MICROHETEROGENEITY IN THE TRANSCRIBED SPACERS OF RIBOSOMAL DNA FROM *XENOPUS LAEVIS*.

Monica A. Stewart

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

Department of Biochemistry. October, 1983.

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

ACKNOWLEDGEMENTS

I would like to thank Dr. Ted Maden for his excellent advice, helpful comments and gentle criticism, throughout the course of this work.

My thanks are also due to Professor R.M.S. Smellie for making the facilities of the department of Biochemistry available for this research. Dr. Chris Darnborough for invaluable assistance in the dissection of *X. laevis*, and Miss Angela Nicol, of the Department of Zoology, for looking after my *X. laevis*. Ian Ramsden of Medical Illustration for photography and Mrs. Anne Mosson for her skilful typing of this thesis.

Special thanks go to all my colleagues in C13 who saw me through both good and bad days and finally, my husband Graeme, the calming force behind the storm - thank-you.

ABBREVIATIONS

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

rRNA	ribosomal RNA
rDNA	genes for ribosomal RNA and associated spacer DNA
RTU	ribosomal transcription unit
pre-rRNA	ribosomal precursor RNA
ITS	internal transcribed spacer
ETS	external transcribed spacer
NTS	non-transcribed spacer
IVS	intervening sequence
bp	base pairs
kb	kilobases (1000 bases)
dNTP	deoxynucleotide-5'-triphosphate
ddNTP	2',3'-dideoxynucleotide-5'-triphosphate
RF	replicative form
cccDNA	covalently closed circular DNA
SDS	sodium dodecylsulphate
PPO	2,5-diphenyloxazole
BCIG	5-bromo-4-chloro-3-indoyl-β-galactoside
IPTG	isopropyl-beta-D-thio-galactopyranoside

CONTENTS

FIGURES AND TABLES

SUMMARY

I have examined the transcribed spacers of Xenopus laevis ribosomal DNA, by direct DNA sequence analysis, for sites of sequence variation. Substantial microheterogeneity was identified in all three transcribed spacers. The ribosomal DNA sources examined included cloned and uncloned amplified ribosomal DNA and cloned chromosomal ribosomal DNA. In addition, amplified rDNA clones constructed from ribosomal DNA isolated and purified from a single Xenopus laevis were studied. The regions covered in DNA from these sources covered the 3' end of the external transcribed spacer and the whole of both internal transcribed spacers - an area comprising approximately 1,100 nucleotides of the ribosomal DNA unit structure. Some twenty sites at which variants occur have been identified. These variants include single base changes and length variants of one to several nucleotides. In addition, a site in the first internal transcribed spacer has been shown to vary by the presence or absence of a block of twenty nucleotides.

The identification of sequence variants of the same type in both cloned and uncloned oocyte ribosomal DNA confirms that heterogeneities in cloned ribosomal DNA are not artefacts introduced during cloning. Comparison of the sequence data for amplified and chromosomal ribosomal DNA revealed very good correspondence between the two sources of DNA. This indicates that chromosomal ribosomal DNA serves as the primary reservoir of variant sequences.

Sequence variation at the level of single nucleolar organisers was inferred in oocyte ribosomal DNA isolated from a single frog. This confirms that variants occur on physically linked repeats within individual chromosomes.

The findings of the transcribed spacers contrast strikingly with those for the ribosomal RNA coding sequences in which no sites of sequence variation were found in an extensive survey of the 18S coding region (Maden et al, 1982b). In addition, no variants were found in the 5.8S coding region in the analysis of cloned amplified and chromosomal ribosomal DNA and in a partial analysis of uncloned amplified ribosomal DNA. A single site of variation was identified near the 5' region of the 28S gene in one amplified and one chromosomal clone.

The results of this study provide novel data relevant to the evolution of ribosomal DNA, especially the transcribed spacer regions, and leads to speculation about the function of these spacer regions in ribosome formation.

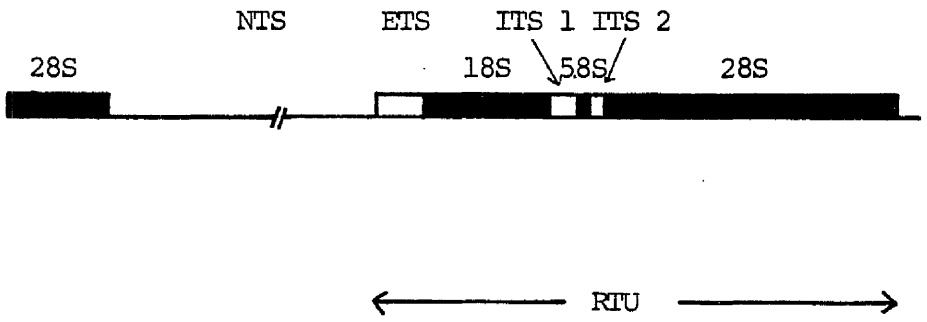


Figure 1.1 rDNA repeating structure in X.laevis with ribosomal transcription unit (RTU) illustrated. NTS, ETS and ITS denote non-transcribed, external transcribed and internal transcribed spacers respectively.

1.1 General Introduction

The aim of this project has been to examine the transcribed spacers of Xenopus laevis ribosomal DNA (rDNA) for sites of microheterogeneity. Transcribed spacers are regions of rDNA which are transcribed into ribosomal precursor RNA (pre-rRNA) and are subsequently eliminated during ribosome maturation (see Figure 1.1). It might be supposed that the transcribed spacers perform some specific role in ribosome maturation. However, a variety of studies culminating in recent sequence analyses have shown that transcribed spacers vary very greatly in primary structure between different eukaryotes, even between quite closely related species. Knowledge of the extent and nature of microheterogeneity in the transcribed spacers within a single species might provide insight into the genetic mechanisms contributing to phylogenetic variation in these regions and may also contribute to an understanding of the function (if any) of these regions in ribosome maturation.

In the remainder of this introduction, the background information on rRNA and its genes which led to the present investigation will be described.

1.2 The Ribosomal Transcription Unit

1.2.1 General features

In both prokaryotes and eukaryotes the ribosomal transcription unit (RTU) contains the rRNA coding sequences and regions of transcribed spacer. In X. laevis, the RTU codes for the 18S, 5.8S and 28S rRNA genes and 3 transcribed spacers, the ETS, ITS1 and ITS2 (Figure 1.1). The order of

the rRNA sequences is the same in eukaryotes and prokaryotes (Lewin, 1976) : a 5' ETS sequence followed by the smaller rRNA component which is succeeded by a transcribed spacer region. The large ribosomal component lies at, or close to, the 3' end. The genes coding for 5S RNA are distinct from the rRNA genes in most eukaryotes. The only exceptions known are in Saccharomyces (Rubin et al, 1973) and Dictostelium (Maizels, 1976). In bacteria, each transcription unit contains genes for 16S, 23S and 5S RNA and in addition some tRNA genes. To date, no tRNA genes have been found in the RTU of any eukaryotic organisms.

