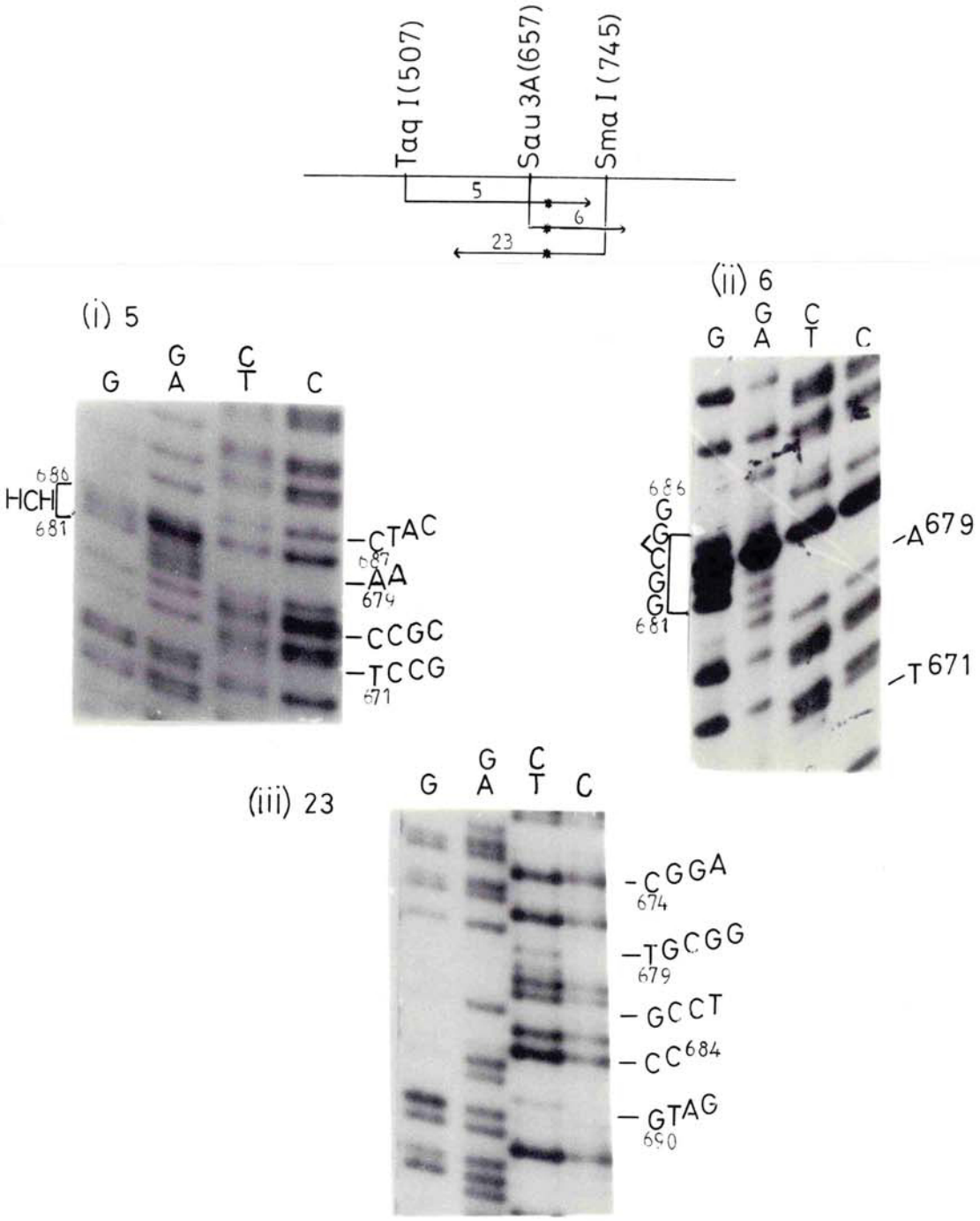
Figures (i) and (ii) show the sequencing runs through this region on the RNA-like strand from Taq I and Sau3A respectively. Both demonstrate an A residue at 679.

Figure (iii) shows the sequencing run through this region on the complementary strand starting from SmaI 745. A T residue is visible at position 679.

Nucleotides are numbered according to the complete sequence (Figure 4.7).

Nucleotides after G684 are numbered to include an A residue at 685. Uncertain nucleotides are described according to Staden (1979). "H" denotes more than one G, "^" denotes that the order is unclear.

All 3 gels demonstrate the complete masking of nucleotide 685.



X.borealis 671 TCCGCCGCAA GGCGAGCTAC 690
X laevis G

FIGURE 4.6

an A residue after G684 on the rightwards strand can be seen in Figures 4.6(i) and (ii). Nucleotides downstream from this point have been numbered accordingly to include the presence of an A residue at position 685. Close examination of the leftwards strand does show a rather thick band in the C + T track at C686 suggesting this band is superimposed with T685 (Figure 4.6(iii)).

4.3.c. Summary of the differences found between the complete 18S rDNA sequence from X. borealis clone pXbr101 and the X. laevis sequence.

The complete 18S gene sequence for pXbr101, obtained from combining this present work with that of Furlong and Maden (1983) is shown in Figure 4.7. There are only two differences on comparison with the 18S rDNA sequence of X. laevis Each of these differences is a base substitution. At nucleotide 679, a G residue in X. laevis is replaced by an A in pXbr101. At nucleotide 1724 (numbering revised to include an A residue at 685), a C residue in X.laevis is replaced by an A in pXbr101.

This initial comparison raises the question outlined in Chapter 2, p. 19. Are these sites of variation fixed within each species or is there intraspecies heterogeneity at these points?

The remainer of this Chapter serves to answer this question.

4.4. Analysis of the sites of variation between X. borealis and X. laevis 18S rDNA in several more clones

4.4.a. Additional clones studied

The clones studied in the further analysis of the two sites of variation are shown in Figure 4.8. Having carried out the initial sequence analysis of the X. borealis sequence on clone pXbr 101, I studied four additional clones from the same series, namely pXbr103, 104, 105 and 106. The additional X. laevis clones studied contained both amplified DNA and chromosomal DNA. Clones containing amplified DNA were pXlr101, 102 and 103. The initial sequence determination of X. laevis

Figure 4.7 Complete 18S rDNA sequence from X. borealis clone pXbr101

The figure shows the complete 18S sequence from clone pXbr101, combining the earlier work of Furlong and Maden, (1983) with this present study. The strand synonomous to RNA is shown. The region analysed in this study is bounded by an Xba I site (TCTAGA) at position 158 and an EcoRI site (GAATTC) at position 1597.

The two differences in the X. laevis sequence are noted by superscripts.

The location of the "extra" A residue (resulting in an extra AluI site) is indicated. Therefore, this A residue becomes nucleotide 685 and all nucleotides downstream are re-numbered "plus-one" with respect to the original numbering of Salim and Maden, (1981).

```
TACCTGGTTG ATCCTGCCAG TAGCATATGC TTGTCTCAAA GATTAAGCCA TGCACGTGTA   60

AGTACGCACG GCCGGTACAG TGAAACTGCG AATGGCTCAT TAAATCAGTT ATGGTTCCTT  120

TGATCGCTCC ATCTGTTACT TGGATAACTG TGGTAATTCT AGAGCTAATA CATGCCGACG  180

AGCGCTGACC CCCAGGGATG CGTGCATTTA TCAGACCAAA ACCAATCCGG GGCCCCCGCG  240

CCCCGGCCGC TTTGGTGACT CTAGATAACC TCGGGCCGAT CGCACGTCCC CGTGACGGCG  300

ACGATACATT CGGATGTCTG CCCTATCAAC TTTCGATGGT ACTTTCTGCG CCTACCATGG  360

TGACCACGGG TAACGGGGAA TCAGGGTTCG ATTCCGGAGA GGGAGCCTGA GAAACGGCTA  420

CCACATCCAA GGAAGGCAGC AGGCGCGCAA ATTACCCACT CCCGACGCGG GGAGGTAGTG  480

ACGAAAAATA ACAATACAGG ACTCTTTCGA GGCCCTGTAA TTGGAATGAG TACACTTTAA  540

ATCCTTTAAC GAGGATCTAT TGGAGGGCAA GTCTGGTGCC AGCAGCCGCG GTAATTCCAG  600

CTCCAATAGC GTATATTAAA GTTGCTGCAG TTAAAAAGCT CGTAGTTGGA TCTTGGGATC  660

X.1.                    G
X.b.GAGCTGGCGG TCCGCCGCAA GGCGAGCTAC CGCCTGTCCC AGCCCCTGCC TCTCGGCGCC  720
                          AluI

TCCCCGATGC TCTTGACTGA GTGTCCCGGG GGCCCGAAGC GTTTACTTTG AAAAAATTAG  780

AGTGTTCCAA GCAGGCCGCG TCGCCTGGAT ACTTCAGCTA GGAATAATGG AATAGGACTC  840

CGGTTCTATT TTGTTGGTTT TCGGAACTGG GGCCATGATT AAGAGGGACG GCCGGGGGCA  900

TTCGTATTGT GCCGCTAGAG GTGAAATTCT TGGACCGGCG CAAGACGAAC CAAAGCGAAA  960

GCATTTGCCA AGAATGTTTT CATTAATCAA GAACGAAAGT CGGAGGTTCG AAGACGATCA 1020

GATACCGTCG TAGTTCCGAC CATAAACGAT GCCGACTAGC GATCCGGCGG CGTTATTCCC 1080

ATGACCCGCC GAGCAGCTTC CGGGAAACCA AAGTCTTTGG GTTCCGGGGG GAGTATGGTT 1140

GCAAAGCTGA AACTTAAAGG AATTGACGGA AGGGCACCAC CAGGAGTGGA GCCTGCGGCT 1200

TAATTTGACT CAACACGGGA AACCTCACCC GGCCCGGACA CGGAAAGGAT TGACAGATTG 1260

ATAGCTCTTT CTCGATTCTG TGGGTGGTGG TGCATGGCCG TTCTTAGTTG GTGGAGCGAT 1320

TTGTCTGGTT AATTCCGATA ACGAACGAGA CTCCTCCATG CTAACTAGTT ACGCGACCCC 1380

CGGCGGTCGG CGTCCAACTT CTTAGAGGGA CAAGTGGCGT TCAGCCACAC GAGATCGAGC 1440

AATAACAGGT CTGTGATGCC CTTAGATGTC CGGGGCTGCA CGCGCGCTAC ACTGAACGGA 1500

TCAGCGTGTG TCTACCCTGC GCCGACAGGT GCGGGTAACC CGCTGAACCC CGTTCGTGAT 1560

AGGGATCGGG GATTGCAATT ATTTCCCATG AACGAGGAAT TCCCAGTAAG TGCGGGTCAT 1620

AAGCTCGCGT TGATTAAGTC CCTGCCCTTT GTACACACCG CCCGTCGCTA CTACCGATTG 1680

X.1.                                                    C
X.b.GATGGTTTAG TGAGGTCCTC GGATCGGCCC CGCCGGGGTC GGCAACGGCC CTGGCGGAGC 1740

GCCGAGAAGA CGATCAAACT TGACTATCTA GAGGAAGTAA AAGTCGTAAC AAGGTTTCCG 1800

TAGGTGAACC TGCGGAAGGA TCATTA                                      1826
```

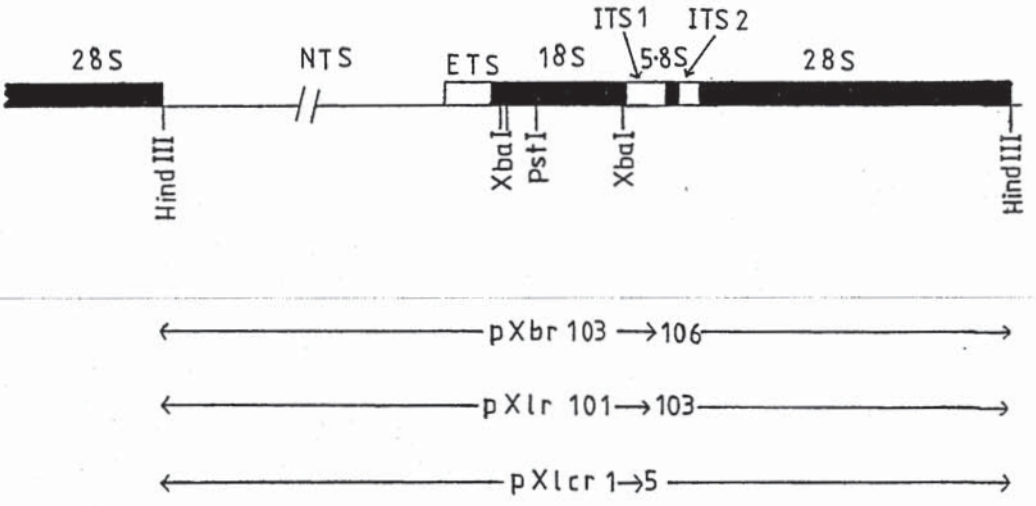


Figure 4.8 Clones studied in the further analysis of the two sites of variation between X. borealis and X. laevis 18S rDNA

Restriction sites relevant to the comparative analysis are noted.

A full description of these clones is given in Figure 2.2 and Table 2.1.

18S rDNA was carried out on clone pXlrl01 (Salim and Maden, 1981). I decided to include this clone in the comparative analysis as further evidence for the original sequence. Five further X. laevis clones, containing chromosomal rDNA, namely pXlcrl-5, (M. Stewart, Ph.D. Thesis, Glasgow University 1983; Stewart et al., 1983) were also studied.

4.4.b. Preliminary examination of the C ——> A substitution at 1724

Not having carried out the initial sequencing of the 3' region of 18S rDNA in pXbrl01, I chose to look at this site of variation first. The Xba I site at 1767 is an ideal starting point for sequencing leftwards through the site of variation at position 1724. I digested all 12 clones with restriction enzyme Xba I to obtain the large Xba I fragment extending from nucleotide 260 through to 1767. All 12 fragments were 5' end-labelled and secondary digested with Pst I in preparation for Maxam-Gilbert sequencing. The X. laevis sequence had shown a C residue at position 1724, therefore sequencing of the complementary strand would show a G residue at this position. Similarly, sequencing of the complementary strand for X. borealis would show a T residue. I carried out G cleavage reactions on the 8 X. laevis clones and full sequencing on the 4 X. borealis clones. G cleavage of the X. laevis clones was unsuccessful, however in all 4 X. borealis clones a T residue was clearly evident at position 1724, complementary to an A residue on the RNA-like strand. The autoradiograph for pXbrl04 is shown in Figure 4.9.

4.4.c. Preparation of DNA fragment containing both points of variation

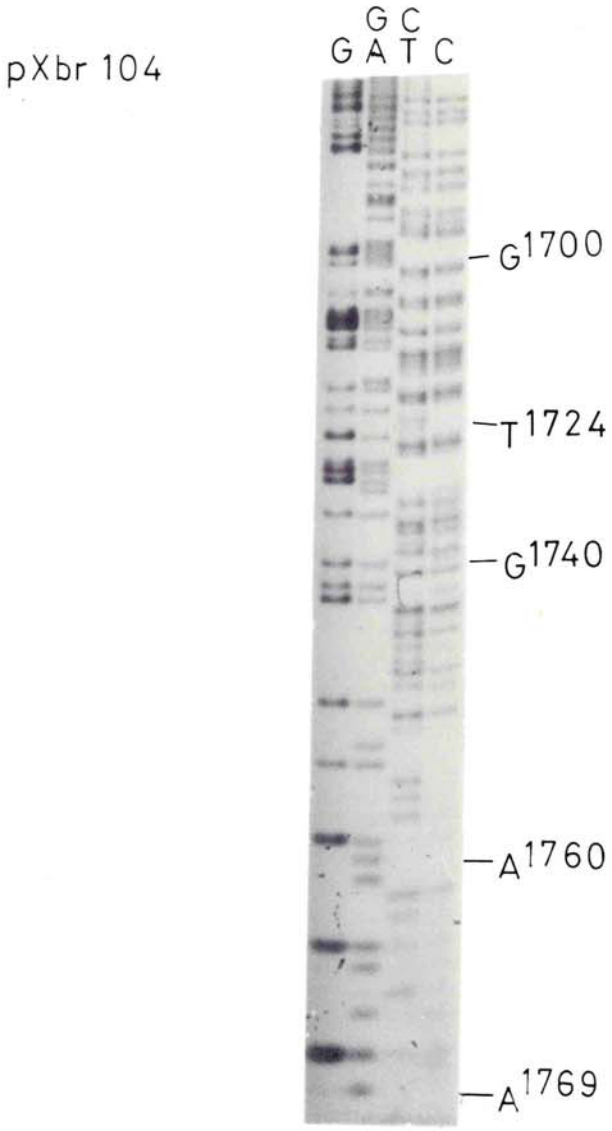
At this time, I realised that the quickest method of looking at both of these sites of variation in several clones was to make use of M13 cloning and dideoxy sequencing. The two sites of variation are conveniently situated towards either end of a Pst I/Xba I fragment (see Figure 4.10). Looking at the list of M13 vectors (Chapter 2, Figure 2.3), pair mpl0/mpll contain unique restriction sites for Pst I and Xba I. By use of this vector pair, it would be possible to clone the Pst I/Xba I

Figure 4.9 Examination of the C ——> A substitution at nucleotide 1724 in X. borealis clones pXbr103-106

For all 4 clones pXbr103-106, the indicated XbaI fragment was 5' end-labelled and secondary digested with PstI. Full Maxam-Gilbert sequencing was carried out from the right hand end, thus resulting in sequence determination of the "C" strand.

The resulting autoradiograph for pXbr104 is shown. Nucleotides are numbered according to the complete sequence (Figure 4.7).

All 4 clones were identical and showed a T residue at nucleotide 1724, indicative of an A residue on the RNA-like strand.



1724

X. borealis 3′ CCGTTGCCGG 5′

X. laevis G

FIGURE 4.9

Figure 4.10 Preparation of DNA fragments containing both points of variation

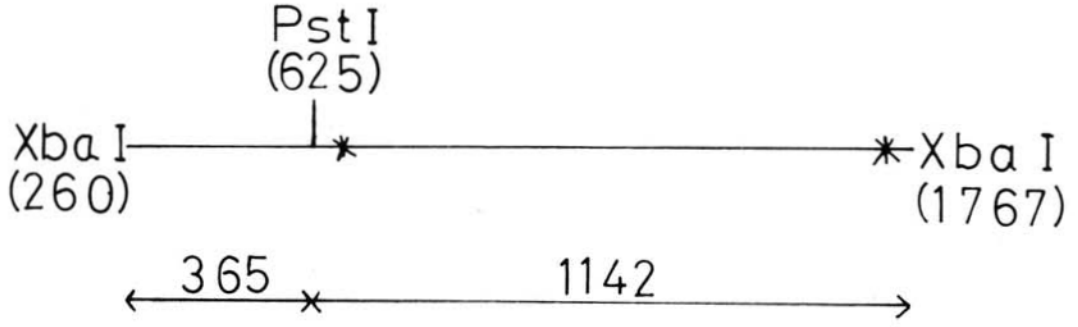
* indicates the positions of the two sites of variation within the large XbaI fragment of 18S rDNA. These sites are situated towards either end of the large PstI/XbaI fragment.

A PstI digest was carried out on the XbaI fragment from all 12 clones and the products of digestion separated on a 4% polyacrylamide gel.

Lanes 1 ——> 12: pXlcr1, 2, 3, 4, 5, pXlr101, 102, 103, pXbr103, 104, 105, 106.

Lane 1 is not visible on the photograph, but the bands could be seen on the actual gel.

The 12 large PstI/XbaI fragments were excised from the gel and prepared ready for subsequent analysis.



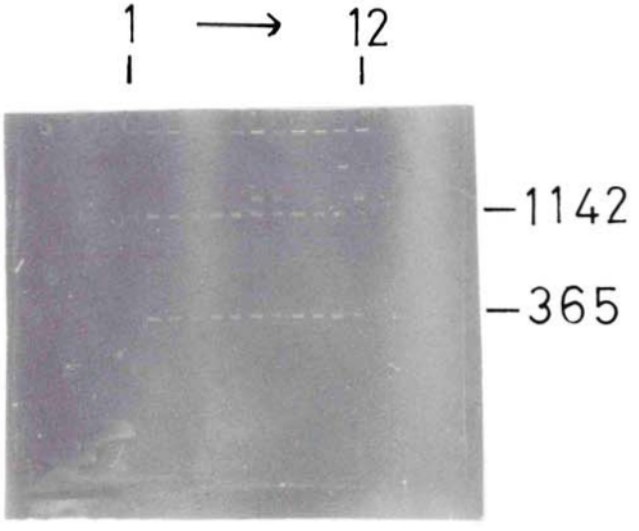


FIGURE 4.10

fragment of interest in opposite orientations and thus sequence from either end. Therefore I digested all 12 of the previously prepared Xba I fragments (4.4.b.) with Pst I and separated on a 4% acrylamide gel (Figure 4.10). The appropriate bands were excised and worked up ready for subsequent cloning into M13.

4.4.d. Analysis of the variation at position 679

Figure 4.11 shows how cloning into mp10 gives the required single-stranded template for the determination of the sequence from the Pst I site through the first point of variation at nucleotide 679. Vector mp10 was cut with Xba I and Pst 1 and run on a 1% agarose gel to separate the small fragment excised from the multiple cloning site region. The large, now linearised, vector band was excised from the gel and worked up ready for subsequent ligation. 12 ligations were carried out, each using 200ng of vector and a 3 x molar excess (100ng) of the Pst I/Xba I fragment. Ligation was carried out in 10µl for two hours before diluting 10X and continuing the ligation overnight. Subsequent transformation into JM103 and plating out resulted in very high numbers of white plaques (recombinant molecules). I prepared the single-strand material from all 12 transformations (Methods, 3.9.)

At position 679, the initial sequence analysis had shown a G residue in X. laevis and an A residue in X. borealis clone pXbr101. I chose to carry out G and A reactions on all 12 clones. This was carried out following the protocol of A. Bankier, MRC, Cambridge and using ^{32}P labelled dATP. A selection of the results are shown in Figure 4.12. By looking at the expected pattern of G and A residues within the full sequence, it is possible to align the patterns on the autoradiographs. In all X. laevis clones there is a G residue at position 679. In all X. borealis clones this G band is missing, being replaced by a band in the A track. I went on to carry out full dideoxy sequencing on all 12 clones, this time using the Amersham sequencing kit and ^{35}S labelled dATP. Two of the resulting autoradiographs, one from each species, are shown in Figure 4.13. In all clones, the sequences reading

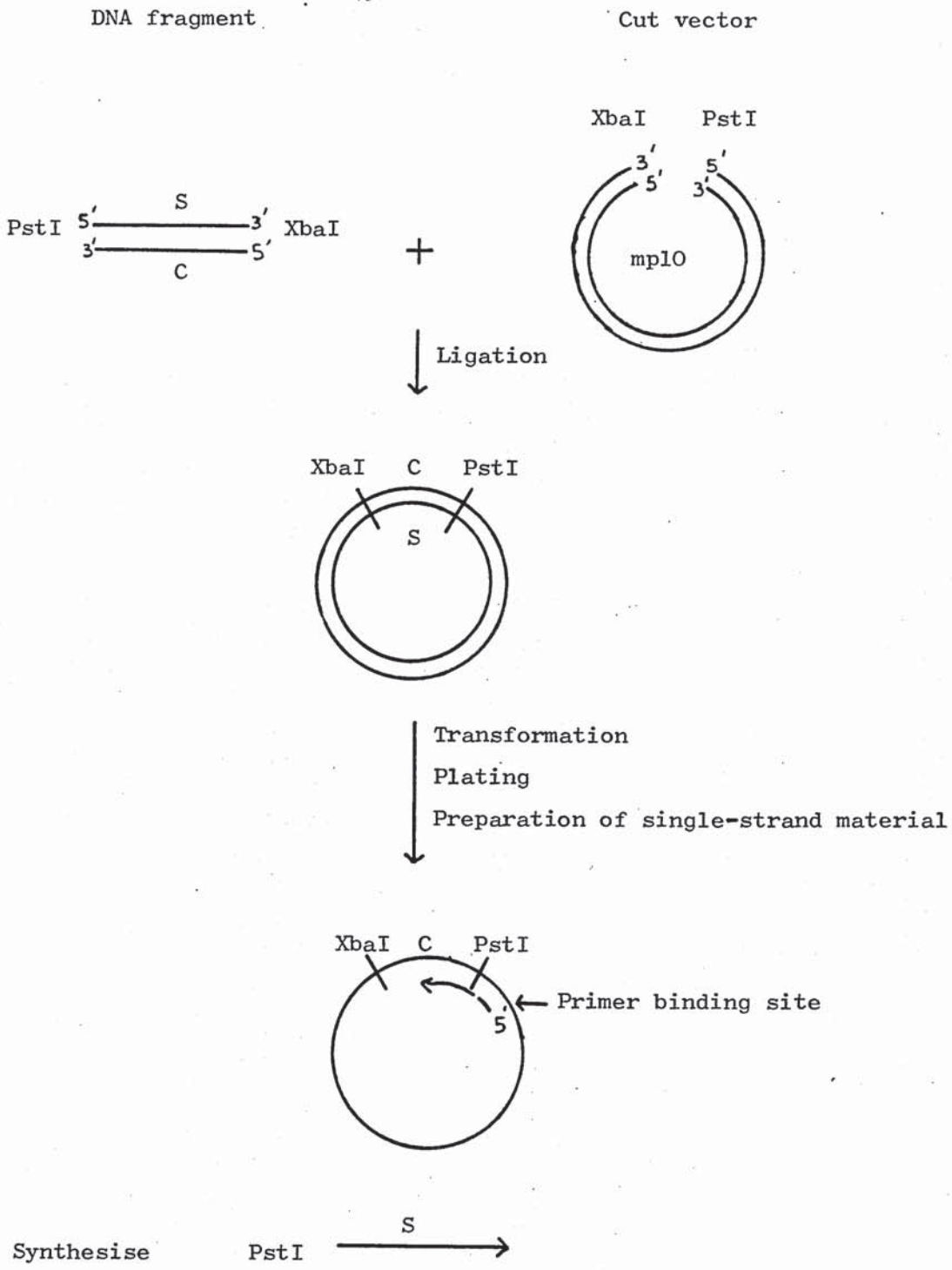
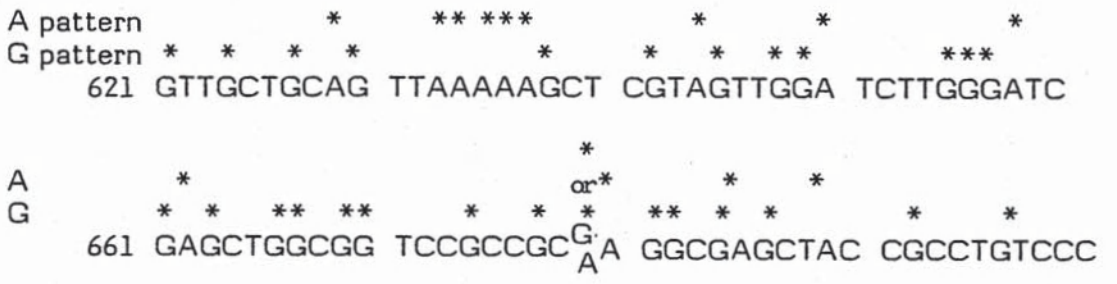


Figure 4.11 Cloning of the PstI/XbaI fragment into M13mp10 to allow sequence determination rightwards from the PstI site

Figure 4.12 Analysis of the variation at nucleotide 679 by dideoxy sequencing: G and A reactions



X. laevis clones pXlr102 and pXlr103 show a G residue at nucleotide 679. X. borealis clones pXbr103, 105 and 106 show an A residue at this position. pXbr104 is a poor quality template, however, the appropriate bands were visible on the autoradiograph. All clones show the "extra" A residue at nucleotide 685.

The compression effect in the region 680-687, experienced in Maxam-Gilbert sequencing, is also evident in the dideoxy method.

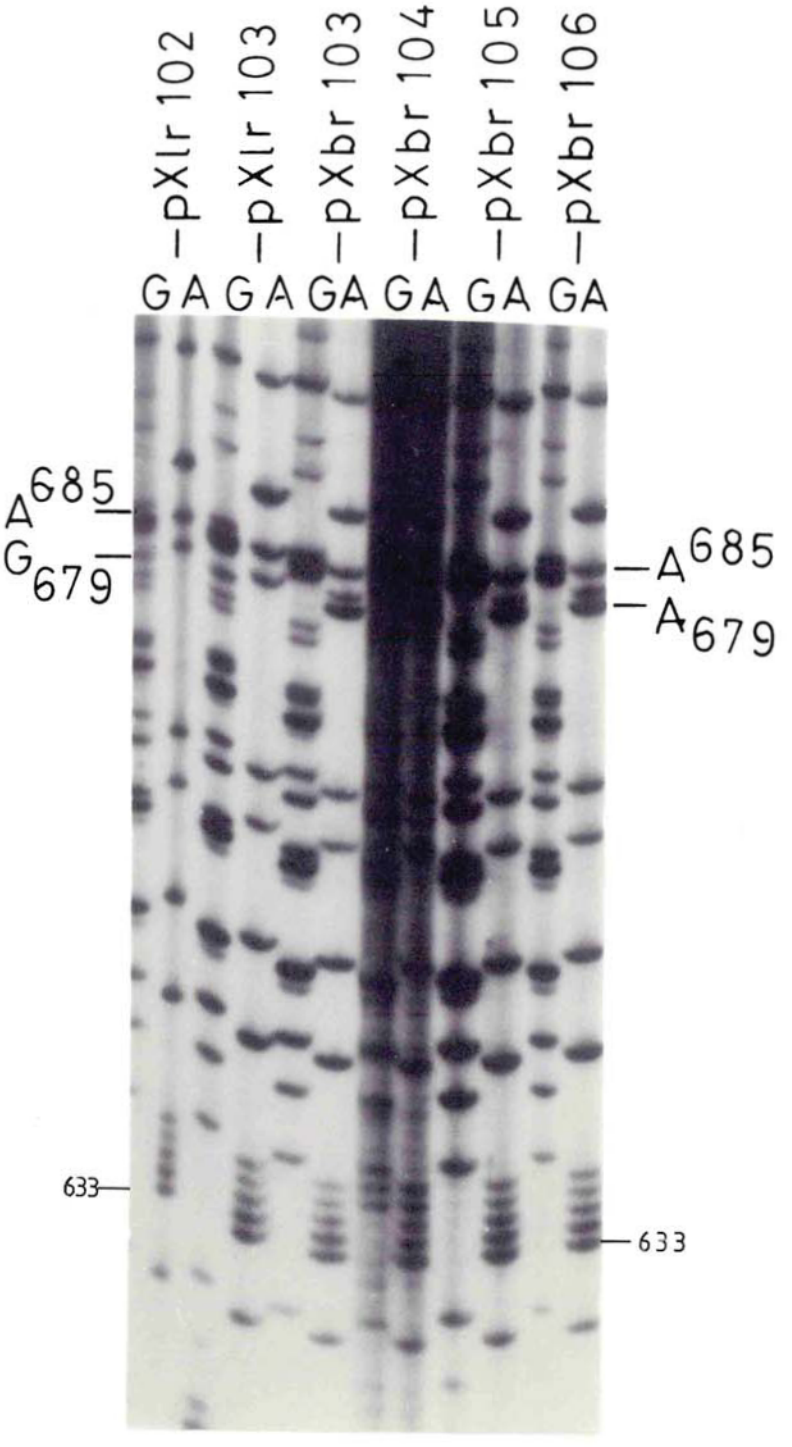

pXlr 102
pXlr 103
pXbr 103
pXbr 104
pXbr 105
pXbr 106
G A G A GA G A G A G A
A685
G679
A685
A679
633
633

FIGURE 4.12

Figure 4.13 Analysis of the variation at nucleotide 679 by full dideoxy sequencing

Full dideoxy sequencing was carried out on X. laevis clones pXlcr 1-5, pXlr101-103 and X. borealis clones pXbr103-106, sequencing rightwards from the PstI site. The relevant regions from the autoradiographs for pXlr102 and pXbr103 are shown.

Nucleotides are numbered according to the complete sequence (Figure 4.7). The sequences are identical, except for the G——>A substitution at nucleotide 679. All other clones are in agreement with the two shown.

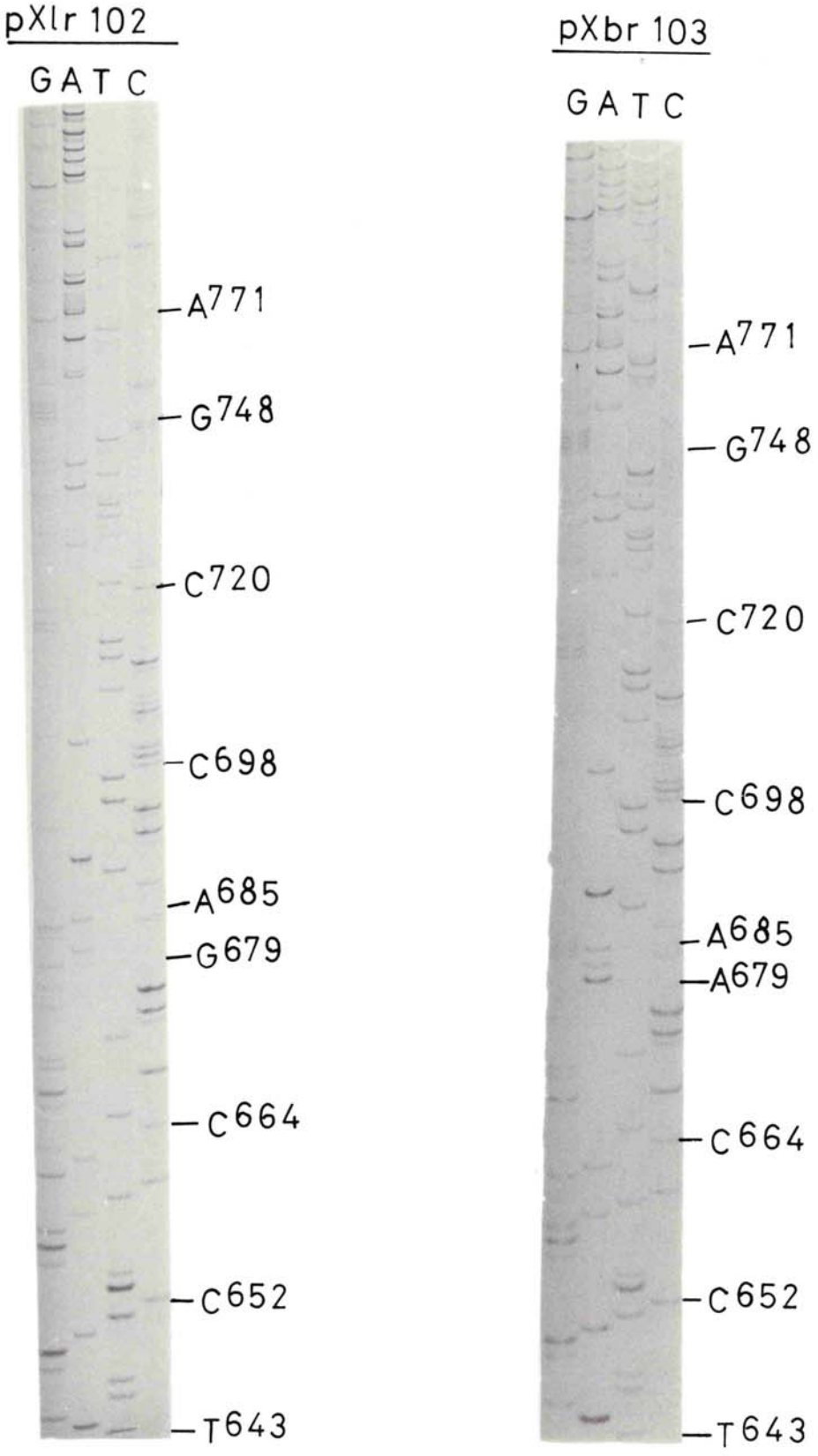

pXlr 102
G A T C
A771
G748
C720
C698
A685
G679
C664
C652
T643
pXbr 103
G A T C
A771
G748
C720
C698
A685
A679
C664
C652
T643

FIGURE 4.13

rightwards from the Pst I site were identical, except for the G ——> A substitution at nucleotide 679.

Comparison of the gels in Figures 4.12 and 4.13 demonstrates the increased resolution in autoradiography on using nucleotides labelled with ^{35}S rather than ^{32}P (discussed in Chapter 2, 2.3.e.).

4.4.e. Analysis of the variation at position 1724

Having already shown the presence of a T residue in X. borealis clones pXbr103, 104 105 and 106 by full Maxam-Gilbert sequencing of the complementary strand through this region (see 4.4.b.), I only required to clone the PstI/XbaI fragment of the 8 X. laevis clones into mp11. Figure 4.14. demonstrates how cloning into mp11 results in the required template for sequence determination leftwards from the XbaI site.

Ligations and transformations were carried out as described in 4.4.d, substituting mp11 for mp10. Again a high number of white plaques were obtained on plating out. Single-strand material was prepared ready for dideoxy sequencing. Full sequencing was carried out on all 8 X. laevis clones using the Amersham sequencing kit and ^{35}S labelled dATP. Two of the gels are shown in Figure 4.15. All 8 clones showed a G residue at 1724, complementary to a C on the RNA-like strand.

4.4.f. Direct evidence for the presence of an A residue at position 685

Because of the unique specificities of dideoxy terminators, it has been possible to obtain direct evidence for the presence of an A residue at position 685 (discussed in 4.3.a.). Sequencing rightwards from the PstI site gives the nucleotide sequence through this region (see Figures 4.12 and 4.13). Dideoxy sequencing again shows compression effects in this area, as was previously found in Maxam-Gilbert sequencing (See Figure 4.6), however an A residue is clearly visible within this compression. It's presence had obviously been obscured by a co-migrating G

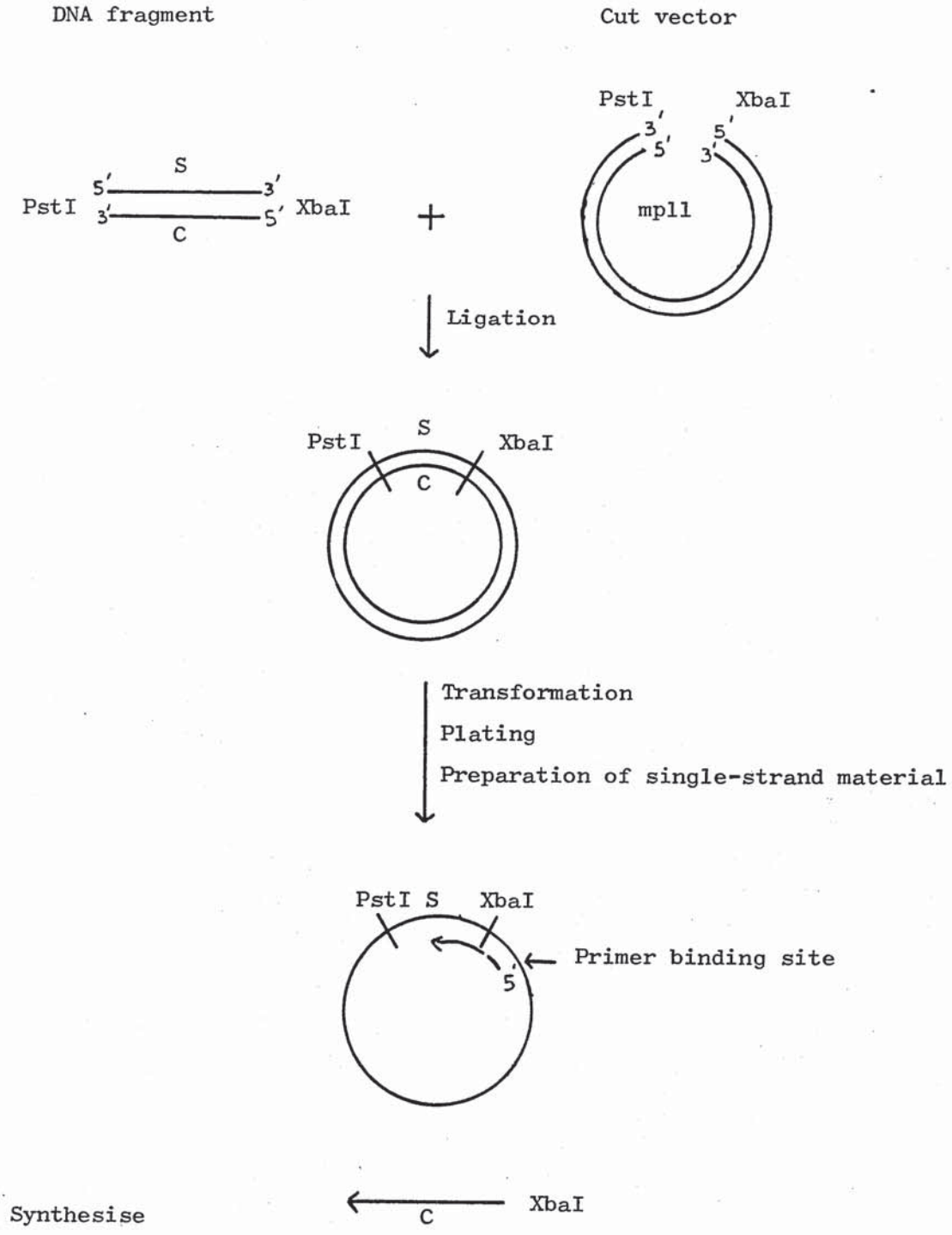


Figure 4.14 Cloning of the PstI/XbaI fragment into M13mp11 to allow sequence determination leftwards from the XbaI site

Figure 4.15 Analysis of the variation at nucleotide 1724 : full dideoxy sequencing of the 8 X. laevis clones

Full dideoxy sequencing was carried out on the X. laevis clones pXlcr1-5, pXlr101, 102 and 103, sequencing leftwards from the XbaI site at 1767.

The resulting autoradiographs for pXlcr2 and pXlcr5 are shown. Nucleotides are numbered according to the complete sequence (Figure 4.7). The sequences are identical and show a G residue at position 1724 complementary to a C residue on the RNA-like strand. The 6 other X. laevis clones agree with the two shown.

Code for uncertain nucleotides: D = more than 1 C

H = more than 1 G.

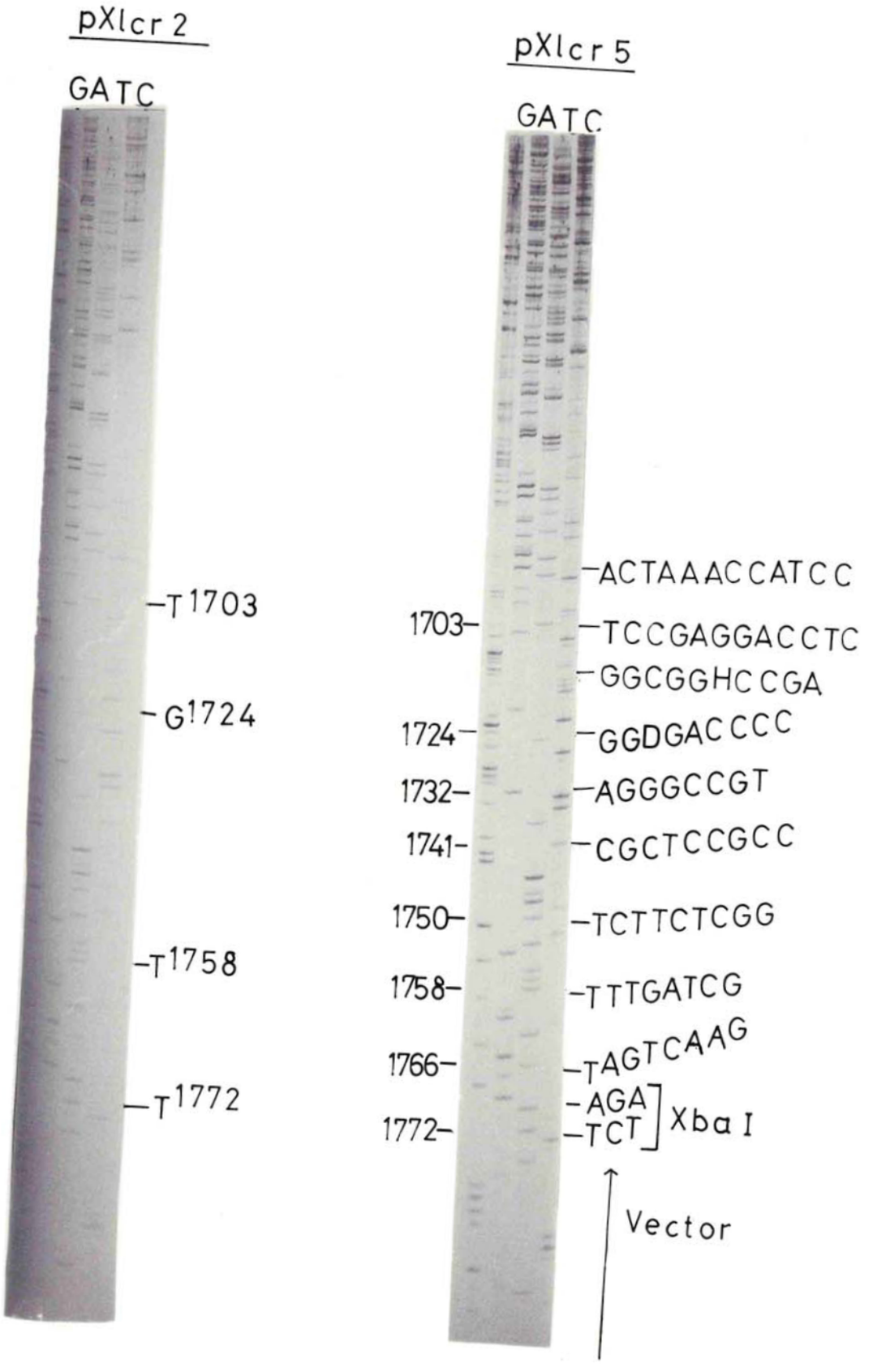

pXlcr 2
GATC
T 1703
G 1724
T 1758
T 1772
pXlcr 5
GATC
1703
1724
1732
1741
1750
1758
1766
1772
ACTAAACCATCC
TCCGAGGACCTC
GGCGGHCCGA
GGDGACCCC
AGGGCCGT
CGCTCCGCC
TCTTCTCGG
TTTGATCG
TAGTCAAG
AGA
TCT
Xba I
Vector

FIGURE 4.15

residue in Maxam-Gilbert sequencing. All 8 X. laevis clones and 4 X. borealis clones analysed in the comparative study demonstrate the presence of this A residue at position 685.

4.5. Summary and Conclusions

The complete sequence of 18S rDNA for X. borealis clone pXbr101 is given in Figure 4.7. Both the X. borealis and the X. laevis sequence are 1826 nucleotides long, one nucleotide longer than that initially reported for X. laevis. The X. borealis sequence differs at only two points from X. laevis 18S rDNA. There is a G —> A substitution at position 679 and a C ——> A substitution at position 1724. Table 4.1 summarises the comparative analysis of these variants carried out on a range of X. borealis and X. laevis clones. At neither site did I find any evidence of intraspecies heterogeneity. Therefore, the initially determined sequences are the predominant ones for the two species. Mutation must have occurred long enough ago (in evolutionary terms) to allow the variants to have attained, or be approaching fixation.

The "extra" A residue at position 685, first demonstrated by restriction analysis of X. borealis clone pXbr101, has subsequently been shown by direct dideoxy sequencing to be present in all X. borealis and X. laevis clones used in this study. Therefore, all nucleotides downstream from A685 have been numbered "plus one" relative to their original numbering in Salim and Maden, (1981).

CLONE	NUCLEOTIDE 679	NUCLEOTIDE 1724	A residue at 685
pXlcr1	G	C	✓
pXlcr2	G	C	✓
pXlcr3	G	C	✓
pXlcr4	G	C	✓
pXlcr5	G	C	✓
pXlr101	G	C	✓
pXlr102	G	C	✓
pXlr103	G	C	✓
pXbr103	A	A	✓
pXbr104	A	A	✓
pXbr105	A	A	✓
pXbr106	A	A	✓

Table 4.1 Summary of the analysis of the two sites of variation and the "extra" nucleotide in a range of X. borealis and X. laevis rDNA clones.

The Table shows the nucleotides present on the RNA-like strand.

CHAPTER 5

ANALYSIS OF THE HUMAN 18S rDNA NUCLEOTIDE SEQUENCE

5.1. Attempt to clone Human rDNA

As outlined in Chapter 2, section 2.1., I initially attempted to prepare human rDNA clones on which to carry out subsequent sequence analysis. I used placenta as a source of human DNA as it was readily available and is a good source of large amounts of DNA. The southern blotting experiments of Wilson et. al. (1978) had shown that restriction with Sal I should generate a fragment containing the 3 complete rRNA genes. I restricted total human DNA with Sal I and electrophoresed through a 1% agarose slab gel along with various rDNA fragments of known size. As expected, restriction of total human DNA gave rise to a long smear as visualised by ethidium bromide staining. The DNA from the gel was transferred to cellulose nitrate and hybridised to partly fragmented and end-labelled 18S and 28S rRNA prepared from HeLa cells (Maizels, 1976). The human rDNA-containing fragment was shown to run coincident with a specific size marker of approximately 11kb. By repeating the agarose gel and cutting out the appropriate band of DNA from the total smear, we had successfully enriched for rDNA. Attempts were then made to clone this fragment into pBR 322 and pAT 153 and to search for positive colonies by hybridisation. It was at this point that the project was unsuccessful. I think the main reason for this was the low efficiency both of ligation of such a large fragment into a much smaller plasmid molecule and of subsequent transformation. To have successfully picked out a human rDNA clone, even after partial purification of rDNA by size selection on agarose gels, it would have been necessary to screen many hundreds if not a few thousand colonies. However, I did not obtain this number of transformants and so the chance of

Figure 5.1 Clones used in the analysis of human 18S rDNA

(a) Shows the regions of ribosomal DNA contained in clones pHrB-SE and pHrA.

(b) Shows the orientation of these fragments in plasmid pBR322.