1.2.2 Structure of ribosomal precursor RNA

Transcription of the RTU results in the formation of pre-rRNA. The existence of pre-rRNA containing both the large and small ribosome components was first demonstrated by pulse labelling experiments of HeLa cells (Perry, 1962; Scherrer and Darnell, 1962). Subsequent partial sequence analysis of pre-rRNAs (Jeanteur et al, 1968) revealed the presence of non-ribosomal type sequences which are lost in the conversion of the precursors to mature rRNA. Studies on the base composition of the 45S pre-rRNA in the HeLa cell nucleolus (Willems et al, 1968) and purified 28S and 18S rRNA (Amaldi and Attardi, 1968) showed the precursor molecule to have a higher G + C content than a putative precursor consisting solely of one 28S and one 18S RNA molecule. These studies implied that the discarded RNA is exceedingly rich in G + C.

The location and approximate size of these non-ribosomal

sequences was first identified by electron microscopic studies (Wellauer and Dawid, 1973) although orientation of the rRNA molecules with respect to transcription was incorrectly defined. Comparison of the secondary structure maps of 45S pre-rRNA, mature 18S and 28S rRNA and various intermediate pre-rRNAs revealed the presence of two transcribed spacer regions, one separating the 18S and 28S genes, the second region being located 5' to the 18S gene. Similar maps were constructed for X. laevis 40S pre-rRNA (Wellauer and Dawid, 1974). The transcribed spacer regions became known as the external and internal transcribed spacers. The arrangement of gene and transcribed spacer regions and of secondary structure loops were compared in mouse-L cell, HeLa cell and Xenopus rRNA (Wellauer et al, 1974a). The overall arrangement of regions and loop pattern was very similar in the rRNA from these three organisms. However, the sizes of the transcribed spacers and the shapes of the loops within the transcribed spacer regions varied greatly between species. Earlier studies (Loening, 1968; Perry et al, 1970) had shown that the molecular weight of 18S and 28S rRNA remains almost constant from lower to higher vertebrates whereas, the molecular weight of the RTU increases considerably. Electron microscopic studies by Schibler et al, (1975) extended those of Wellauer and Dawid (1974) and showed that this increase in size of pre-rRNA during vertebrate evolution was due principally to changes in length of both the internal and external transcribed spacers. The morphology and arrangement of loops in the spacer regions were very similar in closely related species but varied considerably between

different groups of animals.

The RTU of all eukaryotic organisms studied, contains, in addition to the 18S and 28S rRNAs a small RNA species (designated 5.8S) which in the mature ribosome is found hydrogen bonded to 28S rRNA. This RNA species was first characterised in HeLa cell ribosomes (Pene et al, 1968) and was thought to be transcribed as part of the same pre-rRNA containing the 18S and 28S rRNAs. This was verified by studies on the 5.8S RNA of X. laevis (Speirs and Birnstiel, 1974) and its location in the ribosomal pre-rRNA was tentatively designated as somewhere within the internal transcribed spacer region. The precise location of 5.8S has only recently been defined by direct DNA sequencing of the 18S-28S intergene region (Hall and Maden, 1980). This study revealed transcribed spacer regions on either side of 5.8S RNA. These regions were appropriately called ITS1 (the region between 18S and 5.8S) and ITS2 (the region between 5.8S and 28S).

1.2.3 Processing of pre-rRNA

The formation of the mature 18S, 28S and 5.8S rRNAs occurs through a series of processing events which follow a general temporal sequence, varying slightly from organism to organism. Understanding of the order in which these cleavage reactions take place has arisen from the ability to isolate and compare the (metastable) pre-rRNA molecules generated during processing. In E. coli, the sequences flanking the 16S (and 23S) rRNAs interact to form an extensive base-paired structure

which constitutes a processing site for RNase III. Subsequent cleavage by this enzyme releases the almost mature 16S (and 23S) rRNA. However, no regions of extensive base pairing between the ETS and ITS1 of X. laevis rRNA have been found (Maden et al, 1982a). The occurrence of distinct intermediates in ribosome maturation in X. laevis indicates that cleavage on the 5' and 3' side of 18S rRNA are kinetically separate events. It is generally accepted in eukaryotes that the ETS is removed first followed by elimination of the internal transcribed spacers. The precise role (if any) of the internal transcribed spacers during processing are as yet uncertain, although recent reports (Crouch et al, 1983; Bachellerie et al, 1983) suggest a possible specific role for ITS2 in the conversion of 32S rRNA to 28S rRNA and 5.8S rRNA.

1.3 Structure of rDNA

1.3.1 Multiplicity and chromosomal location of rDNA

Multiple copies of the rRNA genes occur in both prokaryotes and eukaryotes (Attardi and Amaldi, 1970). Early studies using mutants of X. laevis suggested the possible clustering of these copies at the nucleolus organiser site of a single chromosome (Brown and Gurdon, 1964). Homozygous anucleolate mutants (0-nucleoli) were unable to synthesise any rRNA, whereas heterozygotes containing one nucleolus only (1 - nu) regulated their wild type genes to produce twice as much 28S and 18S rRNA as the same genes when present in wild type (2 - nu) individuals. Hybridisation of rRNA with DNA isolated from wild type X. laevis (Brown and Weber, 1968)

demonstrated that multipe copies (~450) of the 28S and 18S genes were present for each haploid complement of chromosomes. Clustering was shown to occur at a single chromosome since very low levels of hybridisation were detected with DNA isolated from anucleolate mutants. This intense clustering on the chromosomal DNA and the unique bouyant density of rDNA (Birnstiel et al, 1968) have therefore made the isolation and subsequent analyses of the rRNA genes possible.

The cytochemical location of the rRNA genes on chromosomes is made possible by the ability of rDNA-containing regions to organise a nucleolus (nucleolus organising region, NOR). A simple cytological technique using ammoniacal silver nitrate allows a precise localisation of nucleolus organisers (Goodpasture and Bloom, 1975; Howell et al, 1975). Whereas in situ hybridisation indicates the location of rDNA (Brown and Weber, 1968), silver staining of NOR's reflects activity of ribosomal genes rather than just their presence. Cytochemical tests demonstrate that the silver nitrate binds neither to the rDNA or rRNA, but rather to acidic proteins associated with rRNA transcribed at the rDNA sites. In mammals several chromosomes may contain NOR's, in humans the number is five (Wellauer and Dawid, 1979). In both X. laevis and Saccharomyces cerevisae the rRNA genes are located on a single chromosome. In prokaryotes, the rRNA genes of Escherichia coli have been extensively studied and it is now certain that exactly seven rRNA transcription units exist in the E. coli genome. These units are not clustered at a single location but have been isolated from seven different locations.