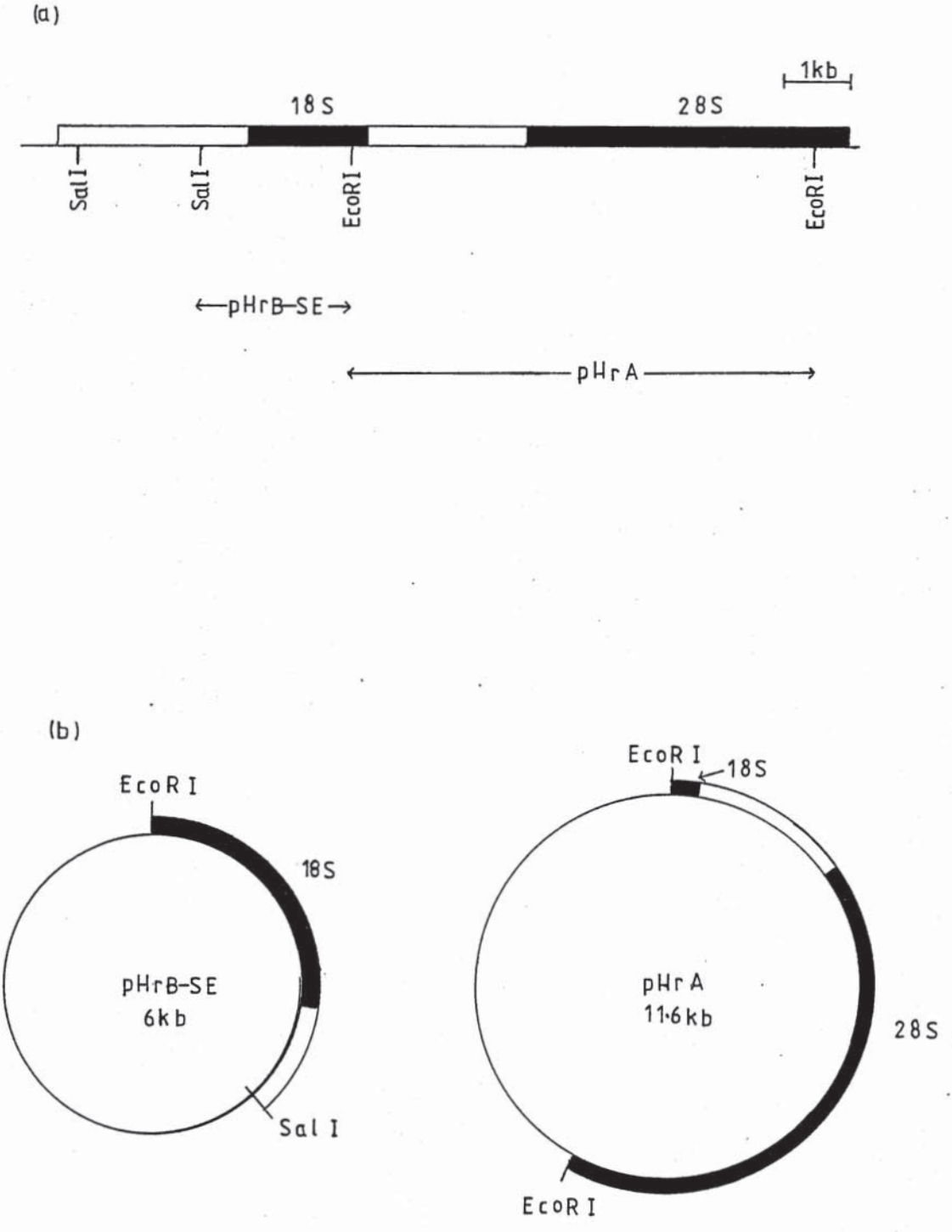

(a)
1kb
18S
28S
SalI
SalI
EcoRI
EcoRI
←pHrB-SE→
pHrA
(b)
EcoR I
18S
pHrB-SE
6kb
Sal I
EcoR I
18S
pHrA
11·6kb
28S
EcoR I

FIGURE 5.1

detecting an rDNA clone was low and the experiment was unsuccessful.

Human clones of the Sal I fragment have recently been successfully prepared in our laboratory by cloning into bacteriophage λ in which the overall cloning efficiency is much higher than in plasmids.

5.2. Clones used in the Sequence Analysis

As discussed in Chapter 2,sequencing of the human 18SrRNA gene was carried out on clones prepared by Wilson's group. A full description of these clones is given in Figure 2.2 and Table 2.1.

The two clones used in this work,pHrA and pHrB-SE, are shown in Figure 5.1. Clone pHrB-SE contains part of the ETS and all of the 18S gene region up to the 3' EcoRI site. Clone pHrA contains the 3' end of the 18S gene region, ITS 1, the 5.8S gene, ITS 2 and the 28S gene region up to the 28S EcoRI site.

5.3. Method of Sequencing

This sequencing objective was met by the method of M13 cloning and dideoxy sequencing (discussed in Chapter 2). On going through the sequencing strategy, I will illustrate the progression from cloning and sequencing particular restriction fragments, to carrying out shotgun cloning and sequencing, thus taking full advantage of the speed of this system over the original system of Maxam-Gilbert where a detailed restriction map is a pre-requisite for sequence analysis.

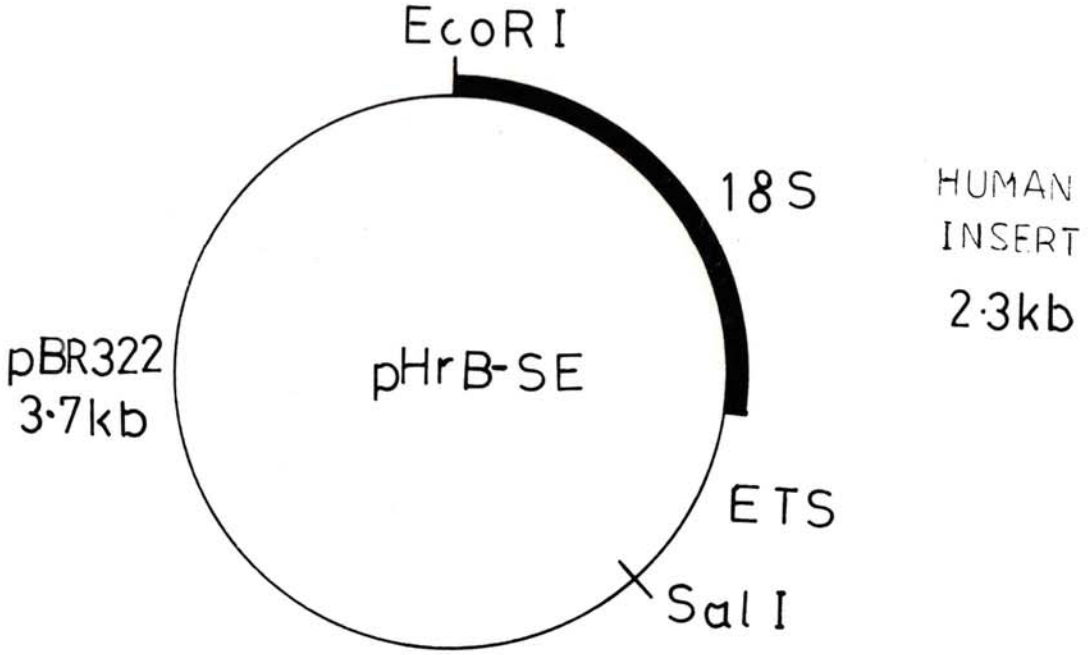
5.4. Isolation of the rDNA fragment from pHrB-SE

To simplify subsequent manipulations, I cut out the SalI/EcoR I human rDNA fragment from plasmid pHrB-SE. The plasmid was digested with restriction enzymes Sal I and EcoRI. The products of digestion were separated by gel

Figure 5.2 Isolation of the rDNA fragment from pHrB-SE

100µg of pHrB-SE was digested with restriction enzymes SalI and EcoRI. The resulting fragments were separated by electrophoresis through a 1% agarose slab gel.

The 2.3kb human DNA fragment was excised from the gel and worked up ready for subsequent analysis.


EcoR I
18S
HUMAN
INSERT
2·3kb
pBR322
3·7kb
pHrB-SE
ETS
Sal I

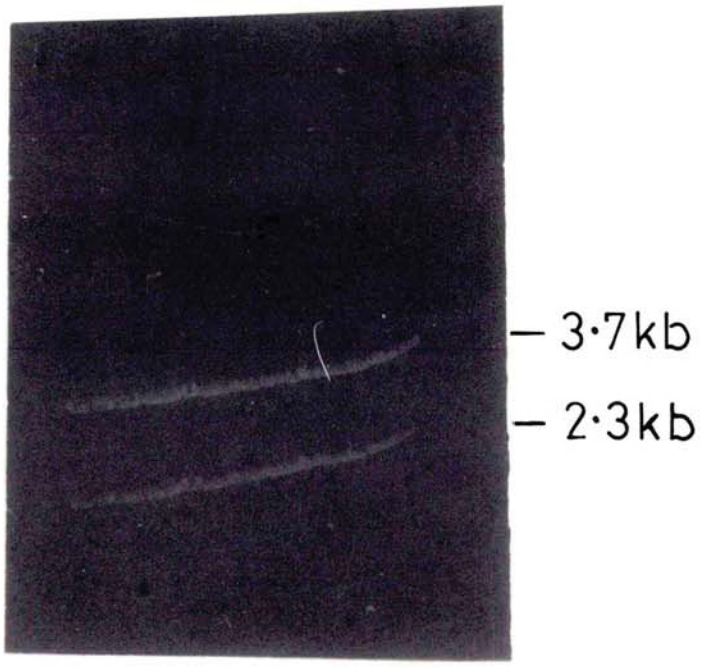

— 3·7kb
— 2·3kb

FIGURE 5.2

electrophoresis (Figure 5.2). The rDNA band was eluted and worked up ready for subsequent analysis.

5.5 Cloning and sequencing specific restriction fragments

To make a start on the human sequence analysis, I decided to make use of the restriction data of Wilson et al (1978, 1982). In particular, I was interested in sequencing from the Xba I and Pst I sites (Figure 5.3(a)). Restriction analysis by Wilson's group suggested the presence of only 1 Xba I site in the 5' region of the 18S gene. However, comparison of this region between frog and rat (Torczynski et al., 1983) shows both species have two Xba I sites in this 5' region. I carried out my initial clonings on the suspicion that human rDNA might also have two closely spaced Xba I sites.

5µg of the Sal I/EcoR I fragment was digested with Pst I and Xba I. After 2 hours digestion, the mix was phenol extracted to remove the restriction enzymes before ligation. Vector pair mp 10/mp 11 (see Figure 2.3) were digested with Sal I and Xba I, Xba I and Pst I, Xba I alone, and Pst I and EcoR I. The restriciton products were run on a 1% agarose gel to purify the linear bands (Figure 5.3.(b)).

Eight ligations were carried out using 200ng of vector and a 3x molar excess (~ 250ng) of the digested Sal I/EcoR I fragment. Each ligation mix therefore contained one of the 8 cut vectors and a mixture of the 4 digestion products of the human rDNA fragment. Table 5.1 shows which fragment is taken up in each ligation.

Transformation and plating were carried out as described in Methods, 3.8.c. No plaques were obtained for transformation number 8. This was not too surprising as very little linear vector was obtained on digestion of mp 11 with Pst I and EcoR I (Figure 5.3.(b)). I prepared single-strand material from a few white plaques for each of the other transformations. Table 5.1 shows the expected sequencing runs from each of the templates. I carried out full dideoxy sequencing and the

Figure 5.3 Restriction of the SalI/EcoRI fragment with XbaI and PstI and preparation of vectors for cloning

(a) Shows the proposed products of digestion of the SalI/EcoRI fragment with Xba I and Pst I.

(b) vector pair mp10/mp11 were both digested with the enzymes shown below and run on a 1% agarose gel.

Lane (1) 0.1µg of uncut mp10

(2) 0.5µg of mp10 cut with SalI and XbaI

(3) 0.5µg of mp10 cut with PstI and XbaI

(4) 0.5µg of mp10 cut with XbaI

(5) 0.5µg of mp10 cut with Pst I and EcoRI

1 empty well

(6) 0.5µg of uncut mp11

(7) 0.5µg of mp11 cut with SalI and XbaI

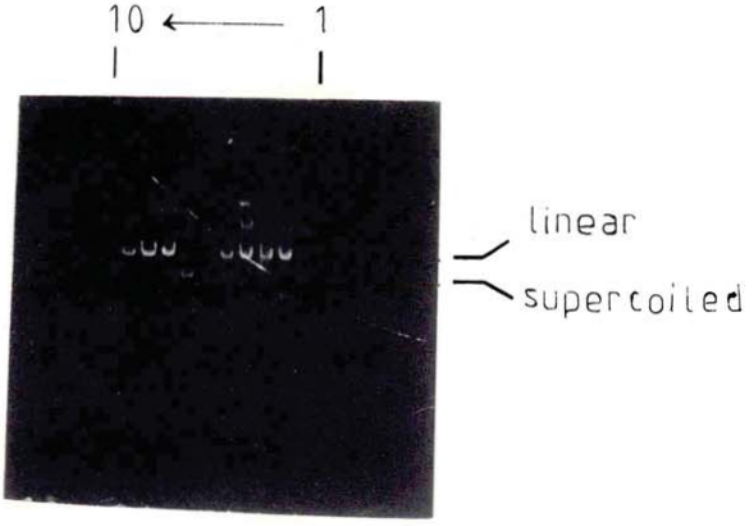
(8) 0.5µg of mp11 cut with PstI and XbaI

(9) 0.5µg of mp11 cut with XbaI

(10) 0.5µg of mp11 cut with Pst I and EcoRI

The linearised vectors were excised from the gel and eluted. Lane (10) showed very little linear material.

(a)



(b)

FIGURE 5.3

LIGATION	VECTOR	INSERT	SEQUENCE
1	mp10 SalI/XbaI	SalI/XbaI	SalI ——>
2	mp10 PstI/XbaI	XbaI/PstI	<—— Pst I
3	mp10 XbaI	XbaI/XbaI	XbaI ——> XbaI OR XbaI <—— XbaI
4	mp10 PstI/EcoRI	PstI/EcoRI	PstI ——>
5	mp11 SalI/XbaI	SalI/XbaI	<—— Xba I
6	mp11 PstI/XbaI	XbaI/PstI	XbaI ——>
7	mp11 XbaI	XbaI/XbaI	XbaI ——> XbaI OR XbaI <—— XbaI
8	mp11 PstI/EcoRI	PstI/EcoRI	<—— EcoRI

Table 5.1 Cloning into mp10/mp11 and subsequent sequence determination

The Table shows which rDNA fragment should be taken up in each ligation. The right hand column gives the direction of sequence determination from the resulting single-stranded template.

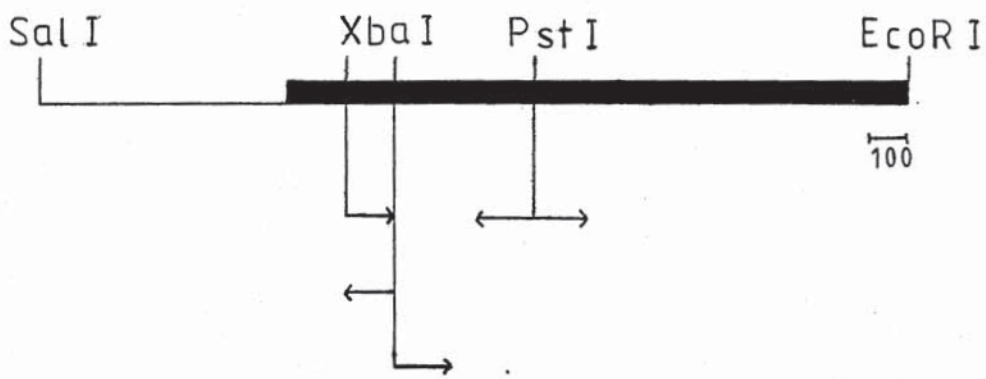


Figure 5.4 Sequence data obtained from the Pst/XbaI digest of the SalI/EcoRI rDNA fragment

The arrows indicate both the direction and length of sequence obtained from a particular restriction site.

No sequencing runs were obtained from the SalI and EcoRI sites, or leftwards from the first XbaI site in the 18S gene. See text for explanation.

combined sequencing map is shown in Figure 5.4.

The first point to note is that the human 18S gene does contain two Xba I sites in the 5' region and not one as shown by Wilson et al., (1982). No sequencing runs were obtained from either end of the Sal I/Xba I fragment. Single-strand templates from transformations 1 and 5 were shown to contain the small XbaI/XbaI fragment. Obviously, in the Sal I/Xba I double digests of mp 10 and mp 11, Sal I had not cut very efficiently, the majority of vector molecules having been cut by Xba I only. Plating of transformations 1 and 5 did result in a high background of blue plaques, indicative of recircularisation of single-cut vector. One point to note, is that in vector pair mp10/mp 11 (Figure 2.3) the recognition sites for Xba I (TCTAGA) and Sal I (GTCGAC) are contiguous. We have found by experience that it is very difficult to successfully carry out double digests with such closely positioned sites.

5.6. Shotgun Cloning and Sequencing

To make most efficient use of M13 cloning and sequencing, it is best to take advantage of the fact that a detailed restriction map is not required prior to sequencing. The task then, is to choose a few enzymes which will cleave the DNA of interest to give a whole range of overlapping fragments of up to a few hundred base pairs. By cloning and sequencing each set of fragments in turn, the complete sequence can gradually be fitted together.

5.6.a. Digestion of pHrB-SE with Sau 3A, Taq I and Hpa II

I chose to use the three restriction enzymes Sau 3A, Taq I and Hpa II to hopefully give a good array of overlapping fragments covering the 18S rRNA gene.

I digested a small amount of pHrB-SE and the purified Sal I/EcoRI fragment with each of the three enzymes to obtain an idea of the number and sizes of fragments obtained. The products of digestion were separated on a 4%

Figure 5.5 Restriction Analysis of pHrB-SE and the purified SalI/EcoRI fragment

The 9 digests shown below were carried out and the products of digestion separated by electrophoresis through a 4% polyacrylamide gel.

(1) 1.5µg of pBR322 cut with Sau3A: Size of fragments in bp. 1374, 665, 358, 341, 317, 272, 258, 207, 105, 91, 78, 75, 46, 36, 31, 27, 18, 17, 15, 12, 11, 8.

(2) 5µg of pHrB-SE cut with Sau3A.

(3) 2µg of SalI/EcoRI fragment cut with Sau3A.

(4) 1.5µg of pBR322 cut with TaqI: size of fragments in bp. 1444, 1307, 475, 368, 312, 141.

(5) 5µg of pHrB-SE cut with TaqI.

(6) 2µg of SalI/EcoRI fragment cut with TaqI.

(7) 1.5µg of pBR322 cut with HpaII: size of fragments in bp. 622, 527, 404, 309, 242, 238, 217, 201, 190, 180, 180, 160, 160, 147, 147, 122, 110, 90, 76, 67, 34, 34, 26, 26, 15, 9, 9.

(8) 5µg of pHrB-SE cut with HpaII

(9) 2µg of SalI/EcoRI fragment cut with HpaII.

Lane (9) did not show any bands. The DNA fragment must have been omitted from the digest.

Lanes (4) and (5) show some extra banding due to incomplete digestion.

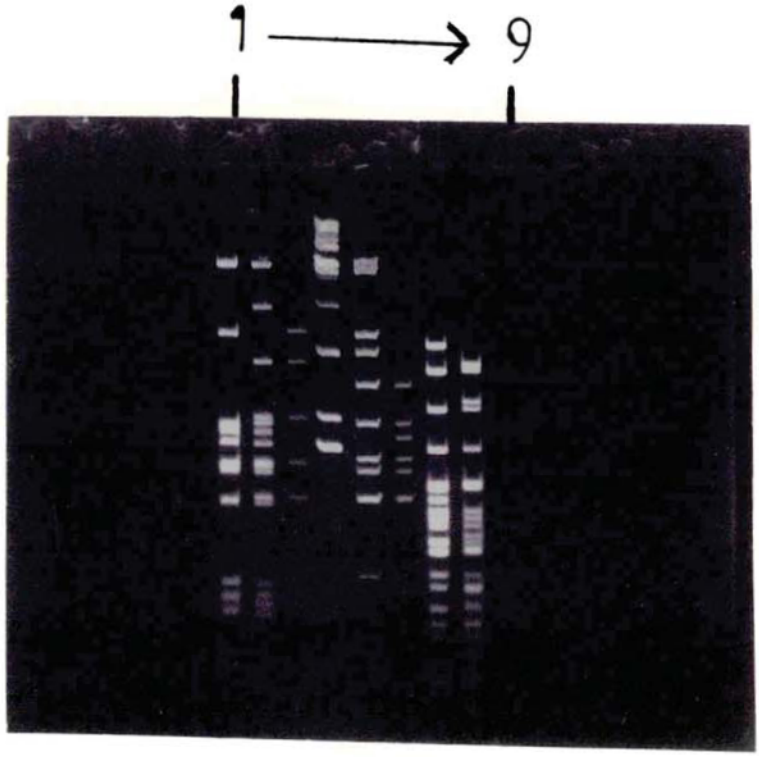


FIGURE 5.5

polyacrylamide gel (Figure 5.5), along with digests of pBR322 to give an indication of the size of fragments obtained. These three digests give a good number of fragments of several hundred base pairs long.

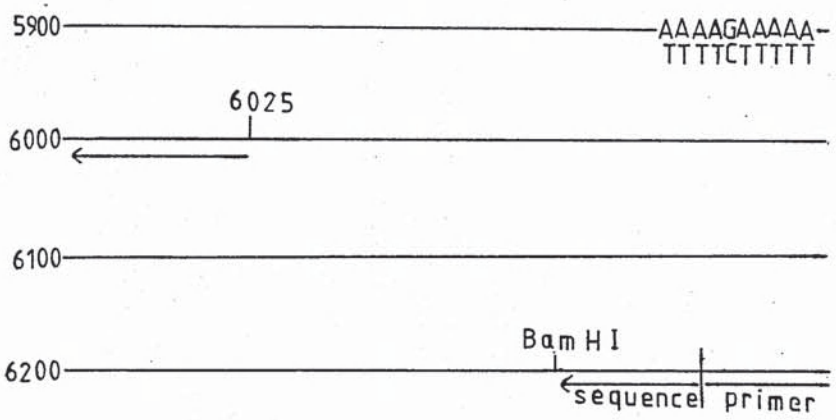
Since the gel gave good resolution of the restriction products, I decided to elute a few of the bands for subsequent cloning and sequencing. I eluted the three largest fragments from the Taq I digest of the Sal I/EcoRI fragment. These were ligated into the Acc 1 site of vector mp 10. Fragment 3 did not give any white plaques on plating. This fragment must have been bounded at one end by the Sal I or EcoRI restriction site and so would have been unable to ligate into mp 10 cut with Acc I alone. Re-inspection of the acrylamide gel (Figure 5.5) does show that this must be the case, as fragment 3 is missing in the corresponding Taq I digest of plasmid pHrB-SE. Fragment 1 was not a good template for sequencing. Fragment 2 was shown to extend rightwards from a Taq I site just outside the start of the 18S gene. The sequence reads through the first Xba I site (see Figure 5.18).

Obviously if all the bands had been eluted from the acrylamide gel, each one could have been cloned and sequenced in turn. However, this would have been a very time-consuming procedure and would have negated the advantage of the speed of the M13 method. I therefore returned to my original strategy of carrying out shotgun cloning.

5.6.b. Cloning Sau 3A fragments into M13 mp 11

The first attempt at shotgun cloning gave a very peculiar result. The Sal I/EcoRI rDNA fragment was digested with restriction enzyme Sau 3A. The resulting mixture of fragments were ligated with mp 11 cut with Bam HI. Subsequent transformation and plating gave the expected white plaque formation. Full dideoxy sequencing was carried out on 9 single-strand templates. Of these, 5 were identical and were found to be a deletion of mp 11. The autoradiograph for one of these templates is shown in Figure 5.6. Reading of the sequence starts as expected just to the left of the primer binding site and proceeds through to the

Figure 5.6 Cloning of Sau3A fragments into mp11 : Identification of a deletion mutant



The very distinctive run of 5 T'S, C, 4 T's aided in the identification of this deletion mutant.

Reading of the sequence starts, as expected, to the left of the primer binding site. However, on reaching the BamHI site (site of expected ligation of a Sau3A fragment), the sequence jumps to nucleotide 6025 in the M13mp11 sequence, giving a deletion of ~ 240 nucleotides.

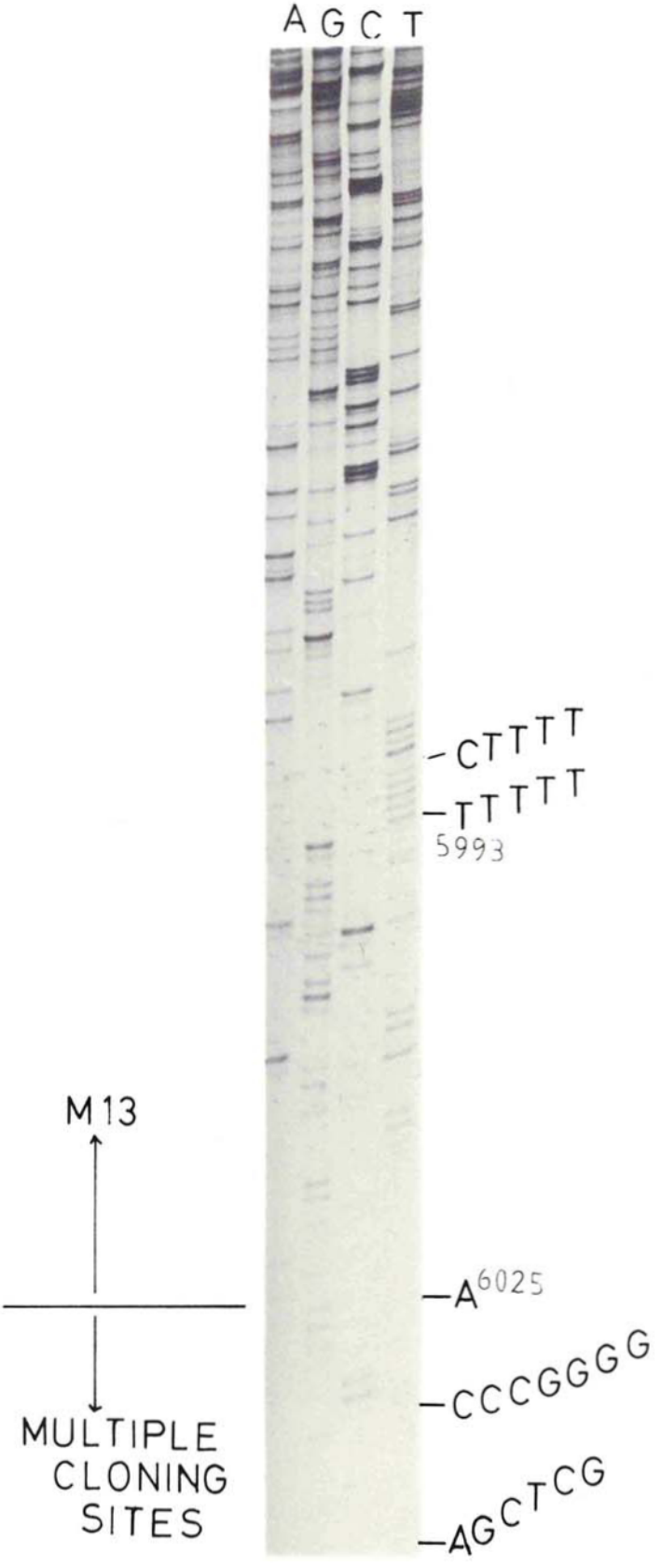

A G C T
-CTTTT
-TTTTT
5993
M13
-A6025
-CCCGGGG
-AGCTCG
MULTIPLE
CLONING
SITES

FIGURE 5.6

Bam HI site within the multiple cloning sites of mp 11. At this point, the sequence jumps to nucleotide 6025 in the M13 sequence. Therefore, there has been a deletion of ~240 nucleotides. This deletion would cause inactivation of the β-galactosidase marker and thus give rise to white plaques (mistaken for recombinant molecules).

The cause of this deletion is unclear. Perhaps the ligation of a particular Sau 3A fragment rendered the mp 11 molecule unstable and led to the subsequent deletion. Whatever the cause, this problem was not encountered on using M13 vector mp 9.

5.6.c. Cloning Sau 3A fragments into M13 mp 9

For this cloning experiment I tried a very simple cloning procedure. The Sal I/EcoRI rDNA fragment was digested with Sau 3A (GATC). M13 vector mp 9 was restricted with Bam HI (GGATCC). After incubating for 90' at 37°C, both digests were heated to 70°C for 15' to inactivate the restriction enzymes before ligation. Ligation was carried out using 200ng of vector and a 3 x molar excess of insert, in a total volume of 10µl, for 3 hours. This mix was then transformed into JM 103 and plated onto minimal agar (Methods, 3.8.c.). The plates were left to incubate at 37°C overnight to allow plaque formation. Inspection of the plates the following morning showed a total of 400 white plaques (recombinant molecules). There was a high background of blue plaques (5 x). This was an expected result as the linear vector had not been gel purified, giving the opportunity for any uncut molecules to transform with high efficiency. Also, the vector was not treated with phosphatase and so retained the ability to self-ligate. However several hundred recombinants was more than ample for subsequent sequence analysis. Therefore by using this much simplified cloning procedure, it is possible to obtain plaques from which to prepare single-strand templates within 24 hours of carrying out the initial vector and insert digests.

Since the ligation mix had contained a whole array of Sau 3A fragments,

Figure 5.7 Cloning of Sau3A fragments : A sample of T-tracking patterns.

The autoradiograph shows the pattern of T residues in templates 25 ——> 36.

By comparing the tracks with each other it is possible to pick out which templates contain the same Sau3A fragments.

25 is the same as 28

26 is unique

27 is unique

29 is the same as 31

34 is unique

36 is unique.

Templates 30, 32, 33 and 35 were not of good quality and so were not considered for futher sequence analysis.

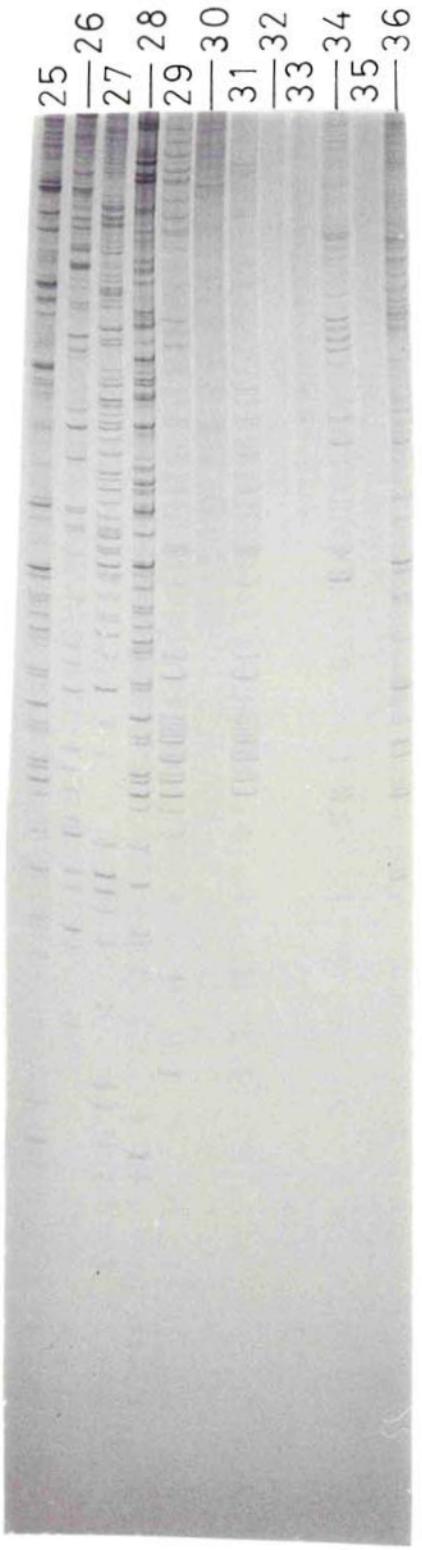


FIGURE 5.7

Templates showing the same T pattern	Template chosen for full sequencing
1, 6, 13,	13
2, 9, 10,	2
3	3
4, 27, 47,	27
5	5
8	8
14	14
15	15
20, 29, 31, 45	20
25, 28	28
26	26
34, 41, 43	41
36, 42	42

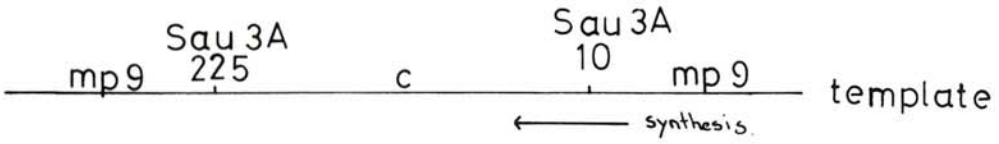
Table 5.2 T-tracking of Sau3A fragments: Identical templates and those chosen for full sequencing

Figure 5.8 Sequencing of Sau3A fragments : An example

The two loadings for template 27 are shown.

The top figure shows how the Sau3A fragment is oriented in M13mp9 to allow sequence determination rightwards from the Sau3A site at position 10, through to the next Sau3A site at position 225.

Nucleotides are numbered according to the complete human 18S rDNA sequence (Figure 5.22).


Sau 3A
225
mp 9
c
Sau 3A
10
mp 9
template
synthesis

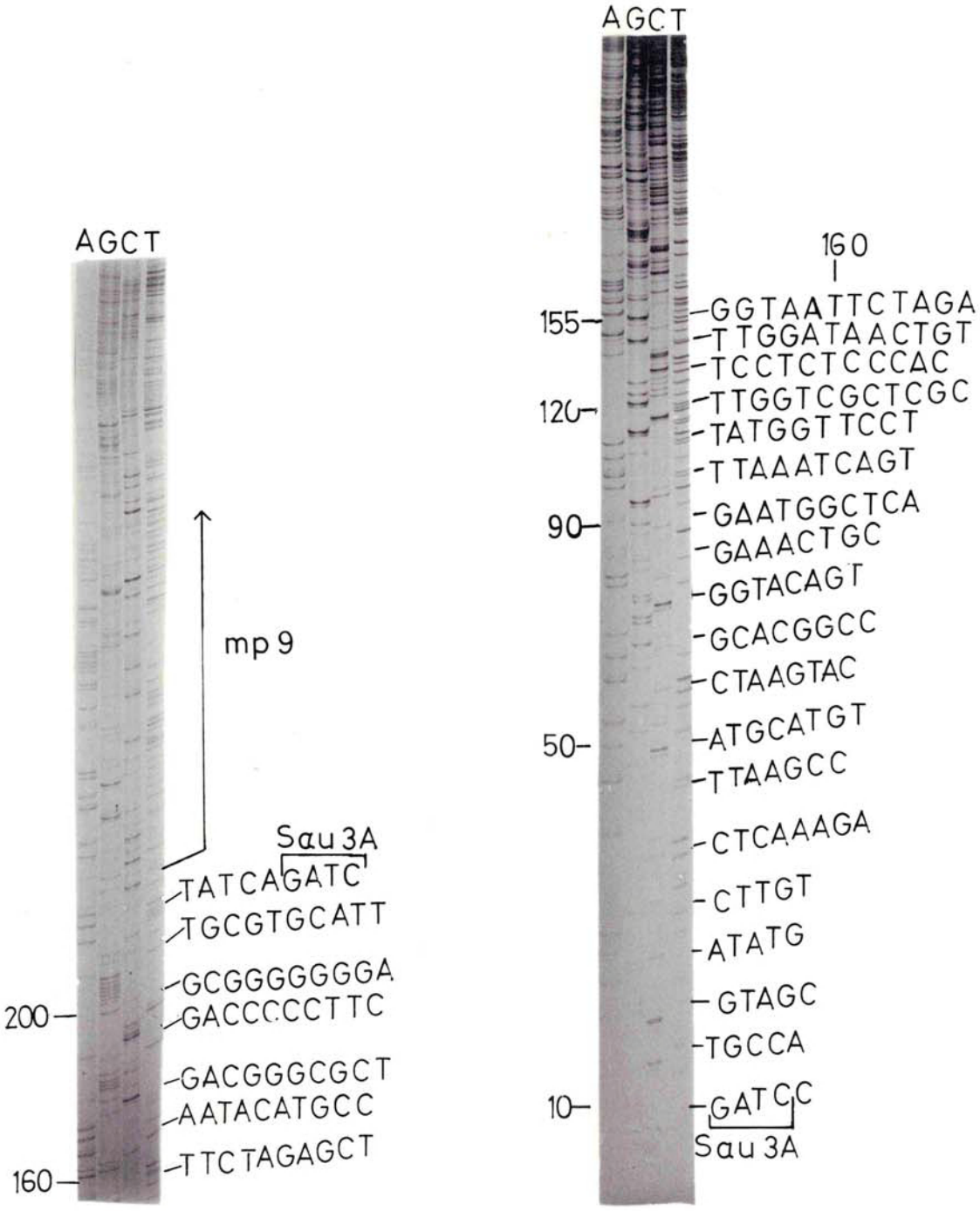

AGCT
mp 9
Sau 3A
TATCAGATC
TGCGTGCATT
GCGGGGGGGA
GACCCCCTTC
GACGGGCGCT
AATACATGCC
TTCTAGAGCT
200
160
AGCT
160
155
120
90
50
10
GGTAATTCTAGA
TTGGATAACTGT
TCCTCTCCCAC
TTGGTCGCTCGC
TATGGTTCCT
TTAAATCAGT
GAATGGCTCA
GAAACTGC
GGTACAGT
GCACGGCC
CTAAGTAC
ATGCATGT
TTAAGCC
CTCAAAGA
CTTGT
ATATG
GTAGC
TGCCA
GATCC
Sau 3A

FIGURE 5.8

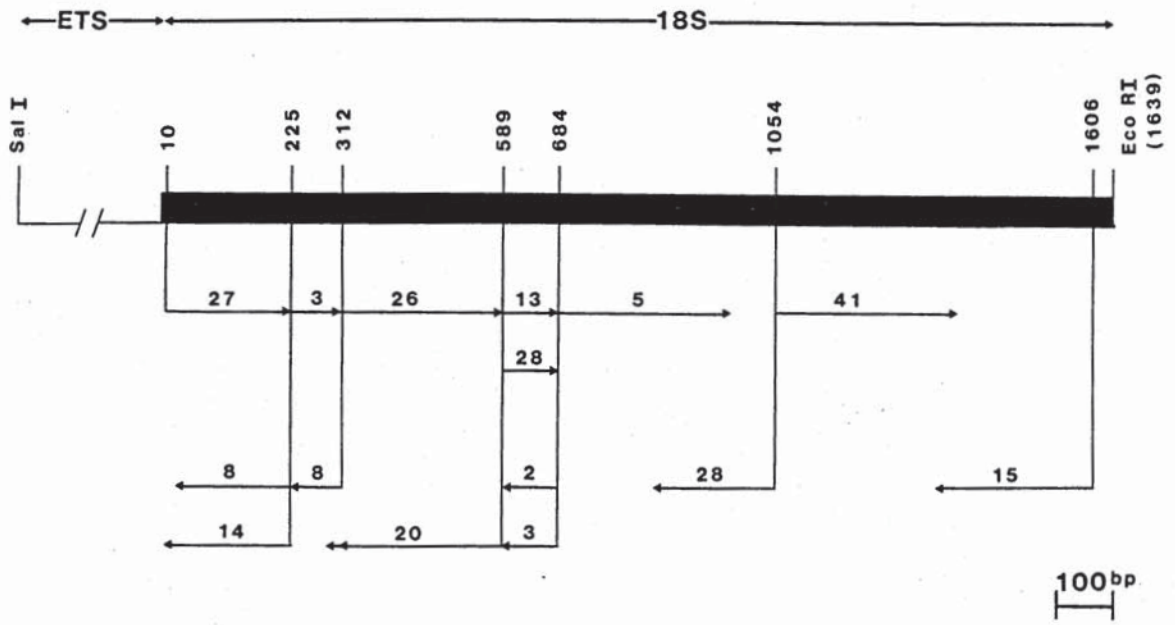


Figure 5.9 Combined sequence data from Sau3A fragments

All Sau3A sites within the 18S gene region are shown. Upper (rightwards) arrows denote sequencing runs on the "S" strand. Lower (leftwards) arrows denote sequencing of the "C" strand. The tips of the arrows indicate the length of sequence obtained from a particular restriction site. Fragments are numbered according to the T-tracking (Table 5.2).

these will each be represented in varying proportions within the total number of white plaques. Each fragment can also be cloned in either orientation. Therefore a sufficient number of plaques have to be picked to give a good representation of the total number of different clones. Therefore, single-strand material was prepared from 48 white plaques (Methods, 3.9.).

To prevent sequencing the same fragment several times, I used the method of T-tracking to identify how many different clones were contained within the 48 templates. (Methods 3.10.d.). Some of the T-tracking patterns are shown in Figure 5.7. Table 5.2. shows the combined results of the T-tracking and which templates were subjected to full dideoxy sequencing. Therefore having carried out 48 T-trackings, the number of full sequencing runs required was narrowed down to only 12.

One of the autoradiographs is shown in Figure 5.8. The complete Sau 3A sequencing map is shown in Figure 5.9. The rat 18S rDNA sequence of Torczynski et al (1983) was used as an aid to aligning the fragments within the human 18S gene.

Figure 5.9 shows that the total possible number of templates were contained within the 48 plaques picked, giving a very good cover of almost the whole of the 18S gene from one enzyme digest alone. This emphasises the speed with which sequence data is obtained using the M13 cloning and sequencing system.

No Sau 3A fragments were picked up from the 5' ETS region.

5.6.d. Cloning and sequencing of Hpa II fragments

The Sal I/EcoRI rDNA fragment was digested with restriction enzyme Hpa II (CCGG). The products of digestion were cloned into mp 9 cut with Acc I (GTCGAC), using the simplified cloning technique. 48 single-stranded templates were prepared and these were subjected to T-tracking. A large proportion of the 48 T-tracks showed uptake of very short inserts. There must be quite a lot of closely spaced Hpa II sites within the Sal I/EcoRI fragment. Table 5.3 shows which

Templates showing the same T pattern	Template chosen for full sequencing
1	1
2, 10, 31	2
3	3
7, 22	7
12	12
13	13
16	16
20, 27	27
26	26
28	28

Table 5.3 T-tracking of HpaII fragments: Identical templates and those chosen for full sequencing

Figure 5.10 Sequencing of HpaII fragments ; An example

The two loadings are shown for template 13.

The top figure shows how the HpaII fragment is oriented in M13mp9 to allow sequence determination rightwards from the HpaII site at position 502. Nucleotides are numbered according to the complete sequence (Figure 5.22).

The insert is too long to allow complete sequence determination right up to the next HpaII site at position 930. Uncertain nucleotides were clarified on the opposite strand.

Code for uncertain nucleotides: H = more than 1 G

D = more than 1 C

^= order unclear

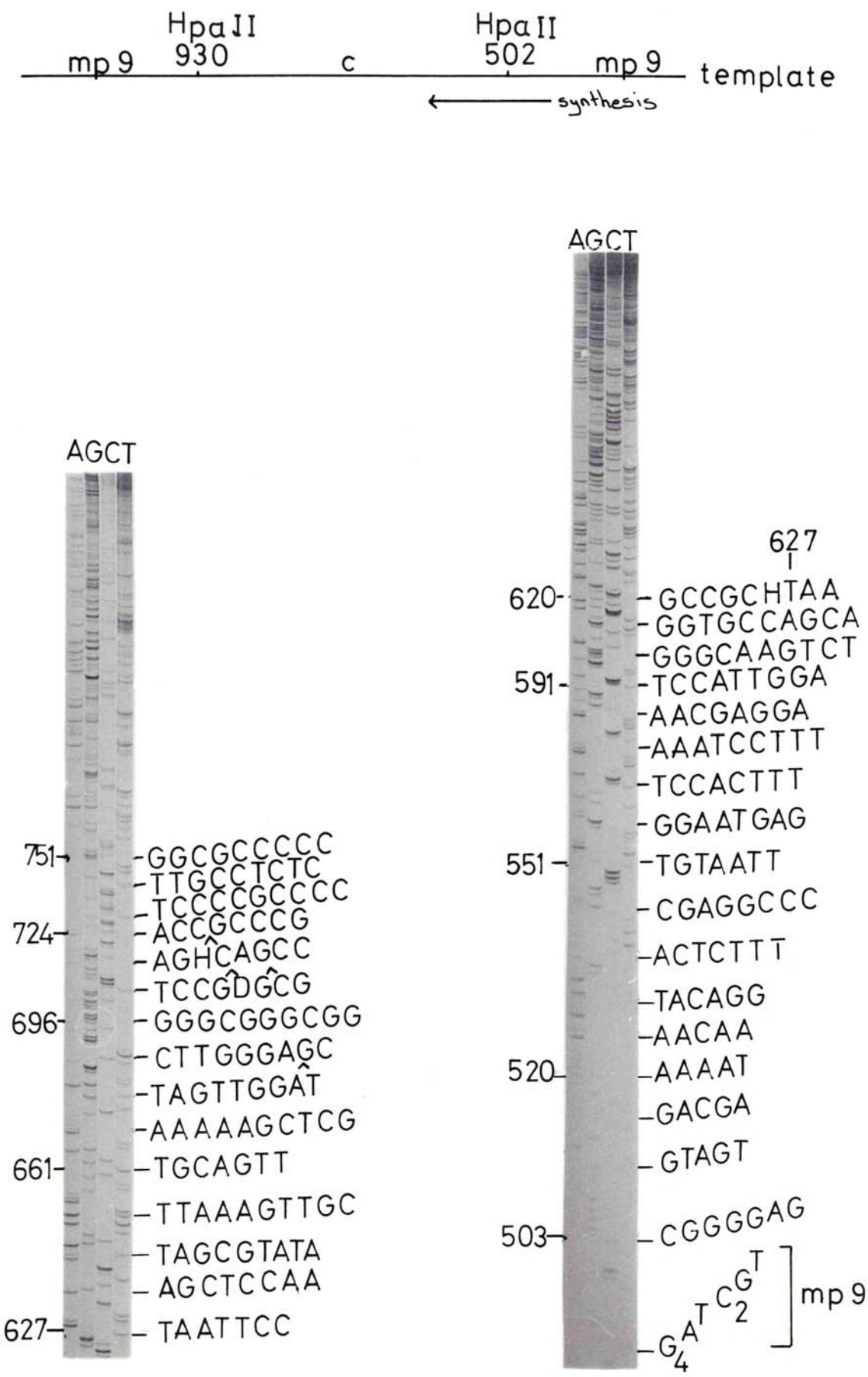

HpaII
930
HpaII
502
mp 9
c
mp 9
template
synthesis
AGCT
751
724
696
661
627
GGCGCCCCC
TTGCCTCTC
TCCCCGCCCC
ACCGCCCG
AGHCAGCC
TCCGDGCG
GGGCGGGCGG
CTTGGGAGC
TAGTTGGAT
AAAAAGCTCG
TGCAGTT
TTAAAGTTGC
TAGCGTATA
AGCTCCAA
TAATTCC
AGCT
620
591
551
520
503
627
GCCGCHTAA
GGTGCCAGCA
GGGCAAGTCT
TCCATTGGA
AACGAGGA
AAATCCTTT
TCCACTTT
GGAATGAG
TGTAATT
CGAGGCCC
ACTCTTT
TACAGG
AACAA
AAAAT
GACGA
GTAGT
CGGGGAG
mp 9

FIGURE 5.10

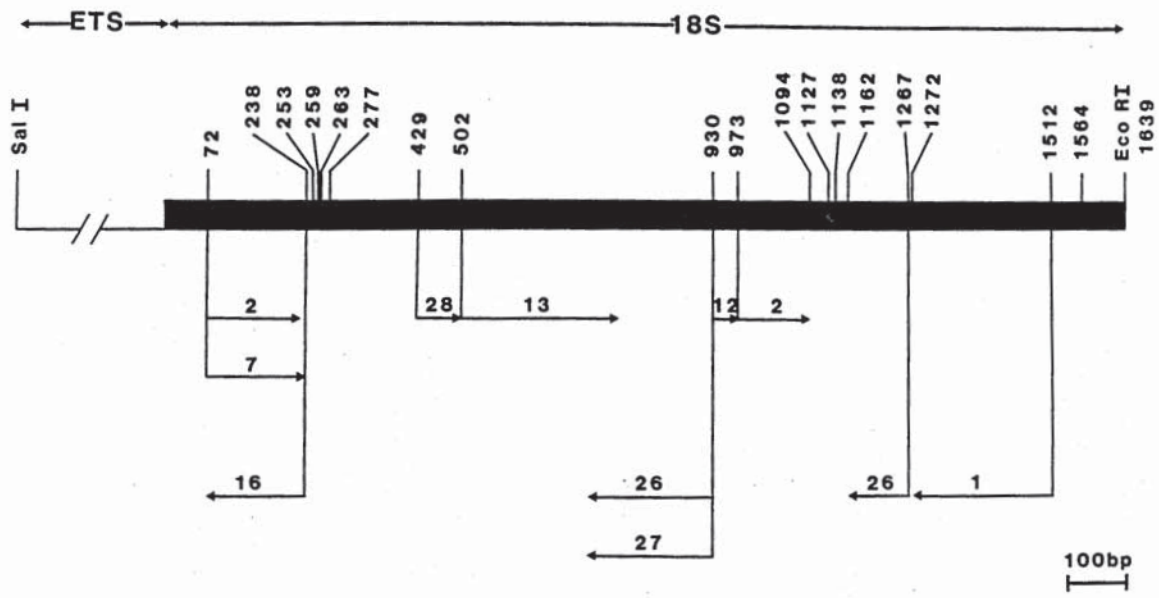


Figure 5.11 Combined sequence data from HpaII fragments

All HpaII sites within the 18S gene region are shown. Upper (rightwards) arrows denote sequencing runs on the "S" strand. Lower (leftwards) arrows denote sequencing of the "C" strand. The tips of the arrows indicate the length of sequence obtained from a particular restriction site. Fragments are numbered according to the T-tracking (Table 5.3).

clones contained larger inserts and those that were subjected to full dideoxy sequencing. The resulting autoradiograph for one of the templates is shown in Figure 5.10. The complete Hpa II map is shown in Figure 5.11. Template 3 could not be fitted into the 18S sequence. This fragment is obviously contained within the ETS region of the Sal I/EcoRI fragment. I have marked on the map the positions of the closely spaced Hpa II sites which will have contributed to the large number of templates having short inserts. I suspect that the ETS region also gave rise to a large number of clones containing short fragments, as this region is known to be very GC rich and so would be expected to possess a high percentage of Hpa II sites (CCGG).

5.6.e. Cloning and sequencing of Taq I fragments

Since I already had the sequence of the Taq I fragment extending rightwards from just outside the start of the 18S gene (Figure 5.18), I wanted to use the Xba I/EcoRI fragment for subsequent analysis with Taq I. This fragment was prepared from an Xba I/EcoRI double digest of the parent plasmid pHrB-SE (Figure 5.12).