1.3.2 The Non-transcribed spacer (NTS) of rDNA

Analysis of purified chromosomal and amplified rDNA from X. laevis (see below for description of different rDNA sources) by a variety of physical and chemical techniques demonstrated the existence of a transcribed region coding for the 40S pre-rRNA and a non-transcribed spacer region. The transcribed and non-transcribed segments form a single unit which is repeated in a tandem fashion within the NOR of a chromosome.

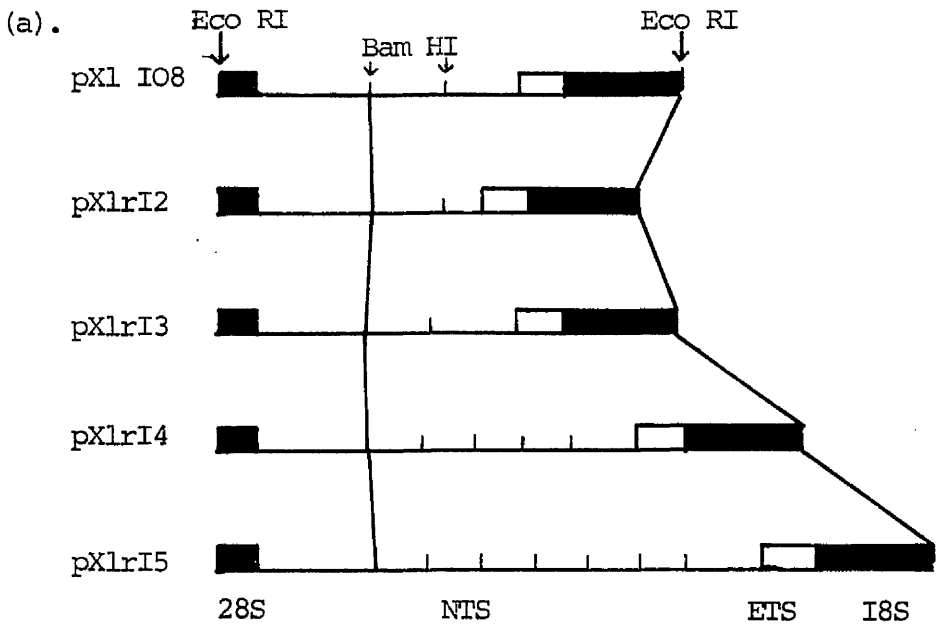
The NTS has been the subject of much detailed analysis. Restriction analysis and electron microscopic studies of amplified rDNA (Wellauer et al, 1974b) and electron microscopic studies on cloned NTS fragments (Wellauer et al, 1976a) showed that the NTS of X. laevis rDNA is heterogeneous in length and is made up of internally repetitious elements. The restriction mapping data of Botchan et al (1977) presented the first detailed analysis of an NTS confirming and extending the earlier studies of Wellauer et al (1974b, 1976a). NTS length heterogeneity in X. laevis rDNA is mainly due to a variable number of internally repetitious sequence elements. The precise organisation of these sequences has now been elucidated by DNA sequencing of a cloned ribosomal NTS (Boseley et al, 1979; Moss and Birnstiel, 1979; Moss et al, 1980). Length heterogeneity in X. laevis NTS manifests itself in the number of recognition sites for the enzyme Bam H1 (Figure 1.2a). This number can vary from two to seven sites in different repeats. Figure 1.2b illustrates the highly repetitive nature of the NTS. A major part of

Figure 1.2.

Length heterogeneity in the NTS of X. laevisrDNA.

(a) Length heterogeneity in a number of NTS-containing rDNA clones. The number of Bam HI recognition sites varies from two to seven (Botchan et al, (1977); Boseley et al, (1979)). Closed boxes denote rRNA coding regions. The open box denotes the external transcribed spacer.

(b) Location of repetitive regions (Rep. Reg.) and Bam-islands in the NTS of pXl108. The sequence around each Bam HI recognition site, the so called Bam-island sequence is the same. These Bam-islands, together with repetitive regions 2 and 3 form a Bam-HI superepeat. Variation in the number of these superepeats gives rise to the major NTS length variation observed in X. laevis rDNA. Boxes are described as above. The data are taken from Moss et al, (1980).


(a).
Eco RI
Bam HI
Eco RI
pXl IO8
pXlrI2
pXlrI3
pXlrI4
pXlrI5
28S
NTS
ETS
I8S

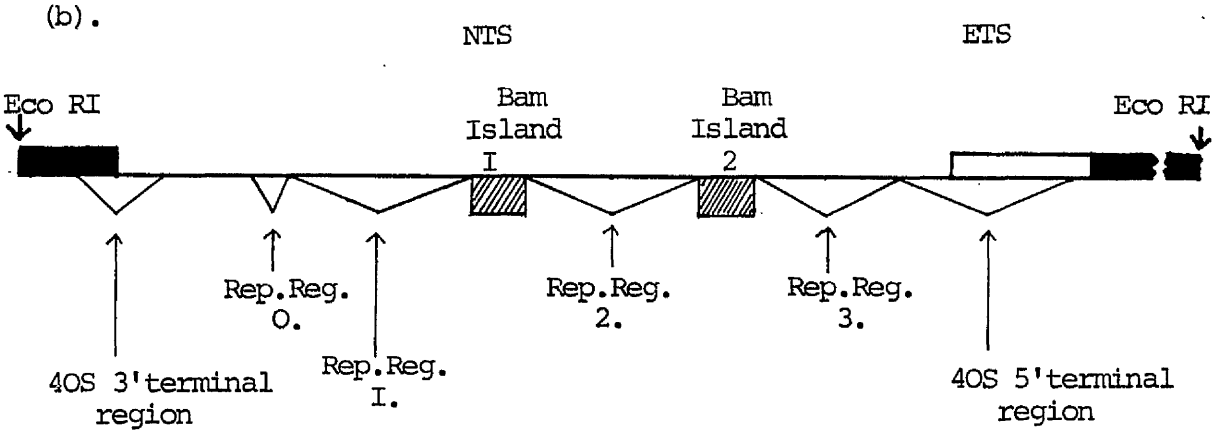

(b).
NTS
ETS
Eco RI
Bam Island I
Bam Island 2
Eco RI
Rep.Reg. O.
Rep.Reg. I.
Rep.Reg. 2.
Rep.Reg. 3.
40S 3'terminal region
40S 5'terminal region

Figure I.2

the NTS comprises a pattern in which sequences called "Bam-islands", which closely resemble the region preceding the site of initiation of transcription (Sollner-Webb and Reeder, 1979), are followed by repetitious regions with multiple Alu I recognition sites.