The XbaI/EcoRI fragment was digested with restriction enzyme Taq I. Since Taq I digestion is carried out at 60°C, the enzyme could not be heat inactivated. Phenol extraction was carried out to remove the enzyme prior to ligation. Three ligations were carried out. One third of the Taq I digest was ligated into mp 8 cut with Acc I, which would allow the uptake of fragments bound by a Taq I site at both ends. The remaining digest mix was split equally between mp 8 and mp 9 doubly digested with Acc I and EcoRI. In this way it was possible to clone the 3' TaqI/EcoRI fragment in both orientations, to allow sequence determination right up to and leftwards from the EcoRI site. Transformation and plating gave rise to the expected white plaque formation. 40 single-strand templates were prepared from the ligation into mp 8 cut with Acc I. In the case of ligation into doubly cut vectors, all plaques should contain the same insert and so I picked only 3 plaques from each of mp 8 and mp 9.

Figure 5.12 Digestion of pHrB-SE with XbaI and EcoRI

The top figure shows the positions of the XbaI and EcoRI sites in clone pHrB-SE.

The products of digestion with XbaI and EcoRI are shown below. The fragments are numbered in order of decreasing size.

Fragment 2 was used for subsequent restriction with TaqI.

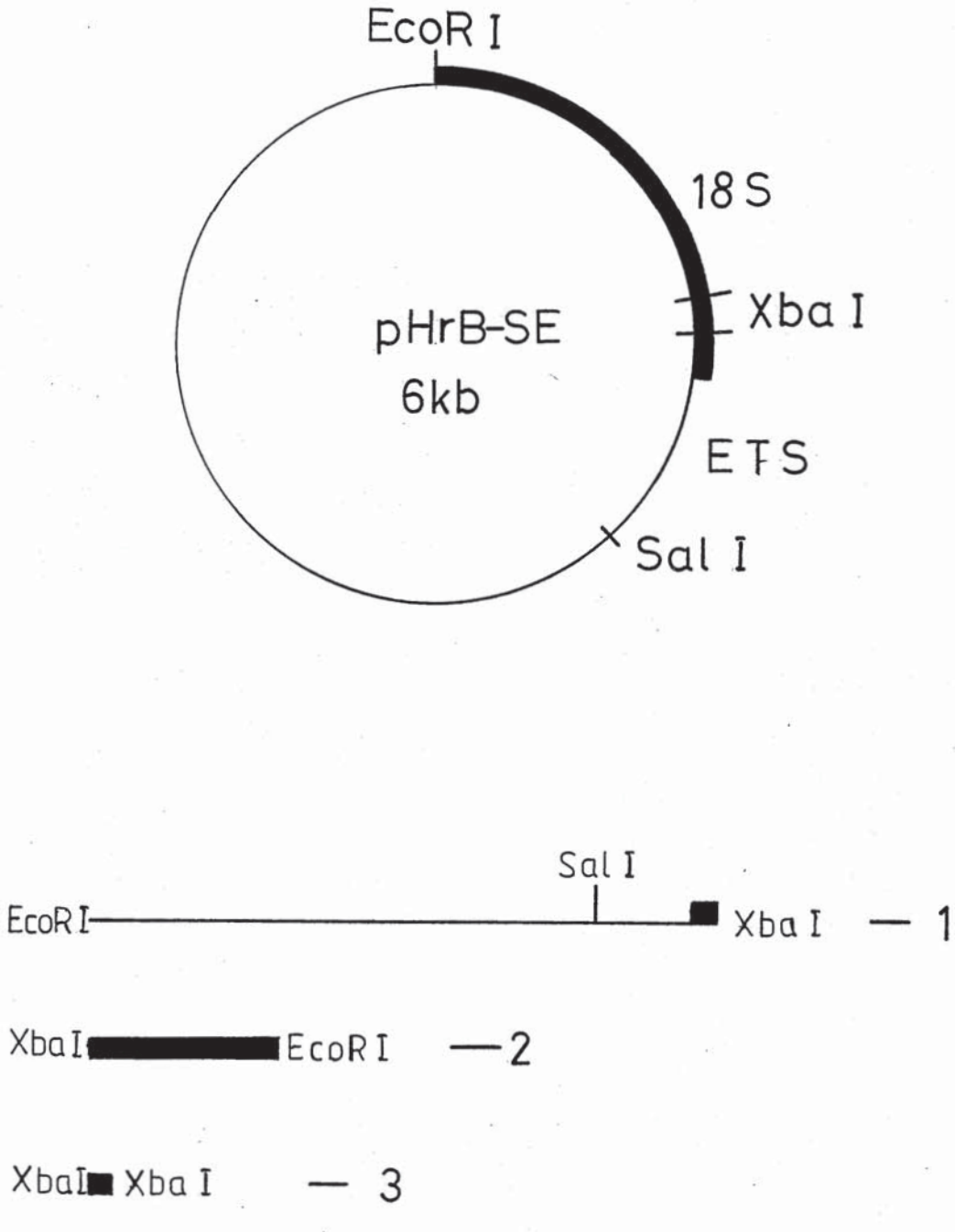

EcoR I
18 S
Xba I
pHrB-SE
6kb
ETS
Sal I
Sal I
EcoR I
Xba I — 1
Xba I
EcoR I — 2
Xba I
Xba I — 3

FIGURE 5.12

Templates showing the same T pattern	Template chosen for full sequencing
1	1
3	3
4, 6, 12	4
7	7
8, 39	8
9, 18	9
10, 17, 19	10
11	11
14	14
20, 37	20
21	21
23	23
24, 28	24
36	36
41, 42, 43	41, 42
44, 45, 46	44, 45

Table 5.4 T-tracking of TaqI fragments: Identical templates and those chosen for full sequencing

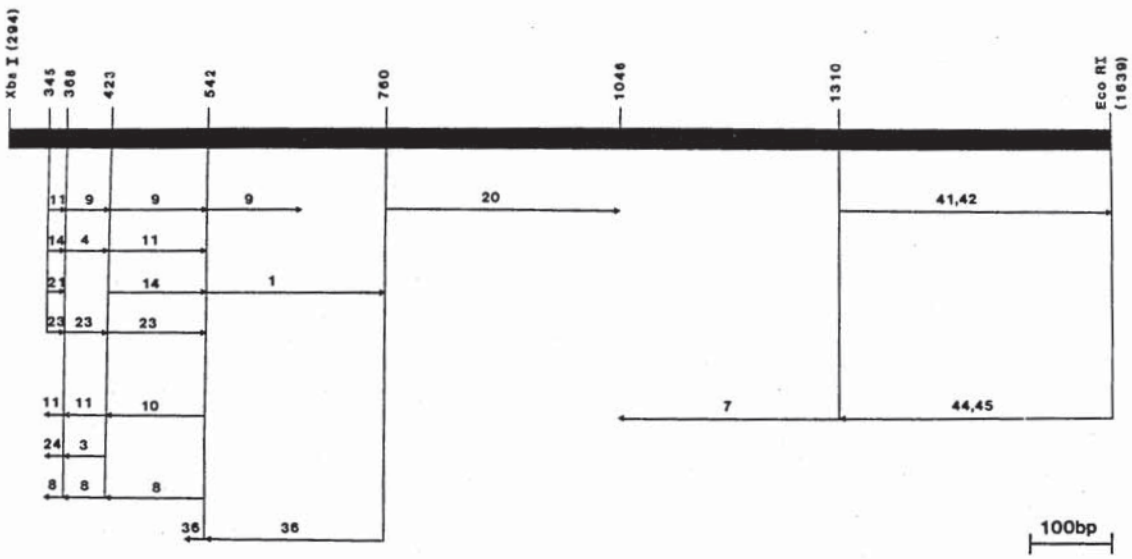


Figure 5.13 Combined sequence data from TaqI fragments

All TaqI sites within the 18S XbaI/EcoRI fragment are shown. Upper (rightwards) arrows denote sequencing runs on the "S" strand. Lower (leftwards) arrows denote sequencing runs on the "C" strand. The tips of the arrows denote the length of sequence obtained from a particular restriction site. Fragments are numbered according to the T-tracking (Table 5.4).

Figure 5.14 Sequencing of the TaqI/EcoRI fragment on the RNA-like strand

The upper figure shows the orientation of the TaqI/EcoRI fragment in vector M13mp8 cut with AccI and EcoRI.

The resulting autoradiographs are shown below. Nucleotides are numbered according to the complete sequence (Figure 5.22).

Uncertain nucleotides were clarified by sequencing the opposite strand.

Code for uncertain nucleotides: H = more than 1 G

D = more than 1 C

^= order unclear

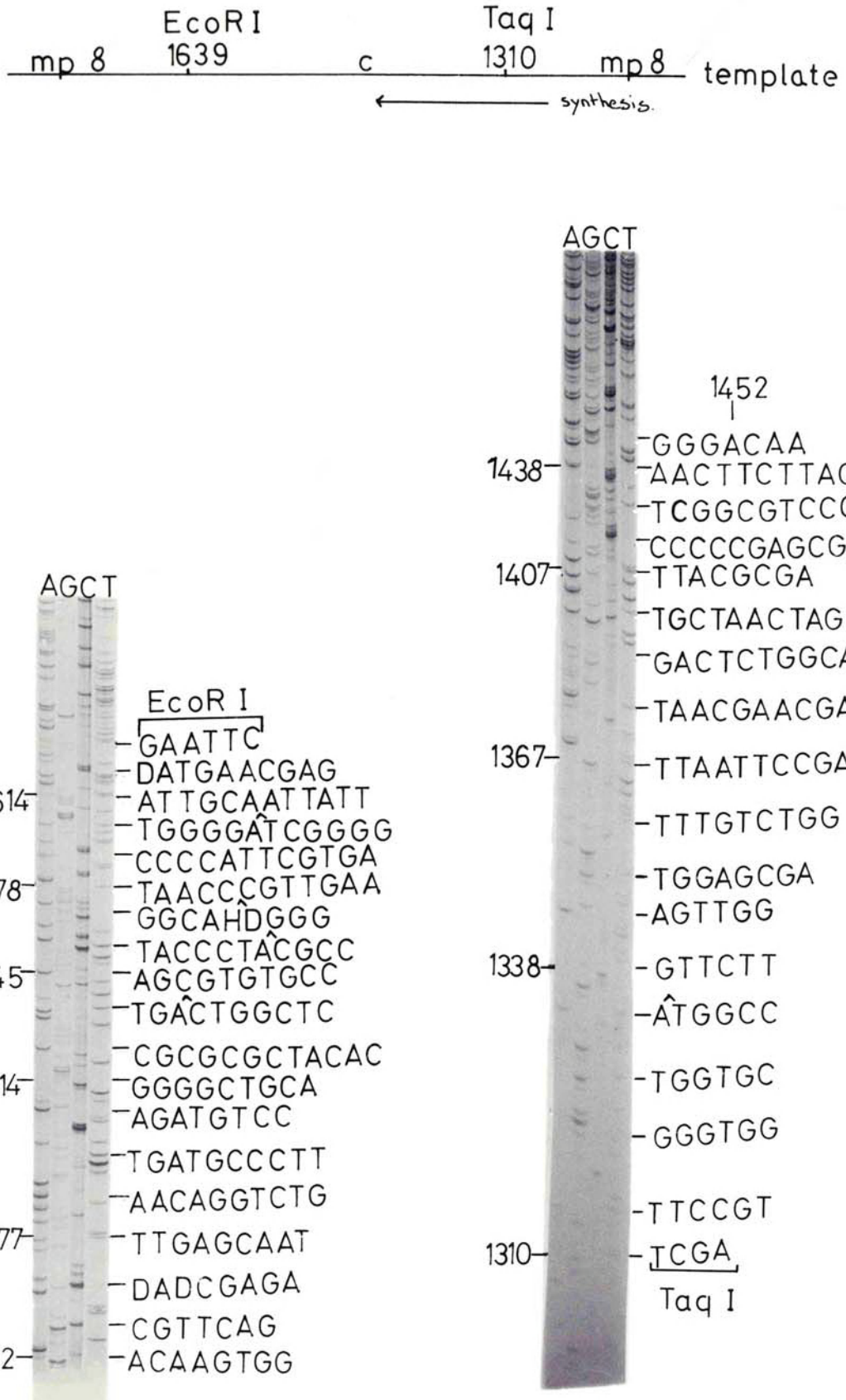

EcoR I
Taq I
mp 8
1639
c
1310
mp 8
template
synthesis.
AGCT
EcoR I
GAATTC
DATGAACGAG
1614
ATTGCAATTATT
TGGGGATCGGGG
CCCCATTCGTGA
1578
TAACCCGTTGAA
GGCAHDGGG
TACCCTACGCC
1545
AGCGTGTGCC
TGACTGGCTC
CGCGCGCTACAC
1514
GGGGCTGCA
AGATGTCC
TGATGCCCTT
AACAGGTCTG
1477
TTGAGCAAT
DADCGAGA
CGTTCAG
1452
ACAAGTGG
AGCT
1452
GGGACAA
1438
AACTTCTTAGA
TCGGCGTCCCCC
CCCCCGAGCGG
1407
TTACGCGA
TGCTAACTAG
GACTCTGGCA
TAACGAACGA
1367
TTAATTCCGA
TTTGTCTGG
TGGAGCGA
AGTTGG
1338
GTTCTT
ATGGCC
TGGTGC
GGGTGG
TTCCGT
1310
TCGA
Taq I

FIGURE 5.14

T-tracking was carried out on all 46 single-strand templates. Table 5.4 shows which templates were subjected to full dideoxy sequencing. The resulting map is shown in Figure 5.13. The map shows that a large number of templates contained the three Taq I fragments at the 5' end of the Xba I/EcoRI fragment joined up in various orientations and combinations. The rightwards sequencing run for the 3' Taq I/EcoRI fragment is shown in Figure 5.14. Only two possible templates were not picked up (both leftwards and rightwards from Taq I (1046)).

5.7. Sequencing of the 3' end of the 18S Gene

The extreme 3' end of the human 18S rRNA gene is contained in pHr A (Figure 5.1.). I wanted to carry out digests on the whole plasmid that would give this remaining 18S rDNA region contained in a fragment of a suitable size for cloning. It is widely believed to be preferential to have fragments of less than 1000 bp for successful cloning into M13. Restriction mapping of this large EcoRI A fragment had been carried out by Erickson et. al. (1981). It contains single sites for restriction enzymes Xba I and Kpn I at approximately 0.2kb and 0.78kb from the 18S EcoRI site respectively (Figure 5.15 (a)). By cloning the EcoRI/XbaI fragment, it would be possible to sequence the whole fragment from both ends. The Xba I site is still within the 18S gene, however by cloning the Xba I/KpnI fragment it would be possible to sequence from the XbaI site through the end of the 18S gene into ITS 1. Cloning of the EcoRI/KpnI fragment would allow sequencing through the XbaI site. Vector pair mp 18 / mp 19 (Figure 2.3) contain restriction sites for the three enzymes EcoRI, Xba I and Kpn I and so can be used to clone the three ribosomal fragments in both orientations.

2µg aliquots of pHrA were restricted with the three double digests, and 0.2µg was run on a 1% agarose mini-gel to check the presence and relative positions of these sites in pHr A (Figure 5.15(b)). The gel showed the expected result. The plasmid digests were heat-inactivated before subsequent ligation.

Figure 5.15 Restriction of pHrA with EcoRI, XbaI and Kpn I

(a) Shows the sites for XbaI and KpnI within the EcoRI A fragment, according to Erickson et al (1978)

(b) 2µg lots of pHrA were digested with EcoRI + XbaI, EcoRI + Kpn I and XbaI + Kpn I. 1/10 of the digests were run on a 1% agarose mini gel.

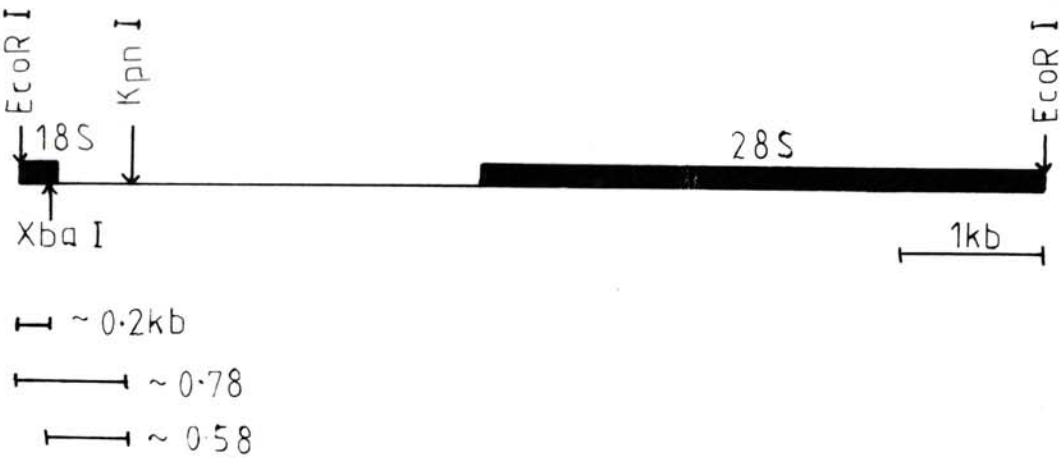
Lane 1 0.2µg of pHrA cut with EcoRI + XbaI

Lane 2 0.2µg of pHrA cut with EcoRI + KpnI

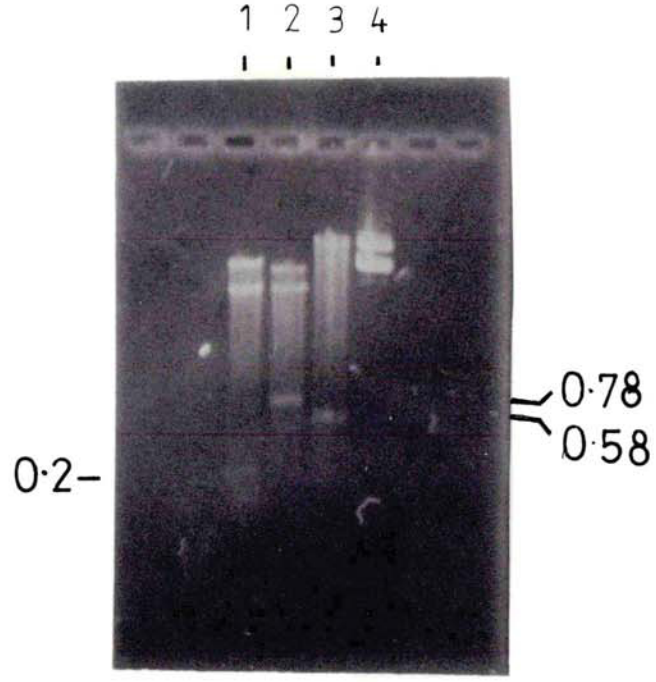
Lane 3 0.2µg of pHrA cut with XbaI + KpnI

Lane 4 Uncut pHrA.

In lanes 1, 2 and 3 the appropriate small bands are shown, giving confirmation of the map of Erickson et al (1978).



(a)



(b)

FIGURE 5.15

Figure 5.16 Cloning into mp18/mp19 : Single stranded templates and sequence determination

(a) Cloning and sequencing of the EcoRI/XbaI fragment.

(b) Cloning and sequencing of the EcoRI/KpnI fragment.

(c) Cloning and sequencing of the XbaI/KpnI fragment.

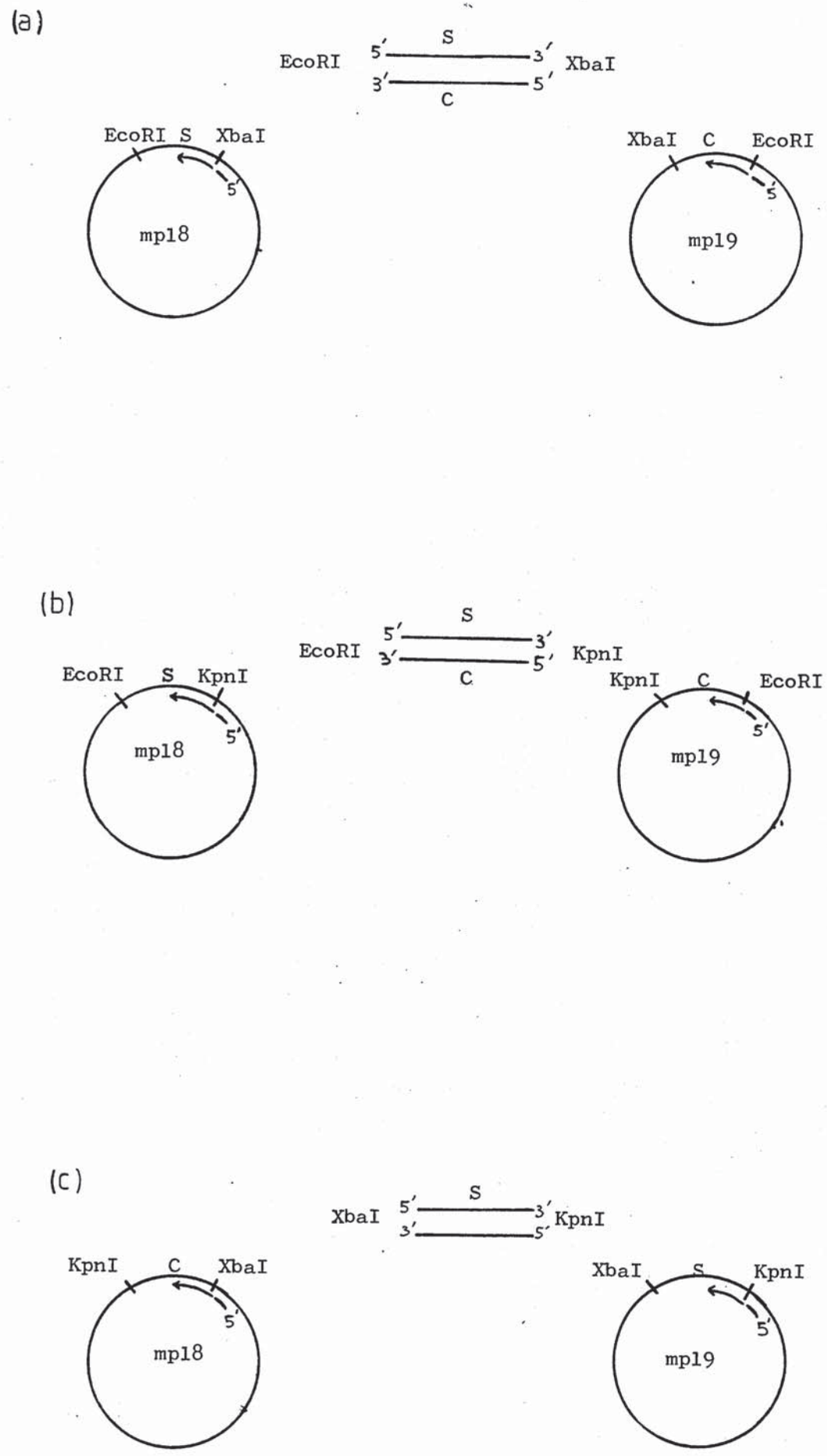

(a)
EcoRI
S
5′
3′
XbaI
3′
5′
C
EcoRI S XbaI
5′
mp18
XbaI C EcoRI
5′
mp19
(b)
S
EcoRI
5′
3′
KpnI
3′
5′
C
EcoRI S KpnI
5′
mp18
KpnI C EcoRI
5′
mp19
(c)
XbaI
S
5′
3′
KpnI
3′
5′
KpnI C XbaI
5′
mp18
XbaI S KpnI
5′
mp19

FIGURE 5.16

Figure 5.17 Sequencing of the 3' EcoRI/XbaI fragment on the complementary strand

The upper figure shows the orientation of the EcoRI/XbaI fragment in vector M13mp18.

The resulting autoradiographs are shown below. Nucleotides are numbered according to the complete sequence (Figure 5.22).

Code for uncertain nucleotides: H = more than 1 G

D = more than 1 C

^= order unclear

? = presence of a nucleotide uncertain

B = more than 1 A

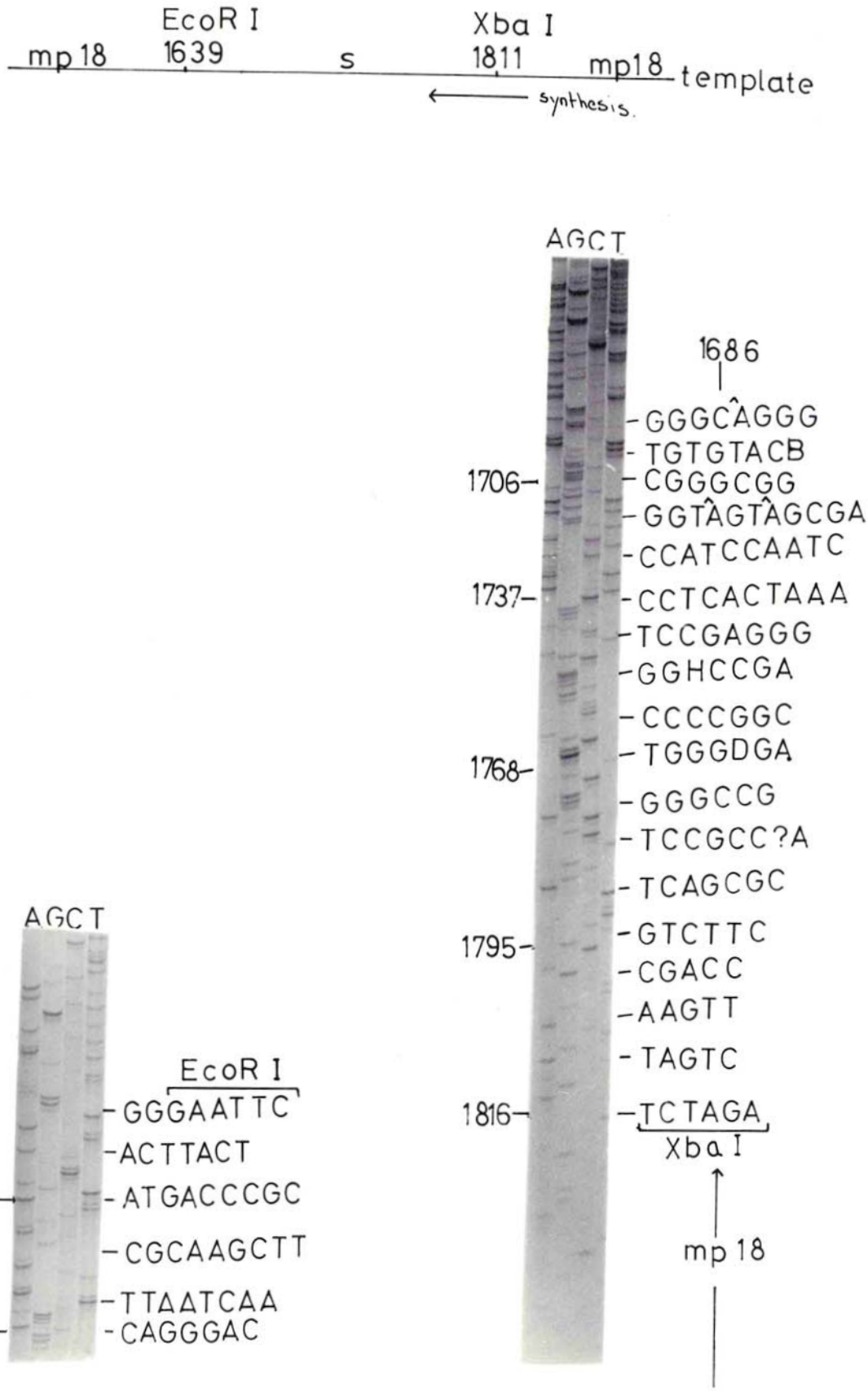

EcoR I
1639
mp 18
s
Xba I
1811
mp18
template
synthesis
AGCT
1686
GGGCÂGGG
TGTGTACB
1706
CGGGCGG
GGTÂGTÂGCGA
CCATCCAATC
1737
CCTCACTAAA
TCCGAGGG
GGHCCGA
CCCCGGC
1768
TGGGDGA
GGGCCG
TCCGCC?A
TCAGCGC
1795
GTCTTC
CGACC
AAGTT
TAGTC
1816
TCTAGA
Xba I
mp 18
AGCT
EcoR I
GGGAATTC
ACTTACT
1662
ATGACCCGC
CGCAAGCTT
TTAATCAA
1686
CAGGGAC

FIGURE 5.17

Six ligations were carried out, with each plasmid digest being split between vector pair mp 18/mp 19 cut with the appropriate pair of enzymes. There was no need to purify the ribosomal fragments of interest before ligation as the other fragments generated in the plasmid digests would be too large to clone. The ligation mixes were transformed into JM 103 and plated onto minimal agar. Figure 5.16 shows the expected single-strand templates and therefore the direction of sequence determination. Figure 5.17 shows the resulting autoradiograph giving the sequence of the EcoRI/XbaI fragment leftwards from the Xba I site.

The ligations involving KpnI gave problems in the identification of true recombinants. Each of the 4 transformations gave rise to white plaques, which on subsequent sequence analysis were shown to be vector molecules with one or two nucleotides deleted. This will have resulted in an alteration of the reading frame and therefore inactivation of the β-galactosidase marker. I suspect this was due to the presence of contaminating exonuclease activity in Kpn I. I did eventually manage to pick plaques which gave the sequence rightwards from the EcoRI site through the XbaI site and rightwards from the Xba I site through the end of the 18S gene. No sequence data was obtained leftwards from the Kpn I site.

5.8. The Complete Nucleotide Sequence of the Human 18S rRNA Gene

The complete sequencing map for the determination of the 18S nucleotide sequence for human is shown in Figure 5.18. A number of the sequencing gels are shown in the text (Figures 5.8, 5.10, 5.14, 5.17). Inspection of these autoradiographs show characteristic features of dideoxy sequencing. In the case of doublets, the upper C is always more intense than the lower C, the upper G is often more intense than the lower G (especially when the doublet is preceded by a T), the upper A is often less intense than the lower A. Remembering these few rules aids in the interpretation of dideoxy sequencing gels.

All of the sequence was determined on both strands, with the exception of the

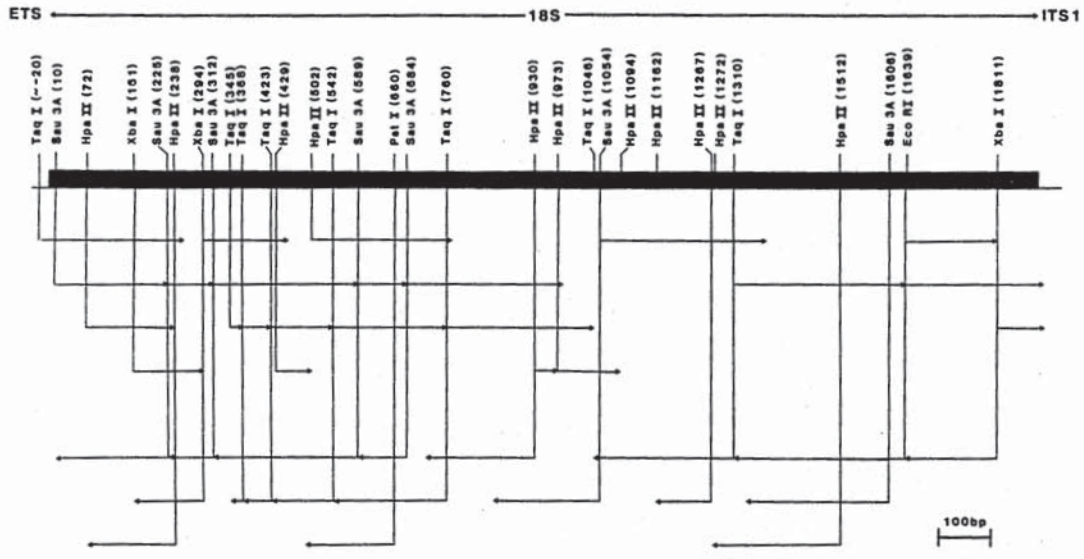


Figure 5.18 Complete sequencing map for human 18S rDNA

Only those restriction sites at the start or end points of sequencing runs are shown. The arrows show the direction and length of sequence obtained from a particular restriction site. The maps of the individual enzymes show the numbering of the single-stranded templates and the number of times each particular fragment was sequenced (Figures 5.9, 5.11 and 5.13).

5' 9 nucleotides and 3' 54 nucleotides which were determined only on the RNA-like strand. The 3' EcoRI site was read up to and from in both directions, but not through, as this point was an initial cloning site in plasmids pHrB-SE and pHrA. However, this site has since been sequenced through in a human rDNA clone containing this EcoRI site (prepared and sequenced by B.E.H. Maden).

As with Maxam-Gilbert sequencing, dideoxy sequencing gels can show regions of unevenly spaced bands, or in the most severe cases, show several bands compressed together making it impossible to interpret these regions of the gel. These compression effects tend to occur in G/C rich areas in the sequence which can form local secondary structures. However, secondary structure effects do not tend to occur at exactly the same position on both strands and so by comparing the complementary gels, the complete sequence can usually be deduced. Therefore any doubtful readings on one strand were always checked on the opposite strand. There was one compression effect where one nucleotide remained uncertain on both strands. This region is shown in three gels (two on the rightwards strand and one on the leftwards strand) in Figure 5.19. Rightwards gels suggest a TT doublet at nucleotides 1775-1776. However sequencing of the leftwards strand could be interpreted as 1 or 2 A residues. This uncertainty was resolved by restriction. The presence of only one T residue would give a recognition site for BstNI (CCTGG) at this point. The small EcoRI/XbaI fragment (identified in Figure 5.15) was prepared from plasmid pHrA. This fragment was shown to be cut into two fragments of very different size on digestion with BstNI (Figure 5.20). Therefore there is only a single T residue at nucleotide 1775 on the RNA-like strand.

Since I had not sequenced the ETS region contained in plasmid pHrB-SE, I wanted to estimate the size of this region (Figure 5.21). pHrB-SE was digested with Sal I and Xba I and the products of digestion separated on a 4% acrylamide gel. By running an AluI digest of pBR322 alongside the digest of pHrB-SE, the SalI/XbaI band is shown to run slightly faster than a 910 bp AluI fragment. I would estimate the Sal I/XbaI fragment to have a size of 820-840bp. From the sequence

Figure 5.19 Compression effect in region 1770-1780

Three gels covering this region are shown.(i) and (ii) give the sequence of this region on the RNA-like strand and (iii) covers the complementary strand.

By comparing the three gels it is possible to determine the nucleotide sequence of this region unambiguously except for 1 nucleotide (shown at the bottom of the figure). The rightwards gels show two closely spaced T residues at nucleotides 1775-1776. The leftwards gel is ambiguous at this point. Nucleotides are numbered according to the complete sequence (Figure 5.22).

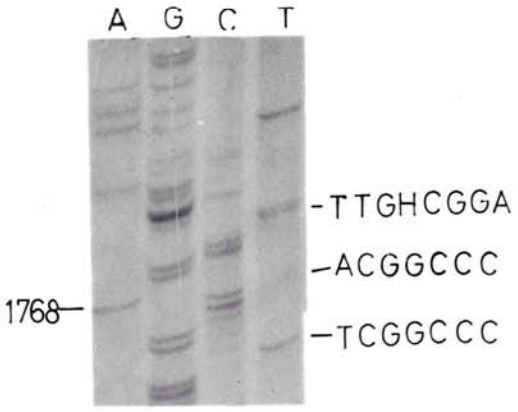
Code for uncertain nucleotides: H = more than 1 G

D = more than 1 C

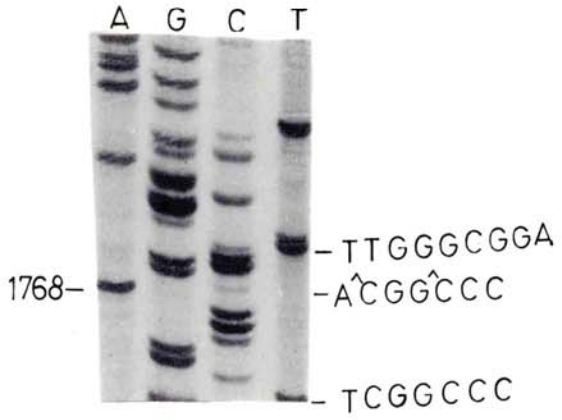
^ = order unclear

? = presence of a nucleotide uncertain.

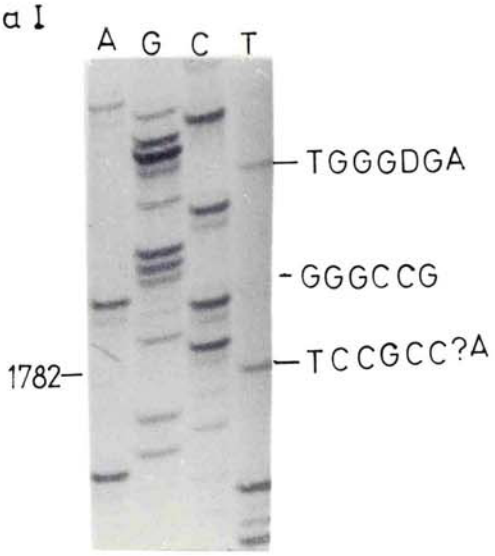
(i) EcoR I ⟶ Xba I



(ii) EcoR I ⟶ Kpn I



(iii) EcoR I ⟵ Xba I



1771 GCCCT(T)GGGCGGA

FIGURE 5.19

Figure 5.20 Identification of a BstNI site at nucleotide 1773

The indicated EcoRI/XbaI fragment was digested with BstNI. The possible restriction products are shown in the top figure.

A HaeIII size marker digest was carried out on pBR322 (sizes in bp): 587, 540, 504, 458, 434, 267, 234, 213, 192, 184, 124, 123, 104, 89, 80, 64, 57, 51, 21, 18, 11, 8.

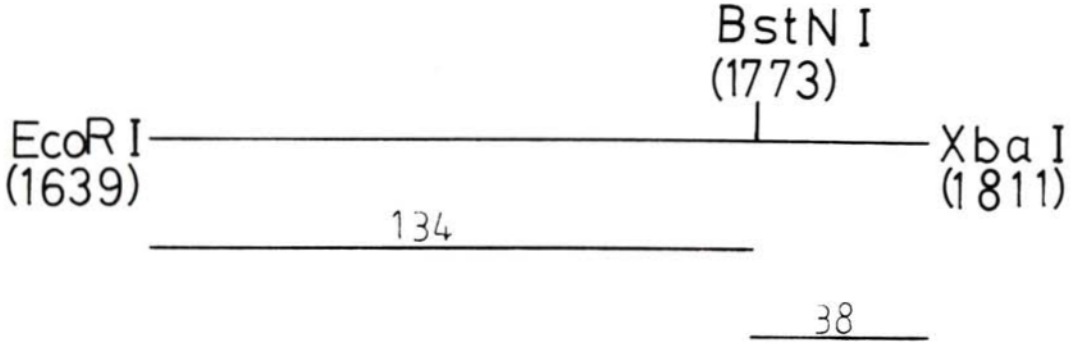
Fragments were separated on a 4% acrylamide gel.

Lane (1) HaeIII digest of pBR322

(2) BstNI digest of EcoRI/XbaI fragment

(3) Uncut EcoRI/XbaI fragment.

The result shows the presence of a BstNI site at position 1773 (CCTGG). Therefore there is only 1 T residue in this region and not a TT doublet as suggested by sequencing of the RNA-like strand.



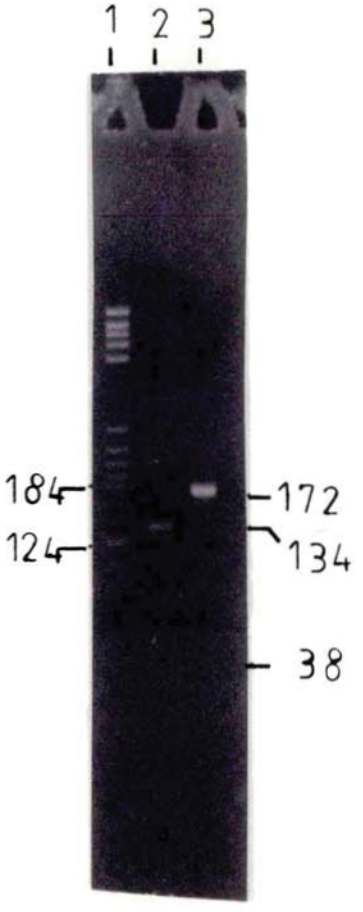


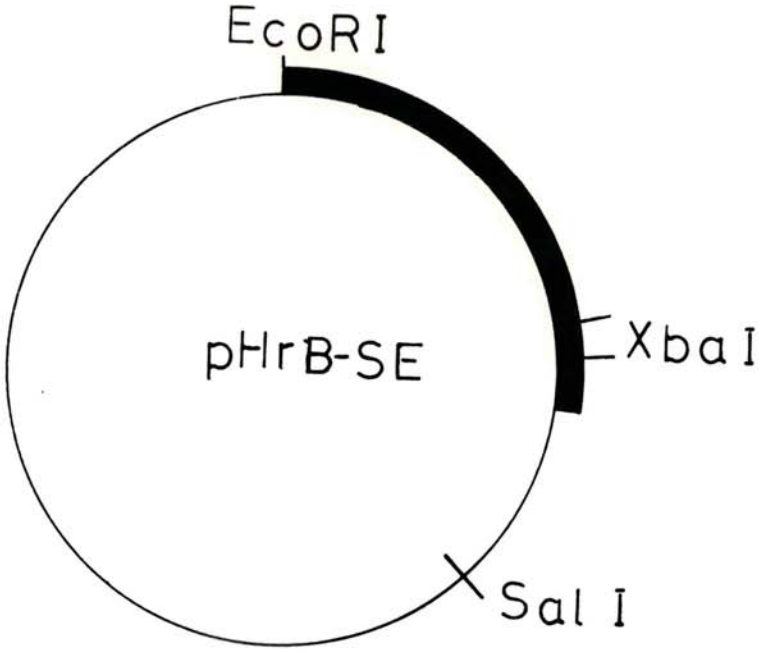
FIGURE 5.20

Figure 5.21 Sizing of the ETS region in pHrB-SE

The positions of the restriction sites for XbaI in plasmid pHrB-SE are shown in the top figure.

0.5µg of pHrB-SE was digested with SalI and XbaI and the products of digestion separated on a 4% polyacrylamide gel. An AluI digest of pBR322 provided suitable size markers.

Band 2 contains the DNA fragment bound by SalI and the first XbaI site in the 18S gene. By comparison with the size markers this fragment has a size of ~820-840 bp. The XbaI site is known to be at nucleotide 161 in the 18S gene. Therefore the ETS region has a size of ~660-680bp.



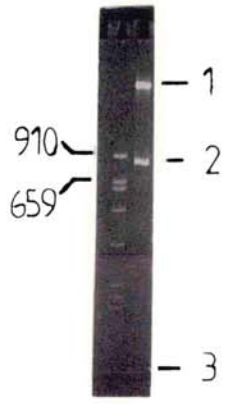


FIGURE 5.21

data, the Xba I site is known to be at position 161 in the 18S sequence. Therefore the ETS region bound by the Sal I site has a size of 660-680bp. This fits in with an earlier observation:- During the sequence determination of the Sau3A fragments of the SalI/EcoRI fragment, no sequence data was obtained for the ETS region (section 5.6.c.). The first Sau3A site in the 18S gene is at nucleotide 10. If no Sau3A sites are present in the ETS region, digestion of the SalI/EcoRI fragment with Sau3A would give a large fragment of 670-690 bp. Looking at Figure 5.5, digestion of the Sal I/EcoRI fragment with Sau3A does indeed give a fragment of ~680 bp. Furthermore, this fragment is missing in the corresponding digest of pHrB-SE and so must have the SalI site at one end. Therefore, I would conclude that there are no Sau3A sites in the human ETS region between the Sal I site and the start of the 18S gene.

Compilation of the total sequence data gives the complete nucleotide sequence of human 18S rDNA bounded by a small amount of ETS and ITS 1 (Figure 5.22). This human gene is 1870 bp long. Comparison of this sequence with other published data for the small subunit rRNA's is discussed in Chapter 6.

Figure 5.22 Complete nucleotide sequence of human 18S rDNA

The figure shows the complete 18S rDNA sequence of human, determined from clones pHrB-SE and pHrA. The strand synonomous to RNA is shown. Clone pHrB-SE provided the sequence from the start of the 18S gene through to the EcoRI site (GAATTC) at nucleotide 1639. The remaining 3' region was determined from clone pHrA. A small amount of sequence data from the adjoining ETS and ITS 1 are also included.

```
ETS                                                AG GTCGCGCGCT CTACCTTACC  -1

18S TACCTGGTTG ATCCTGCCAG TAGCATATGC TTGTCTCAAA GATTAAGCCA TGCATGTCTA   60

    AGTACGCACG GCCGGTACAG TGAAACTGCG AATGGCTCAT TAAATCAGTT ATGGTTCCTT  120

    TGGTCGCTCG CTCCTCTCCC ACTTGGATAA CTGTGGTAAT TCTAGAGCTA ATACATGCCG  180

    ACGGGCGCTG ACCCCCTTCG CGGGGGGGAT GCGTGCATTT ATCAGATCAA AACCAACCCG  240

    GTCAGCCCCT CTCCGGCCCC GGCCGGGGGG CGGGCGCCGG CGGCTTTGGT GACTCTAGAT  300

    AACCTCGGGC CGATCGCACG CCCCCCGTGG CGGCGACGAC CCATTCGAAC GTCTGCCCTA  360

    TCAACTTTCG ATGGTAGTCG CCGTGCCTAC CATGGTGACC ACGGGTGACG GGGAATCAGG  420

    GTTCGATTCC GGAGAGGGAG CCTGAGAAAC GGCTACCACA TCCAAGGAAG GCAGCAGGCG  480

    CGCAAATTAC CCACTCCCGA CCCGGGGAGG TAGTGACGAA AAATAACAAT ACAGGACTCT  540

    TTCGAGGCCC TGTAATTGGA ATGAGTCCAC TTTAAATCCT TTAACGAGGA TCCATTGGAG  600

    GGCAAGTCTG GTGCCAGCAG CCGCGGTAAT TCCAGCTCCA ATAGCGTATA TTAAAGTTGC  660

    TGCAGTTAAA AAGCTCGTAG TTGGATCTTG GGAGCGGGCG GGCGGTCCGC CGCGAGGCGA  720

    GCCACCGCCC GTCCCCGCCC CTTGCCTCTC GGCGCCCCCT CGATGCTCTT AGCTGAGTGT  780

    CCCGCGGGGC CCGAAGCGTT TACTTTGAAA AAATTAGAGT GTTCAAAGCA GGCCCGAGCC  840

    GCCTGGATAC CGCAGCTAGG AATAATGGAA TAGGACCGCG GTTCTATTTT GTTGGTTTTC  900

    GGAACTGAGG CCATGATTAA GAGGGACGGC CGGGGGCATT CGTATTGCGC CGCTAGAGGT  960

    GAAATTCTTG GACCGGCGCA AGACGGACCA GAGCGAAAGC ATTTGCCAAG AATGTTTTCA 1020

    TTAATCAAGA ACGAAAGTCG GAGGTTCGAA GACGATCAGA TACCGTCGTA GTTCCGACCA 1080

    TAAACGATGC CGACCGGCGA TGCGGCGGCG TTATTCCCAT GACCCGCCGG GCAGCTTCCG 1140

    GGAAACCAAA GTCTTTGGGT TCCGGGGGGA GTATGGTTGC AAAGCTGAAA CTTAAAGGAA 1200

    TTGACGGAAG GGCACCACCA GGAGTGGAGC CTGCGGCTTA ATTTGACTCA ACACGGGAAA 1260

    CCTCACCCGG CCCGGACACG GACAGGATTG ACAGATTGAT AGCTCTTTCT CGATTCCGTG 1320

    GGTGGTGGTG CATGGCCGTT CTTAGTTGGT GGAGCGATTT GTCTGGTTAA TTCCGATAAC 1380

    GAACGAGACT CTGGCATGCT AACTAGTTAC GCGACCCCCG AGCGGTCGGC GTCCCCCAAC 1440

    TTCTTAGAGG GACAAGTGGC GTTCAGCCAC CCGAGATTGA GCAATAACAG GTCTGTGATG 1500

    CCCTTAGATG TCCGGGGCTG CACGCGCGCT ACACTGACTG GCTCAGCGTG TGCCTACCCT 1560

    ACGCCGGCAG GCGCGGGTAA CCCGTTGAAC CCCATTCGTG ATGGGGATCG GGGATTGCAA 1620

    TTATTCCCCA TGAACGAGGA ATTCCCAGTA AGTGCGGGTC ATAAGCTTGC GTTGATTAAG 1680

    TCCCTGCCCT TTGTACACAC CGCCCGTCGC TACTACCGAT TGGATGGTTT AGTGAGGCCC 1740

    TCGGATCGGC CCCGCCGGGG TCGGCCCACG GCCCTGGGCG GAGCGCTGAG AAGACGGTCG 1800

    AACTTGACTA TCTAGAGGAA GTAAAAGTCG TAACAAGGTT TCCGTAGGTG AACCTGCGGA 1860

    AGGATCATTA                                                        1870

ITS1 ACGGAGCCCG G                                                      11
```

CHAPTER 6

DISCUSSION

6.1. Xenopus 18S rDNA

The findings of this study are that the previously reported X. laevis sequence has been corrected by the addition of one nucleotide, and that X. laevis and X. borealis 18S rDNA differ at only two points.

6.1.a. Correction to the X. laevis sequence

The previously reported 18S rDNA sequence for X. laevis (Salim and Maden, 1981; Maden et al., 1982a) has been corrected to include an A residue after G684. This nucleotide has remained undetected in previous Maxam-Gilbert sequencing studies due to the unfortunate presence of a compression at this point on both strands, masking therefore both the A residue on the RNA-like strand and the T residue on the complementary strand. The X. laevis sequence was re-examined at this point, on identifying the presence of this A residue in X. borealis clone pXbr101 by restriction analysis and subsequent sequence analysis (discussed in Chapter 4). The location of this A residue gives rise to an AluI recognition site (AGCT). AluI was not utilised in sequencing this region in previous studies (Salim and Maden, 1981; Maden et al., 1982a). This clarifies the anomaly between the number of AluI sites detected by the restriction analysis of Boseley et al., (1979), and the Maxam-Gilbert sequence analysis of Salim and Maden, (1981).

The X. laevis sequence therefore has been corrected to 1826 nucleotides, with nucleotides downstream of A685 being re-numbered plus one according to their original numbering (Salim and Maden, 1981). The presence of this A residue does not alter any other conclusion about 18S rDNA or 18S rRNA. This A residue is

located in a region already known to be phylogenetically variable among a wide range of species (region b in Salim and Maden, 1981) and so does not affect the identification of highly conserved sequences in 18S rRNA. The dideoxy chain terminator method of sequencing has given direct evidence for the presence of this extra nucleotide in all X. laevis clones examined in this study and so does not alter the conclusion that X. laevis 18S rDNA is highly homogeneous (Maden et al., 1982a). Maden et al., (1982a) also showed that X. laevis 18S rDNA does not possess any major open reading frames to allow it to act as a protein coding gene. The inclusion of an extra A does not alter this conclusion.