Recent studies on the detailed structure of D. melanogaster NTS has revealed a similar internally repetitive nature (Coen and Dover, 1982). The overall features of the sequence are thought to be common to several Drosophila species.

It is perhaps unfortunate that this region of the rDNA repeat unit is referred to as the non-transcribed spacer. The detailed studies previously mentioned for a X. laevis and D. melanogaster ribosomal spacer and in addition studies on X. borealis and X. clivii ribosomal spacers (Bach et al, 1981) have suggested an active transcriptional role for the NTS. These studies have shown that a common feature of all NTS regions is the reduplication of the pre-rRNA initiation site sequence. Moss (1983) has shown that the Bam-islands in the X. laevis NTS act as "sinks" for RNA polymerase I molecules. Transcription downstream of the Bam-islands results and is terminated at a site 213bp upstream of the 40S pre-rRNA initiation site. He suggests that this serves to increase the concentration of polymerase molecules at the 40S promoter and 40S initiation would thereby be stimulated or enhanced.

1.3.3 Intervening Sequences in rDNA

In eukaryotes, intervening sequences (IVS's) interrupt the coding regions of many genes specifying mRNA and tRNA (Abelson, 1979; Breathnach and Chambon, 1981). The sequences represented in the rDNA coding regions are usually continuous but in some cases IVS's are present in some or all of the 28S/26S genes. These interruptions have been characterised in a number of organisms e.g. *D. melanogaster* (Glover and Hogness, 1977) and *Tetrahymena thermophilia* (Cech and Rio, 1979).

These IVS's should not be confused with the transcribed spacer sequences found in pre-rRNA from which they differ in 2 respects. Firstly if the IVS's are transcribed at all then they are very short lived, unlike the metastable transcribed spacers. Secondly removal of an IVS requires the recognition of specific splice sites in the 28S gene by splicing enzymes followed by ligation of the gene ends to generate a covalently continuous 28S gene. Removal of the ETS and ITS from pre-rRNA does not result in ligation of the remaining fragments. Rather the individual mature rRNAs are generated.

1.4 Chromosomal and amplified rDNA in *Xenopus laevis*

Amplified rDNA has been used extensively in the detailed characterisation of *X. laevis* rDNA. Two types of rDNA can be purified from *X. laevis* : chromosomal and extrachromosomal amplified rDNA. During amplification (for review see Bird, 1980, and below) the rDNA content of the nucleus of a *Xenopus*

oocyte rises to about 30pg.compared to a total chromosomal DNA content of 12pg. (Macgregor, 1968). Thus almost three-quarters of the DNA in the oocyte nucleus is extrachromosomal. In contrast, chromosomal rDNA comprises about 0.2% of the total DNA of somatic cells of X. laevis (Birnstiel et al, 1966). Because of this large difference in the amount of chromosomal and amplified rDNA much higher yields of amplified rDNA can be purified from X. laevis oocytes than chromosomal rDNA from a somatic source. It is for this reason therefore, that so many of the detailed studies on rDNA in X. laevis have been carried out using amplified rDNA.

Before examining the validity of using amplified rDNA instead of chromosomal rDNA for such detailed characterisations I would first like to describe some of the molecular events which take place during amplification in X. laevis oocytes.

1.4.1 Initial events in rDNA amplification

Amplification of rDNA occurs in oocytes during early oogenesis in a wide variety of animals (Brown and Dawid, 1968, 1969; Gall, 1968, 1969; Lima-de-Faria et al, 1969; Vincent et al, 1969; Dawid et al, 1970). The mature oocyte of X.laevis contains about 4000X as much rDNA as is present in the haploid chromosome set (Brown and Dawid, 1968). This amplified rDNA is extrachromosomal and gives rise to multiple nucleoli within the large oocyte nucleus.

Morphological development of the amphibian oocyte begins with a series of rapid cell cleavages. During the cleavage stages a tremendous amount of protein synthesis occurs. In order to prepare for this rapid protein synthesis many ribosomes must be produced and stored in the oocyte. Thus this

massive amplification of the rRNA genes makes possible the intensive ribosome synthesis required for egg development.

Is amplified rDNA derived from chromosomal rDNA and if so, how is that first copy produced ?

Two alternative mechanisms were considered to try and answer the first part of this question. One possible mechanism, the episome inheritance mechanism, (Wallace et al, 1971; Brown and Blackler, 1972) proposed that amplified rDNA copies are never integrated into the chromosome but pass episomally between generations. Brown and Blackler (1972) also suggested a second mechanism (the chromosome copy mechanism) in which the first amplified copy would be a replica of one or more of the several 100 repeating gene sequences which comprise the chromosomal rDNA at the nucleolus organiser.

It is possible to distinguish between those two mechanisms by taking advantage of the different kinds of inheritance which each model predicts. The chromosomal copy mechanism requires nuclear inheritance of the rDNA, whereas the episome mechanism predicts maternal inheritance. Crosses between X. borealis and X. laevis were carried out (Brown and Blackler, 1972) and the amplified rDNA from female progeny was analysed. Only X. laevis rDNA was found to be amplified, whether the X. laevis parent was male or female. This indicates that the template for the amplified rDNA can be transmitted through the spermatocyte as well as the oocyte. Since sperm cells contain only chromosomal rDNA the findings lead to the conclusion that amplified rDNA is derived from the chromosomal genes. Once this was established, mechanisms whereby the rDNA can make the transition from the integrated to extrachromosomal state were considered.

One possibility involved the excision of a copy of the repeating sequence from the chromosome by a recombination event. This, however, has the difficulty that the number of chromosomal genes would be reduced

at each generation. Another mechanism (Tocchini-Valentini and Crippa, 1971) suggests that the first copy is produced via an RNA intermediate. In this model a transcript of the entire repeating sequence is produced by utilisation of the pre-existing transcription machinery and the transcript is then reverse transcribed to give a double stranded copy of the gene. A third mechanism (Sederoff and Tartoff, cited in Tartoff, 1975), conveniently named the "double-initiation model" by Bird (1980), involves bidirectional replication of the DNA to give a replicative bubble containing a single repeat unit. A second round of replication of the same replication bubble generates the extrachromosomal rDNA segment.