This extra A residue has already been incorporated in the recent secondary structure model for X. laevis 18S rRNA (Atmadja et al., 1984).

6.1.b. X. laevis and X. borealis 18S rDNA differ at only two points

Comparison of the 18S rDNA sequence for X. borealis clone pXbr101 (Furlong and Maden, 1983; and this work) with the corresponding X. laevis sequence (Salim and Maden, 1981; Maden et al., 1982a) has shown that these two species differ from each other at only two points within the 1826 nucleotides. Both changes occur as a result of a base substitution. At nucleotide 679, a G residue in the X. laevis sequence is substituted by an A residue in the X. borealis sequence. Similarly, at position 1724, a C residue in X. laevis is replaced by an A residue in X. borealis.

I am confident that the X. borealis sequence determined from clone pXbr101 is correct, and that no other differences, apart from the two stated above, exist between the two closely related Xenopus species. By carrying out the sequence analysis of both species in the same laboratory, the X. laevis data was at hand for cross-checking purposes and I was able to take particular care in regions already known to be technically difficult.

The two sites of variation occur in regions already known to be phylogenetically variable (Salim and Maden, 1981). Nucleotide 679 is contained within variable region b (encompassing nucleotides 651-761) and nucleotide 1724 is

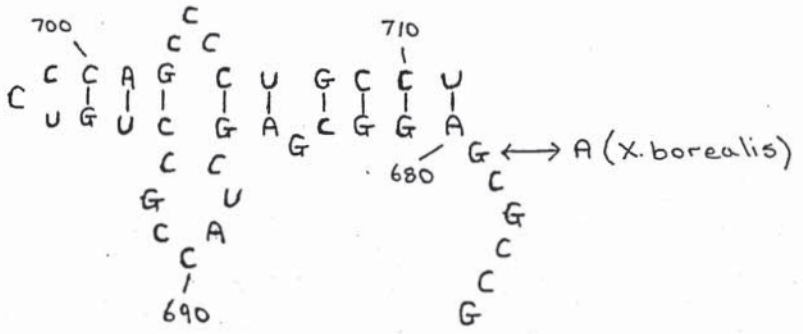
contained within variable region d (nucleotides 1696-1770). Therefore, these sites of minimal divergence between X. laevis and X. borealis 18S rDNA's are in accordance with larger phylogenetic trends.

It was conceivable that these two sites of variation might have indicated sites of sequence heterogeneity within X. borealis 18S rDNA. However, sequence analysis of a range of X. borealis and X. laevis clones through both of these sites (sumarised in Table 4.1) showed the sequence at both sites to be fixed within each species, and there were no signs of sequence heterogeneity within the surrounding phylogenetically variable regions. This complements the findings of Maden et al., (1982a), that X. laevis 18S rDNA is highly homogeneous. I am confident that likewise, the 18S sequence obtained from X. borealis clone pXbr101 is representative of X. borealis. It could be argued that one would have to look at many more clones to seek for signs of heterogeneity. However, I think it can be concluded that although it would be incorrect at this present time to exclude the possibility of very low levels of intraspecies heterogeneity at the two sites of variation, this work does not indicate major heterogeneities.

How do the two sequence changes alter the secondary structure model for X. laevis 18S rRNA? Figure 6.1 shows the relevant regions from the secondary structure model of Atmadja et al., (1984), containing the two points of variation. Nucleotide 679 is contained within a single-stranded region of rRNA at the base of a helical arm. A G ——> A substitution can occur freely without having any consequence on the secondary structure. Nucleotide 1724 is located at the extreme tip of a helical arm. Again, the C ——> A substitution does not destabilise the secondary structure model in any way. Both of these changes occur in regions of the molecule where the details of secondary structure are known to be variable among a wide range of organisms. So again, the minimal divergence between X. laevis and X. borealis in their secondary structure models fits in with regions of much larger scale phylogenetic trends.

Figure 6.1 Location of the two points of variation in the secondary structure model for X. laevis 18S rRNA

The upper figure shows the location of the base substitution at position 679. The lower figure shows the location of the base substitution at position 1724.



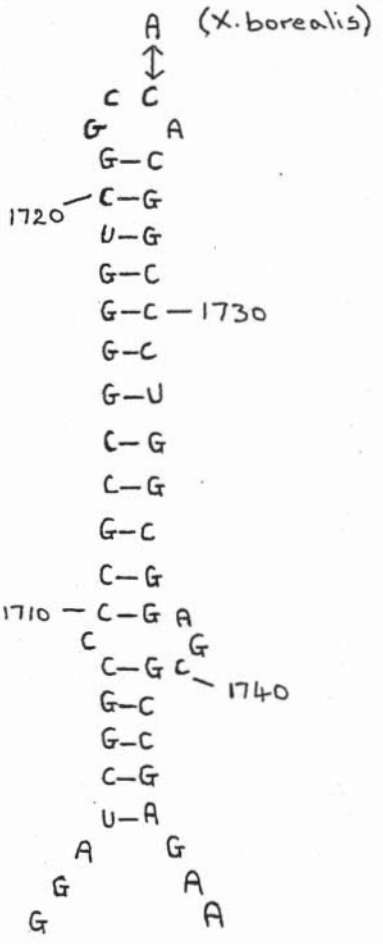


FIGURE 6.1

6.1.c. Very different rates of divergence in the 18S gene regons and transcribed spacers

The comparative data on the 18S gene region obtained from this study contrast greatly with the comparative sequence data on transcribed spacers. As discussed in the Introduction, transcribed spacer regions show much greater interspecies divergence than the gene coding regions. Comparative analysis between the transcribed spacers of X. laevis and X. borealis showed extensive divergence between the two species, with insertions and deletions playing a major role in their divergence (Furlong et al., 1983; Furlong and Maden, 1983). However, limited sequence data within the ribosomal coding regions suggested that there was minimal divergence between these two species. Having completed the 18S gene sequence for X. borealis, we can begin to quantify this very great difference in divergence rates between the different regions of the ribosomal transcription unit.

It is relevant to note that by comparing the serum albumins of X. laevis and X. borealis, the time of divergence of these two species has been estimated to be 10 million years (Bisbee et al., 1977). In this time, the 18S rDNA's of these two species have diverged at only two points within the 1826 nucleotide sequence. That is, they have diverged by only 0.11%. The degree of sequence divergence between the transcribed spacers of X. laevis and X. borealis is very much more difficult to quantify, since divergence has been so rapid that in a large proportion of these regions no measurable degree of homology remains, although there are several short tracts of complete homology. This extensive divergence in the spacer regions has evidently resulted from the cumulative effect of many substitutions, insertions and deletions. It is difficult to put a figure on the rate of divergence of the transcribed spacers. However, the data suggest that it is orders of magnitude greater than the rate of divergence of the 18S gene.

Relatedly, these very different rates of divergence are evident when confining the examination to only one species. The 18S coding regions from cloned and uncloned material from X. laevis show complete homology (Maden et al.,

1982a). In contrast to this, Stewart et al., (1983) showed that the three transcribed spacer regions contain heterogeneities, including single base changes and length variants of one to several nucleotides. These heterogeneities must have arisen from relatively recent mutations which have not yet reached fixation. Within 1,100 nucleotides of spacer sequence there were 20 sites at which variation was seen to occur. Any two clones could differ at several of these sites, giving a divergence of up to 1% between the transcribed spacer regions of individual rDNA units. Therefore, there is up to ten times greater intraspecies variation within the transcribed spacers of X. laevis, than between the 18S coding regions of X. laevis and X. borealis (0.11%). Work is being carried out at present on the transcribed spacers of X. borealis and there is indication that sequence heterogeneity does exist as for X. laevis (B. E. H. Maden, unpublished results).

This very great difference between the rates of divergence of gene coding regions and transcribed spacers does not appear to be unique to Xenopus. Recent analysis on the transcribed spacers of rat and mouse shows extensive divergence between the two species (Michot et al., 1983). In contrast to this, the recently published sequence of the 18S gene coding region of mouse (Raynal et al., 1984) shows only 14 differences from the corresponding rat sequence of Torczynski et al., (1983).

Thus, the accumulating sequence data does stress the obvious requirement to conserve certain regions of 18S rRNA for proper functioning of the mature ribosome. It is doubtful that any function of the transcribed spacers in ribosome maturation is directly related to the nucleotide sequence, although the short tracts of homology contained within the spacers may have a role to play.

6.2. Human 18S rDNA

The work of this project has also led to the elucidation of the complete 18S rDNA sequence for human. This sequence (shown in Figure 5.22) is 1870

(a) Xenopus and Human 18S rDNA base composition data

Species	No. of Nucleotides		T	A	C	G	% G + C
Xenopus	1826	X. laevis	411	433	466	516	53.8
		X. borealis	411	435	465	515	53.7
Human	1870		401	419	500	550	56.1

(b) 18S rDNA base change data from Xenopus ——> Human

Number of deletions	0	
Number of insertions	44	
(A or T insertions)	7	
(G or C insertions)	37	
Number of base substitutions		79
(A or T) ——> (G or C)		52
(G or C) ——> (A or T)		19

Table 6.1

nucleotides long. In the remainder of this Chapter, I shall discuss how this human sequence compares with other 18S data. First, I shall make a comparison with the other species examined in this study, that is Xenopus and then I shall look at the comparison with other mammalian genes.

6.2.a. Comparison between Human and Xenopus 18S rDNA

In Table 6.1(a), I have summarised the base composition data for Xenopus and Human 18S rDNA. The human gene is seen to be 44 nucleotides longer than the Xenopus sequence. The number of A and T residues is greater in Xenopus than in human, and the number of G and C residues increases on going from Xenopus to human, giving a difference in G + C content of ~2.4%. This is in agreement with established knowledge that higher vertebrates have an increased G + C content over lower vertebrates and other eukaryotes.

In Figure 6.2, I have aligned the Xenopus and human 18S sequences and adjacent spacers to give the best fit between the two. Table 6.1(b) summarises the differences. The Xenopus and human 18S sequences differ in total length by 44 nucleotides (1826 ——> 1870). This length difference is accounted for by 44 insertions between Xenopus and human. There are no nucleotides in Xenopus which are not present in human. The number of A or T insertions is much lower than the number of G or C insertions (7 and 37 respectively), contributing to the higher G + C content of human 18S rDNA. All other differences take the form of base substitutions. Between X. borealis and human there are 79 such changes. Again there is a preference for (A, T) ——> (G,C) over (G, C) ——> (A, T) (52 and 19 respectively). One point to note, is that the number of base substitutions between X. laevis and human is only 77. As discussed in 6.1.b., the X. laevis and X. borealis 18S genes differ at two points, at nucleotides 679 and 1724 in the Xenopus sequence. On aligning the human and Xenopus sequences (Figure 6.2), at positions 679 and 1724 (numbering according to the Xenopus sequence), the human sequence is identical to the X. laevis sequence (nucleotides 714 and 1766 in the human

Figure 6.2 Comparison of the 18S rDNA sequences of human and Xenopus

The complete human 18S rDNA sequence is shown. Differences in the X. borealis and X. laevis sequences are noted by subscripts. A dash (-) indicates a deletion. Separate numbering is given for the human and Xenopus sequences. A small amount of adjoining ETS and ITS 1 data are also included.

```
 ETS                                                AG GTCGCGCGCT CTACCTTACC  -1
X.b.                                                GT TCCCCCCCGG AGCCGAGGGC
X.l.                                                CG CCGGGCCCGG GAAAGGTGGC

 18S TACCTGGTTG ATCCTGCCAG TAGCATATGC TTGTCTCAAA GATTAAGCCA TGCATGTCTA  60
                                                                C  G    60

     AGTACGCACG GCCGGTACAG TGAAACTGCG AATGGCTCAT TAAATCAGTT ATGGTTCCTT 120
                                                                       120

     TGGTCGCTCG CTCCTCTCCC ACTTGGATAA CTGTGGTAAT TCTAGAGCTA ATACATGCCG 180
       A      - -- A   GTT                                            177

     ACGGGCGCTG ACCCCCTTCG CGGGGGGGAT GCGTGCATTT ATCAGATCAA AACCAACCCG 240
        A             ---- ----A                      C          T    229

     GTCAGCCCCT CTCCGGCCCC GGCCGGGGGG CGGGCGCCGG CGGCTTTGGT GACTCTAGAT 300
      --G     C ------  ---  -   ------ ------      C                  266

     AACCTCGGGC CGATCGCACG CCCCCCGTGG CGGCGACGAC CCATTCGAAC GTCTGCCCTA 360
                           T    -   A           T A      G T           325

     TCAACTTTCG ATGGTAGTCG CCGTGCCTAC CATGGTGACC ACGGGTGACG GGGAATCAGG 420
                      C TT  T C                          A              385

     GTTCGATTCC GGAGAGGGAG CCTGAGAAAC GGCTACCACA TCCAAGGAAG GCAGCAGGCG 480
                                                                       445

     CGCAAATTAC CCACTCCCGA CCCGGGGAGG TAGTGACGAA AAATAACAAT ACAGGACTCT 540
                            G                                          505

     TTCGAGGCCC TGTAATTGGA ATGAGTCCAC TTTAAATCCT TTAACGAGGA TCCATTGGAG 600
                                  A                          T        565

     GGCAAGTCTG GTGCCAGCAG CCGCGGTAAT TCCAGCTCCA ATAGCGTATA TTAAAGTTGC 660
                                                                       625

     TGCAGTTAAA AAGCTCGTAG TTGGATCTTG GGAGCGGGCG GGCGGTCCGC CGCGAGGCGA 720
X.b.                                     T  A  T                  A    685
X.l.                                                              G

     GCCACCGCCC GTCCCCGCCC CTTGCCTCTC GGCGCCCCCT CGATGCTCTT AGCTGAGTGT 780
       T      T       A       -              T  C             GA      744

     CCCGCGGGGC CCGAAGCGTT TACTTTGAAA AAATTAGAGT GTTCAAAGCA GGCCCGAGCC 840
        -                                          C          - C T    802

     GCCTGGATAC CGCAGCTAGG AATAATGGAA TAGGACCGCG GTTCTATTTT GTTGGTTTTC 900
                 TT                          TC                       862

     GGAACTGAGG CCATGATTAA GAGGGACGGC CGGGGGCATT CGTATTGCGC CGCTAGAGGT 960
           G                                          T                922

     GAAATTCTTG GACCGGCGCA AGACGGACCA GAGCGAAAGC ATTTGCCAAG AATGTTTTCA 1020
                                  A     A                              982

     TTAATCAAGA ACGAAAGTCG GAGGTTCGAA GACGATCAGA TACCGTCGTA GTTCCGACCA 1080
                                                                       1042

     TAAACGATGC CGACCGGCGA TGCGGCGGCG TTATTCCCAT GACCCGCCGG GCAGCTTCCG 1140
                      TA      C                              A         1102

     GGAAACCAAA GTCTTTGGGT TCCGGGGGGA GTATGGTTGC AAAGCTGAAA CTTAAAGGAA 1200
                                                                       1162

     TTGACGGAAG GGCACCACCA GGAGTGGAGC CTGCGGCTTA ATTTGACTCA ACACGGGAAA 1260
                                                                       1222

     CCTCACCCGG CCCGGACACG GACAGGATTG ACAGATTGAT AGCTCTTTCT CGATTCCGTG 1320
                              A                                  T     1282

     GGTGGTGGTG CATGGCCGTT CTTAGTTGGT GGAGCGATTT GTCTGGTTAA TTCCGATAAC 1380
                                                                       1342

     GAACGAGACT CTGGCATGCT AACTAGTTAC GCGACCCCCG AGCGGTCGGC GTCCCCCAAC 1440
                   CTC                                  -              ---  1398

     TTCTTAGAGG GACAAGTGGC GTTCAGCCAC CCGAGATTGA GCAATAACAG GTCTGTGATG 1500
                                      A      C                         1458

     CCCTTAGATG TCCGGGGCTG CACGCGCGCT ACACTGACTG GCTCAGCGTG TGCCTACCCT 1560
                                            AC   A           T         1518

     ACGCCGGCAG GCGCGGGTAA CCCGTTGAAC CCCATTCGTG ATGGGGATCG GGGATTGCAA 1620
     G     A     T             C          G          A                1578

     TTATTCCCCA TGAACGAGGA ATTCCCAGTA AGTGCGGGTC ATAAGCTTGC GTTGATTAAG 1680
          T                                          C                1638

     TCCCTGCCCT TTGTACACAC CGCCCGTCGC TACTACCGAT TGGATGGTTT AGTGAGGCCC 1740
                                                                   T   1698

     TCGGATCGGC CCCGCCGGGG TCGGCCCACG GCCCTGGGCG GAGCGCTGAG AAGACGGTCG 1800
X.b.                           A-           -          C          A  A 1756
X.l.                           C

     AACTTGACTA TCTAGAGGAA GTAAAAGTCG TAACAAGGTT TCCGTAGGTG AACCTGCGGA 1860
                                                                       1816

     AGGATCATTA                                                        1870
                                                                       1826

ITS1 ACGGAGCCCG G                                                       11
X.b. ACGAGAGAGG G
X.l. ACGAGACCCC C
```

sequence). In total then, the human and Xenopus sequences show 123 differences (44 insertions and 79 base substitutions) within 1870 nucleotides. That is a divergence of 6.6%. Xenopus and mammalian sequences are thought to have diverged some 300 million years ago. Therefore, a divergence between Xenopus and human of 6.6% is consistent with the previous discussion for X. laevis and X. borealis, where divergence has only taken place within the last 10 million years giving a sequence divergence of only 0.11%.

Where do the 123 differences lie within the nucleotide sequence? The differences are not spread evenly throughout the molecule, but are seen to be concentrated in specific areas. Within the 9 nucleotide stretch encompassing nucleotides 197-205 (numbering according to the human sequence), there is complete divergence between the two species, comprising 1 base substitution and 8 insertions. Also, within the 35 nucleotides (242-276) there are 25 differences, including 23 insertions. Both of these regions fall within variable region a between Xenopus and yeast (nucleotides 176-278 in the Xenopus sequence, Salim and Maden, 1981). This region is shown to have an increased G plus C content on going from yeast ⟶ Xenopus ⟶ human. Looking at Figure 6.2 again, a substantial proportion of the differences between Xenopus and human occur within nucleotides 690-880 (numbering according to the human sequence). This region contains another region shown to be variable between Xenopus and yeast (variable region b (651-761), Salim and Maden, 1981).

There are several completely conserved regions between Xenopus and human containing ~100 nucleotides or more: 1, nucleotides 408-501, 94 nucleotides long; 2, nucleotides 594-693, 100 nucleotides long; 3, nucleotides 992-1094, 103 nucleotides long; 4, nucleotides 1131-1282, 152 nucleotides long. Regions 1 and 2 correspond to tracts A and B shown to be conserved between yeast, Xenopus and rat (Torczynski et al., 1983). The 3' 70 nucleotides are also conserved between the two species, the last 49 of which correspond to region C, again shown to be completely conserved between yeast, Xenopus and rat (Torczynski et al., 1983).

Several of the differences between Xenopus and human 18S rDNA result in a change in restriction patterns between the two species. Nucleotides are numbered according to the human sequence. For example, at position 227, there is a C ——> T substitution from Xenopus to human. This gives a Sau3A site (GA$\overset{*}{T}$C) in the human sequence. At nucleotide 502, there is a G ——> C substitution, giving rise to an SmaI site (C$\overset{*}{C}$CGGG) in the human sequence. At position 593, there is a T ——> C substitution, giving a BamHI site in the human sequence (GGATC$\overset{*}{C}$). At position 785, there is a C insertion in the human sequence, resulting in the loss of an SmaI site (CCCG$\overset{C}{}$GG) present in the Xenopus sequence. At position 1668, a C ——> T substitution results in the presence of a HindIII site (AAGCT$\overset{*}{T}$) in the human sequence. Restriction analysis of human 18S rDNA has shown the presence of restriction sites for SmaI and BamHI (Wilson et al., 1982).

For completeness, I should like to note that on extending the comparative analysis between Xenopus and human outside the 18S gene region to the adjoining ETS and ITS 1 regions, the limited amount of data available shows almost complete divergence between the two species on leaving the 18S gene. This is as expected from our previous discussion on the much higher rates of divergence of spacer regions in contrast to gene regions.

6.2.b. Comparison of human 18S rDNA with other mammalian data

I shall now go on to compare the human 18S rDNA sequence with other more closely related species. In Table 6.2, I have listed the differences between human and rat 18S coding regions. It is difficult to put an exact figure on the number of differences between human and rat, as there are two published sequences for rat 18S rDNA which differ from each other (Torczynski et al., 1983; Chan et al., 1984).

Comparing the human sequence with the rat sequence, as determined by Torczynski et al., (1983), there are 23 differences comprising 9 insertions or deletions and 14 base substitutions. These changes favour a slightly higher G + C content in human than rat 18S rDNA. 23 differences between the two species gives

NUCLEOTIDE	HUMAN (This work)	RAT (Torczynski et al., 1983)	RAT (Chan et al., 1984)	Definite Differences
123	G	-	G	
140	C	T	T	✓
196a	-	C	C	✓
198a	-	C	C	✓
199a	-	C	C	✓
201	C	T	T	✓
208	G	G	A	
210	T	C	C	✓
250	T	C	C	✓
251	C	T	T	✓
252	T	C	C	✓
256	G	-	G	
258	C	T	T	✓
270	G	T	T	✓
278a	-	-	C	
280a	-	A	-	
321	C	C	T	
324	C	T	-	
407	G	A	G	
720	A	A	G	
721	G	G	C	
722	C	T	T	
725	C	C	G	✓
730a	-	-	T	
736a	-	-	A	
743	T	T	-	
845	G	A	G	
986	G	G	A	
1095	C	T	T	
1228	A	-	A	✓
1295	A	A	G	
1392a	-	-	C	
1472	C	C	-	
1537a	-	-	A	
1542	C	C	T	
1774	C	T	T	BstNI site
1777	G	-	G	
1778	G	-	-	✓
1784	C	C	G	
1785	G	G	C	

Table 6.2 Base change data between human and rat 18S rDNA

Nucleotides are numbered according to the human sequence.

a divergence of 1.2%. This is consistent with the closer relatedness of human and rat (in evolutionary terms), as opposed to human and Xenopus whose 18S gene sequences have diverged by 6.6%.

The two rat sequences differ from each other at 25 sites. How can we account for the differences between the two published rat sequences? During the work carried out on Xenopus 18S rDNA in our own laboratory, there was no sign of intraspecies heterogeneity in the 18S gene coding regions (Maden et al., 1982a). However, I cannot rule out the possibility that there may be some minimal divergence of gene coding sequences in mammalian species. Only by carrying out more sequence analysis on several clones from individual species (for example human) will we be able to answer this question. However, at present I would speculate that the majority, if not all of the differences between the two rat sequences, may turn out to be minor sequencing errors in one or both sequences. By looking at the locations of some of the differences between the two rat sequences, I think the latter suggestion is likely to be correct.

In sequencing the 3' end of the human 18S gene, I had difficulty in determining the nucleotide sequence in the region 1770-1780, where a difficult compression and an apparent TT doublet could only be clarified by restriction analysis (discussed in Chapter 5, section 5.8). This region was shown to contain a restriction site for BstNI (CCTGG). Neither of the rat sequences contain a BstNI site. Both sequences show a TT doublet, and the sequence of Torczynski et al., (1983) shows only 1 G residue. However, since this BstNI site is shown to be conserved between Xenopus and human, I would speculate that determination of the rat sequence has encountered similar difficulties to those I have detailed for the human analysis, and that restriction with BstNI would show the presence of this site in rat 18S rDNA. Recently published sequence data for mouse 18S rDNA (Raynal et al., 1984) demonstrates the presence of a BstNI site at this position in the mouse sequence.

In summary, I would suggest that where the two rat 18S sequences agree with

each other, and the corresponding site in the human sequence is different, then those are likely to be real differences between the two species. Looking at Table 6.2, I would say that there are probably only as few as 14 real differences between human and rat 18S rDNA. So the true extent of divergence between human and rat 18S rDNA could be as low as 0.75%, as opposed to the higher estimate of 1.2%.

The recently published mouse 18S sequence (Raynal et al., 1984) contains 14 differences from the rat sequence of Torczynski et al., (1983). However, again some of these are at sites where the two rat sequences differ. So, the real number of differences is probably less than 14. Complete accuracy of sequence determination will be an absolute requirement in continuing studies on closely related species, because in instances where there are expected to be only a small number of differences between the complete 18S genes, incorrect sequence data at only one or a few points will yield quite a marked difference from the true rate of divergence.

Despite the slight anomalies between the rapidly accumulating data for 18S rDNA's, all the present data support a common arrangement within 18S rDNA structure. The highly variable regions originally identified between yeast and Xenopus (Salim and Maden, 1981), are shown to be retained on the addition of the mammalian data. The corresponding regions in 18S rRNA structure are unlikely to be vital for ribosome formation and function. The addition of the human and mouse sequences to data on the comparative analysis between yeast, Xenopus and rat (Torczynski et al., 1983), supports the identification of 3 large tracts which are completely conserved among all these species (regions, A, B and C in Torczynski et al., 1983). Complete conservation of these regions over a long period of evolutionary time, strongly suggests their involvement in ribosome function. Embedded within these tracts are sequences which are common to both eukaryotic 18S rRNA's and prokaryotic 16S rRNA's. So perhaps these regions in 18S rRNA play a similar role in eukaryotic ribosome function, as the corresponding regions of E.coli 16S rRNA do in prokaryotic ribosome function.