The latter two models have an advantage over the first in that they both leave the parental chromosome unchanged. To date however, this primary replication event in ribosomal gene amplification remains one of the major unresolved areas in this field.

1.4.2 Replication of amplified rDNA

The mechanism whereby the first amplified rDNA copy is replicated to give many thousands of copies is now well characterised. rDNA purified from amplifying ovaries revealed, by electron microscopy, that 2-5% of the molecules are circular (Hourcade *et al*, 1973). The contour lengths of the circles suggested that they were monomer, dimer, trimers etc. of the rDNA repeat unit. In addition a small proportion of the circular molecules examined contained tails. Because the circles and tailed circles were found to be only a small fraction of the total amplified rDNA analysed it was important to be sure that those molecules were, in fact, rDNA. Partial denaturation mapping revealed that these circular molecules displayed the same melting pattern characteristics of rDNA and

therefore represent rDNA.

The tailed circle configuration suggested that these molecules were rolling circle intermediates (Gilbert and Dressler, 1968) in rDNA replication. This conclusion was shown to be correct by incubating the amplifying ovaries with ^{3}H-deoxycytidine prior to electron microscopic autoradiography (Bird et al, 1973; Rochaix et al, 1974). These labelled molecules preferentially enter tailed circles, showing that they are indeed replication intermediates. In addition, the pattern of pulse and pulse-chase experiments conforms exactly with that expected for a rolling circle mechanism: As the pulse length is increased the labelled region in the tail grows from the replication fork towards the free end of the tail. Secondly after a pulse-chase, the labelled region is displaced towards the free end of the tail and no label remains associated with the rolling circle.

Studies by Wellauer et al (1976b) also support this rolling circle mechanism. In analysis of rDNA from individual frogs, the amplified rDNA showed little if any intramolecular length heterogeneity, whereas in chromosomal rDNA repeats of different lengths were scrambled to a considerable extent. The rolling circle mechanism would automatically generate such intramolecular length homogeneity in amplified rDNA.

1.4.3 Comparison of amplified and chromosomal rDNA

Dawid et al (1970) isolated chromosomal and amplified rDNA from X. laevis erythrocytes and ovaries respectively and compared them by a variety of physical and chemical methods. The comparison was undertaken because amplified rDNA was found to be of higher bouyant density than somatic rDNA (Brown and Dawid, 1968). The single difference observed in all the studies carried out between these DNAs was in their content of

5-methyldeoxycytidilic acid (5-MeC). Chromosomal rDNA has 4.5% of its residues as 5-MeC whereas amplified rDNA, in contrast, is not detectably methylated. The presence of 5-MeC is known to lower the density of DNA and the difference in bouyant density of the two DNAs was due entirely to the different methyl content.

Subsequent work by electron microscopy (Wensink and Brown, 1971; Wellauer et al, 1976b described above) has confirmed the gross similarity between chromosomal and amplified rDNA. However, see also the next paragraph.

1.4.4 How representative of chromosomal rDNA is amplified rDNA?

Analysis of rDNA from the whole ovary indicates that individual frogs show varying degrees of preferential amplification of rDNA units containing particular length variants (Wellauer et al, 1976b). This phenomenon was studied further by analysing the amplified rDNA from single oocytes (Bird, 1977). Digestion of Xenopus rDNA with Eco RI had shown clearly that the tandemly arranged repeat units are not homogenous in length, rather the large Eco RI fragments obtained were heterogeneous in size due to length variation in the NTS (Morrow et al, 1974; Wellauer et al, 1974b).

Amplified rDNA from individual oocytes from one female X. laevis was purified and restricted with Eco RI. The patterns obtained were compared with a Eco RI digest of blood cell rDNA purified from the same female. Analysis of the restriction products showed that each oocyte amplifies only a small subset of the repeats present chromosomally. Bird (1977) showed that individual oocytes differ in their choice of amplified subset.

In the frog studied (Bird, 1977) the spacers amplified most often were those most common in the somatic rDNA. However, studies by Wellauer *et al* (1976b) on the *whole* ovaries of individual frogs showed that some animals selectively amplified repeat lengths rarely found in their chromosomal rDNA whereas others amplify their most abundant size class.

Although individual oocytes and individual frogs show a preference to the repeat units amplified, when large preparations of amplified rDNA are made from many frogs, the frequencies of different length variants approximate those in chromosomal rDNA. Thus, such amplified rDNA preparations contain a representative and large sample of the rDNA units present in chromosomal rDNA. Therefore, on the basis of gross criteria, such as electron microscopy and restriction with Eco RI, the use of amplified rDNA in the detailed characterisation of *X. laevis* rDNA appears to be justified. However, some actual sequence studies on chromosomal rDNA would also be desirable and have been carried out as part of this project.

1.5 Interspecies Comparisons

The earliest comparison between the ribosomal gene repeating units of different species was made using purified rDNA from *X. laevis* and *X. borealis*. Brown *et al*, (1972) observed that the 18S and 28S sequences of both species were virtually indistinguishable whereas cross hybridisation involving the NTS regions showed that only a small fraction of the NTS sequences hybridised. Subsequent restriction

Species/rRNA gene	Reference
Human mitochondrial 16S, 12S	Eperon et al, (1980)
Mouse mitochondrial 16S, 12S	van Etten et al, (1981)
Zea mays chloroplast 16S	Schwarz & Kössel, (1980)
" " " 12S	Edwards & Kössel, (1981)
E. coli 23S	Brosius et al, (1980)
" " 16S	Brosius et al, (1978)
S. cerevisae 25S	Georgiev et al, (1981)
" " 17S	Rubstov et al, (1980)
X. laevis 18S	Salim & Maden (1981)
Rat 18S	Torczynski et al, (1983)

Table 1(a) Complete sequence data for some large and small subunit rRNAs.

	Size(bp)			G + C (%)		
Species	ETS	ITS1	ITS2	ETS	ITS1	ITS2
Drosophila	n.d.a.	n.d.a.	400	n.d.a.	n.d.a.	20
Yeast	652	360	230	43	35	38
X. laevis	712	557	262	83	84	88
X. borealis	613	554	336	81	76	88
Mouse	n.d.a.	999	1089	n.d.a.	70	74
Rat	n.d.a.	1067	765	n.d.a.	74	80

Table 1.1(b) Phylogenetic variability of transcribed spacer rDNA in relation to size and % G + C content.

n.d.a. means no complete sequence data are available

References : Drosophila: taken from Michot et al, (1983).
Yeast : for ETS, Skryabin et al, (1979a);
for ITS, Skryabin et al (1979b),
Veldman et al, (1980),(1981).
X.laevis : ETS, Maden et al, (1982a);
for ITS, Hall and Maden (1980).
X.borealis: ETS Furlong et al, (submitted for publication);
for ITS Furlong and Maden, (1983).
Mouse : Michot et al, (1983).
Rat : Subrahmanyam et al, (1982).

analyses and extensive electron microscopic studies (Wellauer et al, 1974b, 1976a, 1976b) of X. laevis rDNA revealed that the NTS shows considerable intraspecific length heterogeneity (the molecular basis of which is described in Section 1.3.2), whereas the RTU shows apparent length homogeneity. Heterologous nucleic acid hybridisation studies (Sinclair and Brown, 1971; Gerbi, 1976; Gourse and Gerbi, 1980) demonstrated that even over a wide range of eukaryotic species some degree of homology is observed in the rRNA coding regions.

With the advent of molecular cloning and rapid DNA sequencing techniques, large sections of the RTU in a number of species have been investigated in detail and sequence comparisons have been made. Comparative data are important in trying to identify regions of the RTU which may have some functional importance in, for example, ribosome formation.

1.5.1 Sequence comparisons in the rRNA coding regions

Complete sequence data are now available for a number of eukaryotic, prokaryotic and organelle ribosomal genes (see Table 1(a)). Comparison of the small rRNA reveals a high degree of homology between prokaryote/organelle rRNAs and between eukaryote/eukaryote rRNAs. Homology between prokaryote 16S/eukaryote 18S rRNA is much lower although interspersed throughout the sequences are many short tracts of clear homology (Zwieb et al, 1981; Stiegler et al, 1981). In addition to the primary structure, secondary structures have been defined for prokaryotic 16S rRNA (Woese et al, 1980; Glotz and Brimacombe, 1980) and eukaryotic 18S rRNA (Zwieb et al, 1981; Stiegler et al, 1981). One of the most important conclusions

to emerge from these models is that the secondary structure of prokaryote, eukaryote and mitochondrial small subunit rRNA all conform to a common plan. All of the rRNAs consist of four domains, each domain bound by a short helix. Each of these helices is formed by a long range interaction between rRNA regions that are distant from each other in the primary structure.

Similar studies have been carried out on the large subunit rRNA although they are not so extensive because of the limited amount of sequence data. From all these studies, however, it appears that in certain regions of the gene, the nucleotide sequence is of utmost importance and must be under intense selective pressure to remain conserved; in other regions mutations can accumulate gradually but the ability to form a particular secondary structure must be maintained.

1.5.2 Sequence comparisons in the transcribed spacer regions

Whereas the mature rRNAs have been strongly conserved during evolution, ribosomal transcribed spacer regions have evolved more rapidly and show dramatic differences in size and G + C content among distant eukaryotes (Table 1(b)).

Complete sequence data on the 18S-28S intergene region are available now from yeast (Skryabin *et al*, 1979b; Veldman *et al*, 1980, 1981) *Xenopus* (Hall and Maden, 1980; Furlong and Maden, 1983), Mouse (Michot *et al*, 1983) and rat (Subramanyam *et al*, 1982). Sequence comparison reveals little homology between the phylogenetically distant species.

Nucleic acid hybridisation (Brown et al, 1972; Forsheit et al, 1974) showed that the transcribed spacers of X. laevis and X. borealis were comparable in length in the two species but that their nucleotide sequences differed. Comparison of the actual transcribed spacer sequences confirms these findings. (Furlong and Maden, 1983; Furlong et al, submitted for publication). Both the ETS and the ITS, although differing extensively, do contain short tracts of completely, or almost completely, conserved sequences in both regions. Similar findings have been reported for the ITS1 of mouse and rat (Michot et al, 1983). It is interesting to observe that some of the conserved tracts found in Xenopus show 100% homology with specific conserved tracts in the mouse ITS. Considering the overall lack of homology in the ITS, it is surprising to find such conservation, although restricted to small regions, between these distantly related species.

Thus the transcribed spacer sequences show a high degree of evolutionary fluidity and appear to be under little selective pressure to maintain primary structure.

1.6 Why search for heterogeneities in the transcribed spacers?

In the introductory paragraph to this thesis, I suggested that searching for microheterogeneity in the transcribed spacers of X. laevis rDNA might help to improve our understanding of how these spacer regions have diverged so extensively even between closely related species.

Classically, the rDNA repeat units within a single species have been considered to contain two types of sequence : a highly variable NTS region which shows considerable length