Therefore, perhaps now we are beginning to see more precisely which regions of 18S rRNA are important for proper ribosome functioning, and it will not be too long in the future before precise roles are assigned to these sequences.

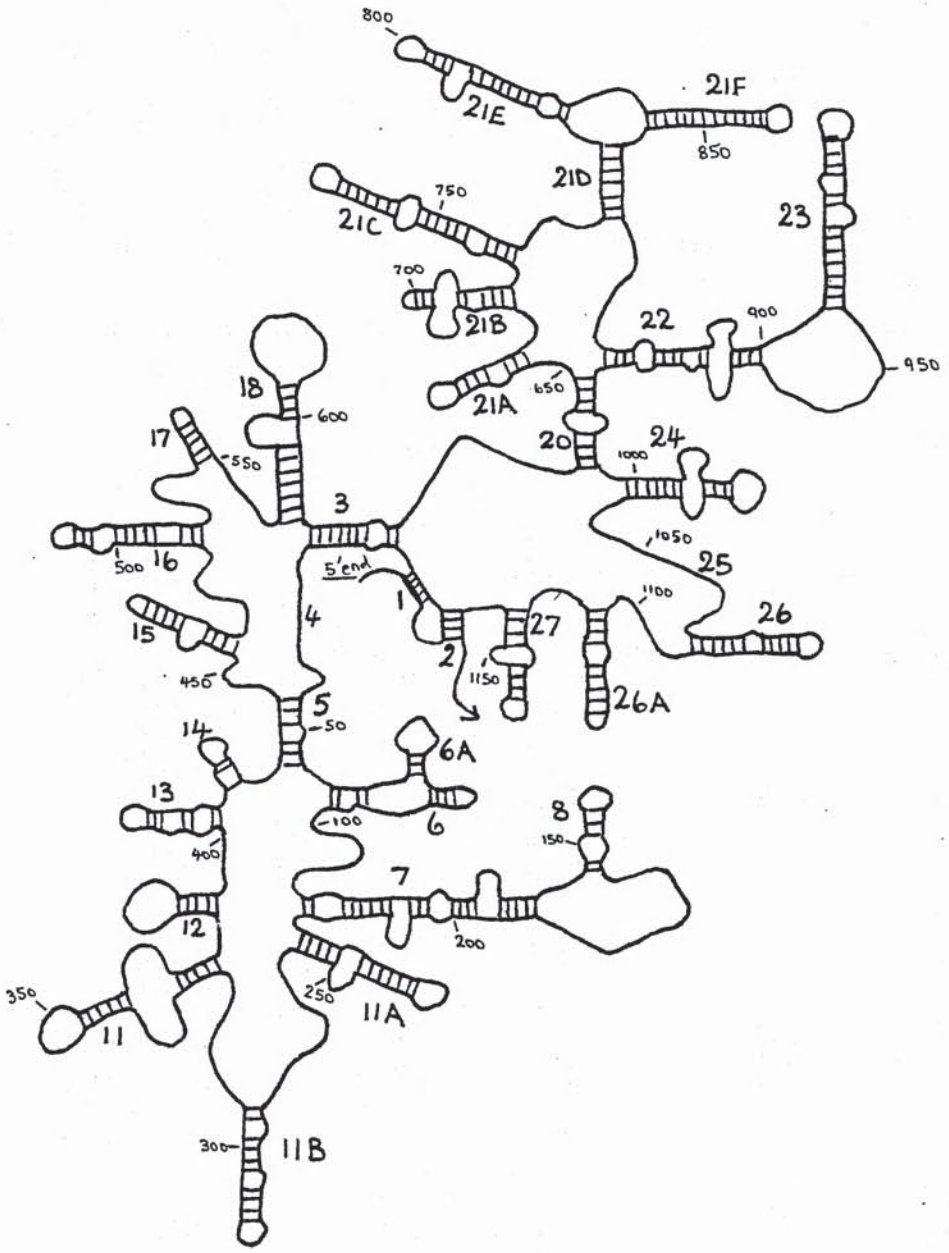
6.3. Secondary structure model for 18S rRNA

How do the accumulating sequence data help to define a consensus model for eukaryotic 18S rRNA? Two new models have recently been published : a revised model for X. laevis 18S rRNA (Atmadja et al., 1984) and a model for rat 18S rRNA (Chan et al., 1984).

In this present study, I have shown that the human and rat 18S sequences may differ at as few as 14 positions. These differences between the two species can be accomodated quite readily in the secondary structure model for rat. Almost half of the differences occur in loop structures and so do not affect the secondary structure in any way. Of the changes that occur in helical regions, only a few alter a base pair. However, the general structure is identical for the two species. Also, the human sequence shows only 121 differences from the corresponding X. laevis sequence and so it would seem reasonable to expect the Xenopus and mammalian models to be very similar to each other. However, several areas of the two present models (Xenopus and rat) show quite different secondary structure arrangements. I shall discuss these differences and suggest, where possible, which arrangement is likely to be more correct.

In Figure 6.3(a), I have drawn a diagram of the A domain of X. laevis 18S rRNA (Atmadja et al., 1984). In Figure 6.3(b), I have drawn the complete model for rat 18S rRNA (Chan et al., 1984). I have not depicted domain B in the Xenopus model, since the two models are in very close agreement in the arrangement of this region. There are some minor discrepancies over the precise folding of some helical regions, but the overall structure is very similar between the two.

I shall concentrate the remainder of this discussion on domain A. For ease of



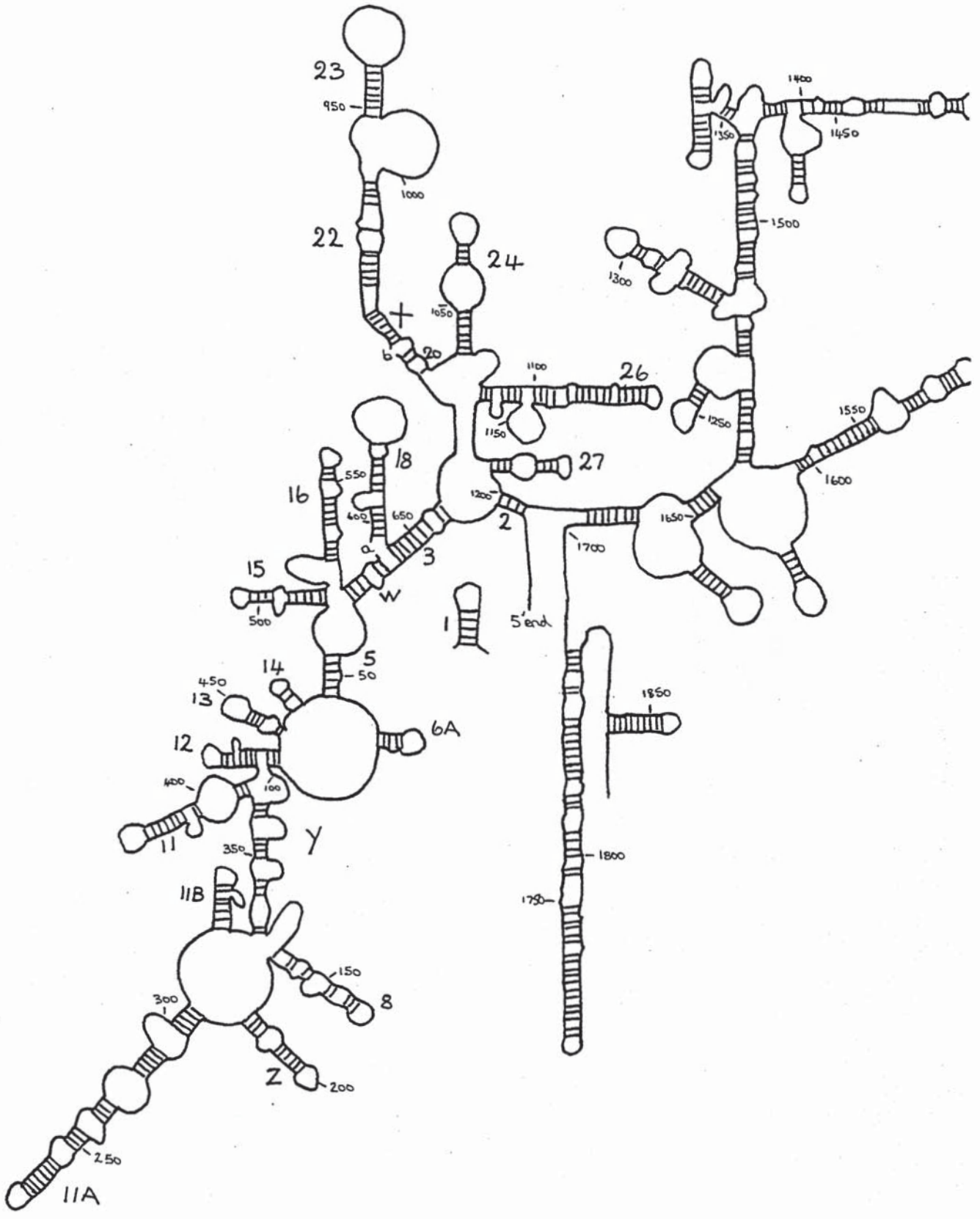
(a)

Figure 6.3 Secondary structure models for 18S rRNA

(a) shows the A domain of Xenopus laevis 18S rRNA, according to Atmadja et al., (1984).

(b) shows the complete model for rat 18S rRNA, according to Chan et al., (1984).

Labelling of regions within each model is explained in the text.



(b)

Note: the separate hairpin structure in the centre shows an alternative arrangement for the first 21 residues.

cross-reference I have labelled regions that are common to both models, according to their identification by Atmadja et al., (1984) in the Xenopus model.

In the rat model, region a (nucleotides 573-597) has not been assigned a secondary structure. In the rat model, a helix (W) has been formed between nucleotides 32-38 and 565-572. By pulling these two regions together, helices 16 and 18 come into close proximity. Thus, a structure similar to helix 17 in the Xenopus model cannot be fitted into the rat model. However, region a in the rat sequence would allow a similar structure to that for Xenopus at this point, and so I would predict that helix W is wrong in the rat model, and that a structure similar to that for Xenopus is correct.

A second region of the rat sequence has not been assigned a secondary structure. Region b (nucleotides 680-917) has not been included. This region is equivalent to nucleotides 642-875 in the Xenopus sequence. In the Xenopus model this sequence encompasses structures 20, and 21A-21F. This is still regarded as a tentative arrangement in the Xenopus model, however base pairing data support the presence of 21C. (Atmadja et al., 1984). I would suggest that something similar to this structure is formed by region b in the rat model. To accommodate this, I would say that helix X in the current rat model has been paired wrongly.

Both models contain similar structures for regions 22, 23 and 24. However on coming towards the end of domain A, there is a large helical structure in the rat model encompassing nucleotides 1093-1164. This corresponds to nucleotides 1051-1122 in the Xenopus model. In Xenopus, there is a much more loose structure encompassing region 25, and helices 26 and 26A. Atmadja et al., (1984) had originally expected to find a structure similar to that depicted in the rat model. However, experimental data has shown the presence of base paired fragments which form helix 26A. Atmadja et al., (1984) looked at this region in close detail and have identified a possible phylogenetically conserved "switch" region between E. coli 16S and eukaryotic 18S rRNA. If this hypothesis is correct, I would suggest that helix 26A is also present in the rat model. This then leads into helix 27 which

is conserved in both models, and then into domain B.

The greatest structural differences between the Xenopus and rat models occur within the 5' 400 or so nucleotides. Starting at the 5' end, the Xenopus model shows helix 1, formed between residues 3-21. The rat model shows this as an alternative structure. Helices 2 and 3 are conserved. Having discussed previously that helix W is incorrect in the rat model, region 4 in the Xenopus model will be common to both. Helix 5 is conserved between the two models.

After this point, the two models differ quite markedly. In the rat model there is a very loose region encompasing nucleotides 54-95, except for the small helical region equivalent to 6A in the Xenopus model. However, in the Xenopus model, nucleotides 78-81 are base paired with 85-88 to give helix 6, and also nucleotides 56-59 are linked to nucleotides 92-96. The rat sequence does have the potential to form a similar structure. From this alone, I cannot make any preference between either model. However, I shall return to this point.

Opposite to this region, helix 14 is identical in both models. Helix 13 is also present in both, although there are some discrepancies between how these nucleotides are paired. In the Xenopus model, helices 13 and 12 are separated by a short tract encompassing nucleotides 397-400, and helix 12 starts with the pairing of G396 with C374. However, in the rat model, the much looser structure in region 54-95 (except for helix 6A, discussed previously), allows the pairing of nucleotides 96-98 with 435-437 (96-98 and 397-399 in the Xenopus sequence). By pulling over nucleotides 96-98, region Y in the rat model can be formed. This gives a long interrupted helical region between nucleotides 106-129 and nucleotides 341-358 (equivalent to 106-129 and 303-320 in the Xenopus model). Thus, these two regions are brought much closer together in the rat model, and a much more ordered structure results, as opposed to a rather "open" arrangement in the Xenopus model. In Figure 6.4, I have drawn out this region Y for both rat and Xenopus. Figure 6.4 shows that the 5 base substitutions between the two species in this region are readily accommodated in this proposed arrangement of secondary structure. There

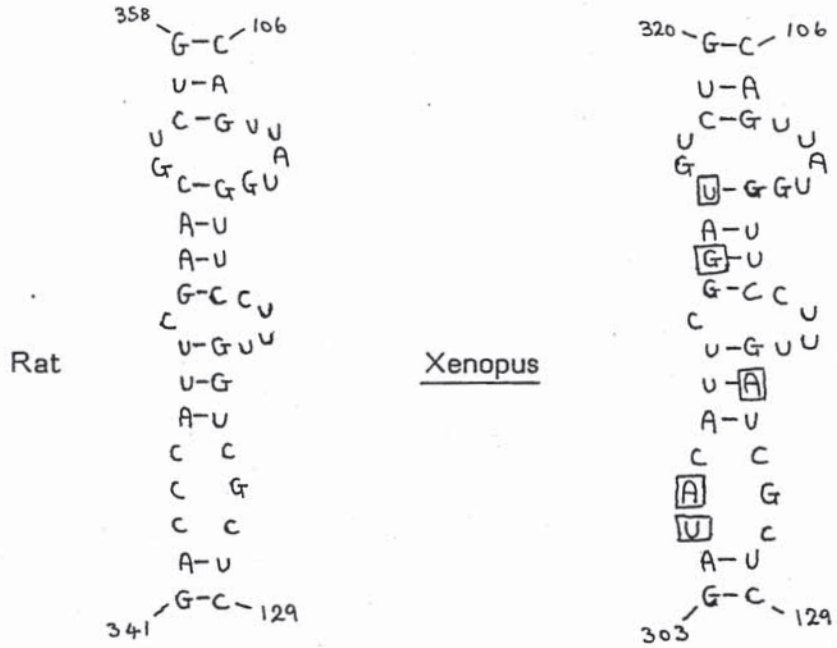


Figure 6.4 Secondary structure model : region Y

The left hand figure shows region Y in the rat model of Chan et al., (1984). In the right hand figure the corresponding regions of Xenopus 18S rRNA have been drawn in the same arrangement. Base changes between the two species are boxed in the Xenopus model.

are no differences between rat and human 18S rRNA in this region. I would suggest that the formation of this region Y in the rat model is a much better arrangement than the corresponding Xenopus model. Relatedly, the arrangment of nucleotides 54-95 must be as depicted in the rat model to enable region Y to form.

The formation of region Y will lead to the loss of helix 7 in the Xenopus model. Helix 8 is present in both. As a consequence of the loss of helix 7, there is the introduction of helix Z. This helix will be variable in length, to accommodate variable region ~195-205, where there is a GC rich insertion in mammals. Helix 11A is common to both models. Again, this helix is variable in size to accommodate the large number of GC insertions in region 240-280. Helix 11B is a feature common to both models. This then leads back into region Y.

By comparing the secondary structure models for Xenopus and rat 18S rRNA's, I hope to have been able to define the structure of a model that can accommodate both species, and hopefully be representative of a wide range of eukaryotic 18S rRNA's.

REFERENCES

Amaldi, F. (1969) Nature, 221, 95-96.

Amaldi, F. and Attardi, G. (1968) J. Mol. Biol. 33, 737-755.

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

Arnheim, N. (1979) Gene 7, 83-96.

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

Atmadja, J., Brimacombe, R., and Maden, B. E. H. (1984) Nucleic Acids Research 12, 2649-2667.

Bach, R., Allet, B., and Crippa, M. (1981) Nucleic Acids Research 9, 5311-5330.

Bachellerie, J-P., Michot, B., and Raynal, F. (1983) Molec. Biol. Rep. 9, 79-86.

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

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

Bisbee, C. A., Baker, M. A., Wilson, A. C., Hadzi-Azimi, I., and Fishberg, M. (1977) Science 195, 785-787.

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

Boseley, P., Moss, T., Mächler, M., Portman, R., and Birnstiel, M. (1979) Cell 17, 19-31.

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

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

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

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

Brown, D. D. and Weber, C. S. (1968a) J. Mol. Biol. 34, 661-680.

Brown, D. D. and Weber, C. S. (1968b) J. Mol. Biol. 34, 681-697.

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

Burgers, P. M. J. and Eckstein, F. (1978) Proc. Nat. Acad. Sci. USA 75, 4798-4800.

Burgers, P. M. J., and Eckstein, F. (1979) J. Biol. Chem. 254, 6889-6893.

Chan, Y-L., Gutell, R., Noller, H. F. and Wool, I. G. (1984) J. Biol. Chem. 259, 224-230.

Chan, Y-L., Olvera, J. and Wool, I. G. (1983) Nucleic Acids Research 11, 7819-7831.

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

Coen, E. S. and Dover, G. A. (1982) Nucleic Acids Research 10, 7017-7026.

Connaughton, J. F., Rairkar, A., Lockard, R. E. and Kumar, A. (1984) Nucleic Acids Research 12, 4731-4745.

Crouch, R. J., Kanaya, S., and Earl, P. L. (1983) Molec. Biol. Rep. 9, 75-78.

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

Duckworth, M. L., Gait, M. J., Goelet, P., Hong, G. F., Singh, H., and Titmas, R. C. (1981) Nucleic Acids Research 9, 1691-1706.

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

Eckstein, F., Romaniuk, P. J. and Connolly, B. A. (1982) Methods in Enzymology 87, 197-212.

Erickson, J. M., Rushford, C. L., Dorney, D. J., Wilson, G. N., and Schmickel, R. D. (1981) Gene 16, 1-9.

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

Fuke, M., Dennis, K. J. and Busch, H. (1981) Mol. Gen. Genet. 182, 25-30.

Furlong, J. C., Forbes, J., Robertson, M. and Maden, B. E. H. (1983) Nucleic Acids Research 11, 8183-8196.

Furlong, J. C., and Maden, B. E. H. (1983) The EMBO J. 2, 443-448.

Georgiev, O. I., Nikolaev, N., Hadjiolov, A. A., Skryabin, K. G., Zakharyev, V. M. and Bayev, A. A. (1981) Nucleic Acids Research 9, 6953-6958.

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

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

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

Greenberg, H. and Penman, S. (1966) J. Mol. Biol. 21, 527-535.

Groneborn, B. and Messing, J. (1978) Nature 272, 375-377.

Grummt, I., Soellner, C., and Scholz, I. (1979) Nucleic Acids Research 6, 1351-1369.

Hall, L. M. C. (1981) Ph.D. Thesis, University of Glasgow.

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

Heidecker, G., Messing, J., and Gronenborn, B. (1980) Gene 10, 69-73.

Henderson, A. S., Warburton, D., and Atwood, K. C. (1972) Proc. Nat. Acad. Sci. USA 69, 3394-3398.

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

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

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

Klootwijk, J. and Planta, R. J. (1973) Eur. J. Biochem. 39, 325-333.

Kunkel, T. A., Eckstein, F. Mildvan, A. S., Koplitz, R. M., and Loeb, L. A. (1981) Proc. Nat. Acad. Sci. USA 78, 6734-6738.

Lewin, B. (1980) In Gene Expression 2, Second Edition, Eukaryotic Chromosomes 865-906, Wiley Interscience.

Leoning, U. E. (1968) J. Mol. Biol. 38, 355-365.

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

McDonnell, M. W., Simon, M. N. and Studier, F. W. (1977) J. Mol. Biol. 110, 119-146.

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

Maden, B. E. H., Forbes, J. M., Stewart, M. A., and Eason, R. (1982a) The EMBO J. 1, 597-601.

Maden, B. E. H., Moss, M. and Salim M. (1982b) Nucleic Acids Research 10, 2387-2398.

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

Maden, B. E. H., Salim, M., and Summers, D. F. (1972) Nature New Biol. 237, 5-9.

Maden, B. E. H. and Reeder, R. H. (1979) Nucleic Acids Research 6, 817-830.

Maizels, N. (1976) Cell 9, 431-438.

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

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

Messing, J., Crea, R., and Seeburg, P. H. (1981) Nucleic Acids Research, 9, 309-321.

Messing, J., Groneborn, B., Müller-Hill, B. and Hans Hofschneider, P. (1977). Proc. Nat. Acad. Sci. USA 74, 3642-3646.

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

Michot, B., Bachellerie, J-P., and Raynal, F. (1983) Nucleic Acids Research 11, 3375-3391.

Moss, T. (1983) Nature 302, 223-228.

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

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

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

Otsuka, T., Nomiyama, H., Yoshida, H., Kukita, T., Kuhara, S., and Sakaki, Y. (1983) Proc. Nat. Acad. Sci. USA 80, 3163-3167.

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

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

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

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

Perry, R. P., (1962) Proc. Nat. Acad. Sci. USA, 48, 2179-2186.

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

Putney, S. D., Benkovic, S. J., and Schimmel, P. R. (1981) Proc. Nat. Acad. Sci. USA 78, 7350-7354.

Raynal, F., Michot, B., and Bachellerie, J-P., (1984) FEBS Letters 167, 263-268.

Ross, A. and Brimacombe, R. (1979) Nature 281, 271-276.

Rubin, G. M., and 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. and Bayev, A. A. (1980) Nucleic Acids Research 8, 5779-5794.

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

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

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

Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H., and Roe, B. (1980) J. Mol. Biol. 143, 161-178.

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

Scherrer, K. and Darnell, J. E. (1962) Biochem. Biophys. Research Comm. 7, 486-490.

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

Sinclair, J. H. and Brown D. D. (1971) Biochemistry 10, 2761-2769.

Skryabin, K. G., Zakharyev, V. M., Rubstov, P. M. and Bayev, A. A. (1979a) Dokl. Akad. Nauk. USSR 247, 1275-1277.

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

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

Southern, E. M. (1975) J. Mol. Biol. 98, 503-517.

Speirs, J. and Birnstiel, M. L. (1974) J. Mol. Biol. 87, 237-256.

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

Steffensen, D. M. and Duffey, P. (1974) Nature 252, 741-743.

Stewart, M. A. (1983) Ph.D. Thesis, University of Glasgow.

Stewart, M. A., Hall, L. M. C. and Maden, B. E. H. (1983) Nucleic Acids Research 11, 629-646.

Stiegler, P., Carbon, P., Ebel, J-P., and Ehresmann, C. (1981) Eur. J. Biochem. 120, 487-495.

Subrahmanyam, C. S., Cassidy, B., Busch, H. and Rothblum. L. I. (1982) Nucleic Acids Research 10, 3667-3680.

Torczynski, R., Bollon, A. P. and Fuke, M. (1983) Nucleic Acids Research 11, 4879-4890.

Valenzuela, P., Bell, G. I., Venegas, A., Sewell, E. T., Maziarz, F. R., Degennaro, L. J., Weinberg, F. and Rutter, W. J. (1977) J. Biol. Chem. 252, 8126-8135.

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

Veldman, G. M., Klootwijk, J., van Heerikhuizen, H., and Planta, R. J. (1981) Nucleic Acids Research 9, 4847-4862.

Ware, V. C., Tague, B. W., Clark, C. G., Gourse, R. L., Brand, R. C. and Gerbi, S. A. (1983) Nucleic Acids Research 11, 7795-7817.

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

Wellauer, P. K. and Dawid, I. B. (1973) Proc. Nat. Acad. Sci. USA 70, 2827-2831.

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

Wellauer, P. K. and Dawid, I. B. (1979) J. Mol. Biol. 128, 289-303.

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

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

Wellauer, P. K. and Reeder, R. H. (1975) J. Mol. Biol. 94, 151-161.

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

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

Willems, M., Wagner, E., Laing, R. and Penman, S. (1968) J. Mol. Biol. 32, 211-220.

Wilson, G. N. (1982) in The Cell Nucleus, Vol. 10, Eds. Busch, H. and Rothblum, L., pp. 287-318, Academic Press, New York.

Wilson, G. N., Hollar, B. A., Waterson, J. R., Schmickel, R. D. (1978) Proc. Nat. Acad. Sci. USA 75, 5367-5371.

Wilson, G. N., Szura, L. L., Rushford, C., Jackson, D., and Erickson, J. (1982) Am. J. Hum. Genet. 34, 32-49.

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

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

Zwieb, C., Glotz, C. and Brimacombe, R. (1981) Nucleic Acids Research 9, 3621-3640.

* Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) Proc. Nat. Acad. Sci. USA. 80, 3963-3965.