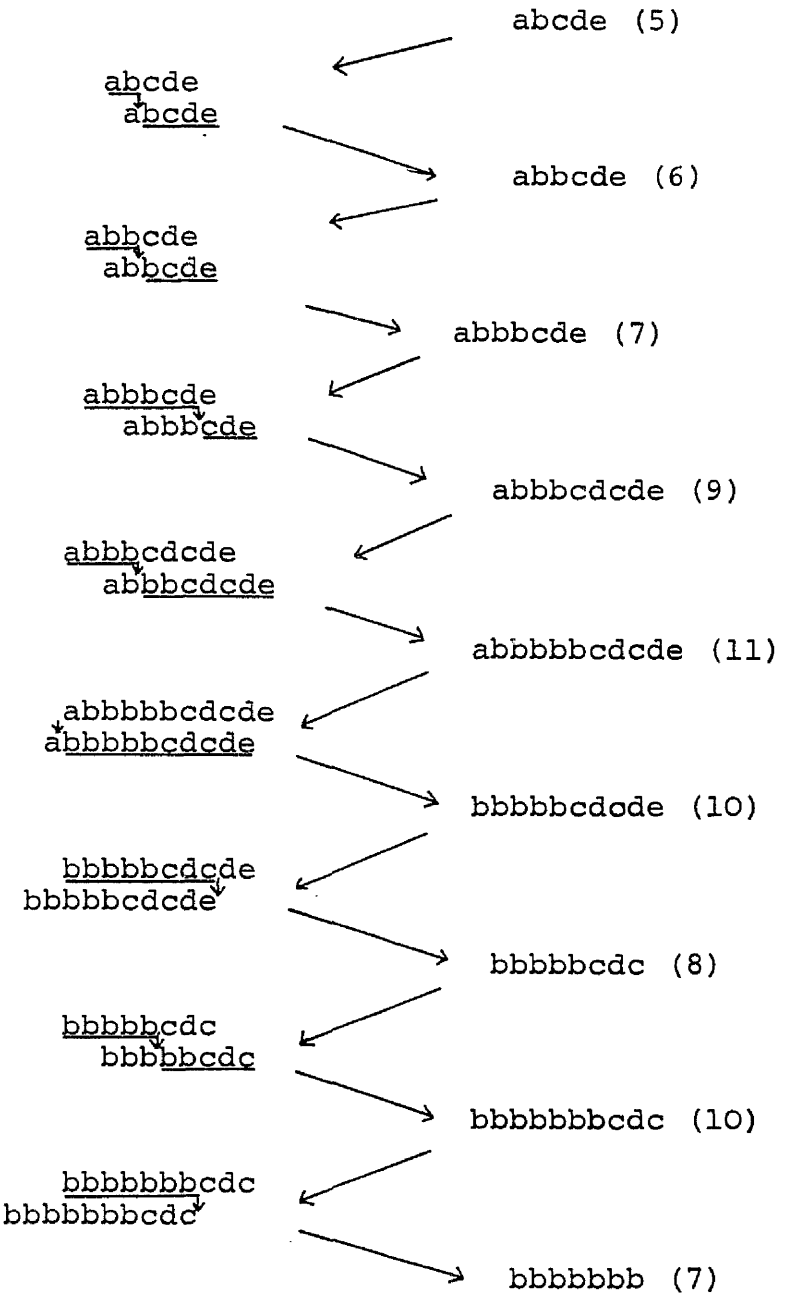


Figure 1.3 Schematic illustration of Unequal Recombination.

Each letter represents an individual unit in a tandem array. When two recombining sequences align out of register with respect to the total array length, a genetic crossover may change the size and constitution of the array. Although reciprocal products are generated by each event, the fate of only one is shown. Crossovers within the array generate deletions and duplications resulting in an increase or decrease (as indicated by the numbers in brackets) of the size of the array. Eventually one sequence (b) predominates. In this way a unit containing a new mutant will either be lost through deletion or will be duplicated and gradually encompass the entire array.

heterogeneity and a highly constant transcribed region which, by restriction analysis, shows no length variation. When different species are compared, both the non-transcribed and transcribed spacers are highly variable and only the ribosomal coding regions are relatively highly conserved. How can different parts of the rDNA repeat unit evolve at different rates while individual members of the rDNA multigene family appear to evolve in parallel to retain homogeneity within species, at least in their transcribed regions ?

Unequal recombination provides one mechanism whereby such"concerted evolution" might occur (Smith 1973, 1976). Suppose a mutation occurs at a site in one member of a multigene family. After several cycles of replication and unequal recombination, a number of progeny bearing different numbers of copies of the mutation in the multigene family will now be present in the animal population (Figure 1.3). Such unequal crossovers, which must occur in the germ line for their effect to be propagated into new generations of animals, might happen at meiosis or at any one of the many germ line mitoses. Computer simulation shows that the result of successive random crossovers is to eventually impose uniformity on the region involved due to the random increase of one variant and decrease of the other. Any new variants created by mutation will either be eliminated or expand to occupy the entire array. This process Smith named "crossover fixation".

Smith also suggests that different regions of the repeat might differ in their respective rates of acceptance of mutations. If a deleterious mutation were introduced into a gene region then natural selection would rapidly eliminate

chromosomes in which a repeat harbouring such a mutation is on the way to fixation. On the other hand mutations in the spacer portion might have no damaging effect, would not be eliminated by selection, and thereby would be able to achieve fixation.

A further feature of the model is that short tandem repeats can evolve to occupy any region of DNA that is not subject to selective pressure. This may partly explain the internally repetitive nature of both Xenopus and Drosophils NTS (However, see also the preceding discussion on a possible role of the NTS for RNA polymerase loading). The currently recognised source of intraspecific variation among repeat units is length variation caused by different numbers of NTS sub-repeats. Unequal exchange can account for this since out of register crossover within repetitious internal segments of a complete repeat would automatically generate length variants.

How has interspecific variation arisen in the transcribed spacers ? Clearly many individual sequence variants must have arisen and become established in the respective rDNA multigene families to generate the observed degree of sequence divergence between the transcribed spacers of closely related species (e.g. X. laevis and X. borealis : Furlong and Maden, 1983, Mouse and rat : Michot et al, 1983). Restriction analysis of the transcription unit of several Drosophila species of the Melanogaster group (Coen et al, 1982) has revealed that each species has only one predominant ITS length implying intraspecies homogeneity. However, the degree of homogeneity/heterogeneity in the transcribed spacer cannot

be critically assessed by such techniques and has never been examined at the nucleotide sequence level in a single species. It is in fact possible that regions of the transcribed spacers may contain small scale variants, not detectable by restriction analysis, which are in the process of undergoing fixation or elimination from the multigene family. The aim of this project was to determine whether any such microheterogeneity exists within the transcribed spacers of X. laevis rDNA.

2. Strategy of Research

2.1 Starting point

At the outset of this study the complete nucleotide sequence of the 18S-28S intergene region of a rDNA clone called pX1r101 (See Figure 2.1) was known (Hall and Maden, 1980). In addition a small amount of sequence data was available from a different clone pX1212 (see Figure 2.1) in which the 5.8S gene and flanking regions of the two internal transcribed spacers had been analysed (Boseley et al, 1978). Comparison of the sequence data revealed several apparent sites of sequence variation between the respective parts of the two ITS regions thus providing a focal point for the start of my search for variants. In fact these apparent sequence differences were not confirmed - see Chapter 4.

To undertake systematic screening of the transcribed spacers of X. laevis rDNA for microheterogeneities, I carried out extensive DNA sequencing on X. laevis rDNA from several sources using the chemical method of Maxam and Gilbert (1980) and the "dideoxy" method of Sanger et al (1977).

2.2 Sources of rDNA

The rRNA coding sequences were the first eukaryotic genes to be cloned in E. coli (Morrow et al, 1974). Since those initial experiments many further rDNA clones have been constructed. Figure 2.1 illustrates the different plasmid based rDNA clones which I have used during the course of this study. Some clones contain a complete repeat unit of rDNA (e.g. pX1101-103, pxlcr1-5) whereas others contain only a segment of the rDNA repeating structure (e.g. pX1212, pX1108). Moreover, the clones illustrated in Figure 2.1 fall into two categories with respect to the rDNA source; clones which contain amplified rDNA

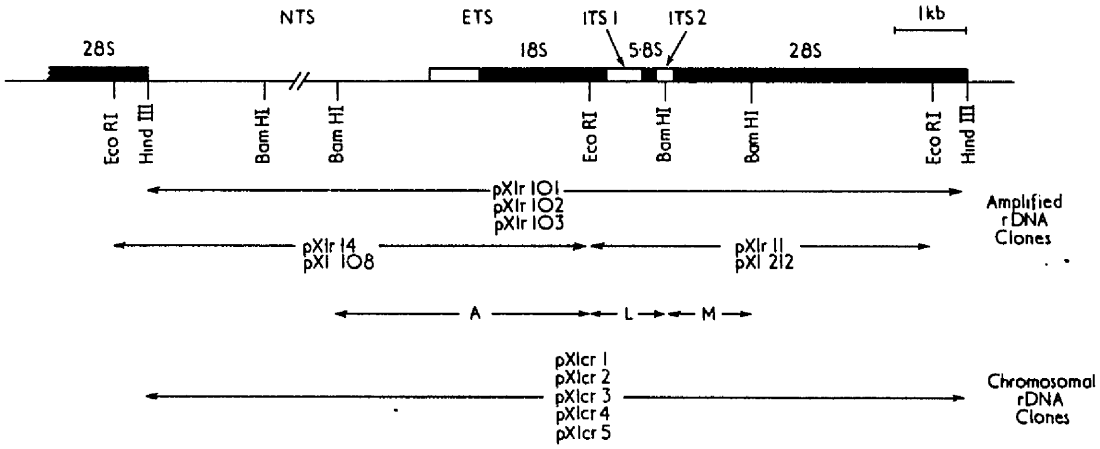


Figure 2.1. Description of the plasmid-based rDNA clones used in this study.

The structure of rDNA from X.laevis showing slightly more than one unit of repeating structure is illustrated. The restriction sites used for cloning and the rDNA fragment contained in each clone is shown. The number and spacing of Bam HI sites in the NTS differs between different clones. Subclones containing regions A, L and M from some of the clones were used in this work.

purified from X. laevis oocytes and clones which contain chromosomal rDNA purified from X. laevis erythrocytes. These latter clones were constructed during this project (see Chapter 6). The original rDNA (either amplified or chromosomal) was isolated and purified after pooling from several frogs.

Uncloned amplified rDNA was also studied. This rDNA was part of a large preparation from the ovaries of 50-100 small frogs. It was purified by banding in CsCl according to Dawid et al (1970) and was donated by D.D. Brown.

Finally amplified rDNA, isolated and purified from a single X. laevis frog as part of this project, was also analysed.

2.3 DNA Sequencing

The two most commonly used rapid methods of DNA sequencing are the chemical method of Maxam and Gilbert and a primed synthesis method using chain-terminating inhibitors.

2.3.1 Chemical method

In the chemical method end-labelled DNA is partially cleaved at each of four nucleotide bases in four different reactions following by chain cleavage with piperidine. The products are ordered by size by gel electrophoresis and the sequence is read off an autoradiograph by noting which base-specific agent induced cleavage at each successive nucleotide along the strand. Figure 2.2 describes the general steps involved in this method. Either single or double stranded DNA uniquely labelled at either the 5' or 3' end can be sequenced.

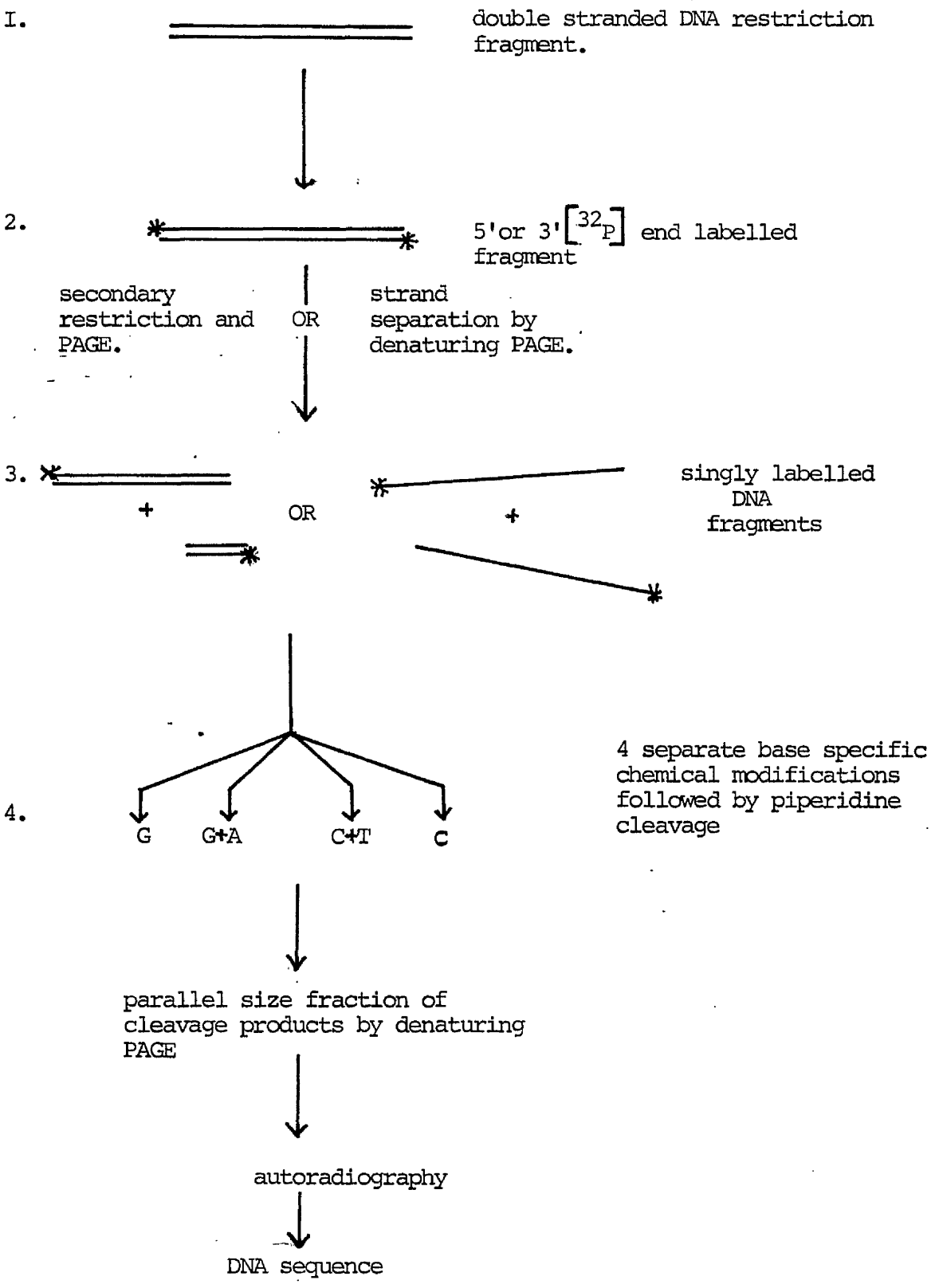


Figure 2.2. Summary of the general steps in the Maxam and Gilbert method of DNA sequencing.

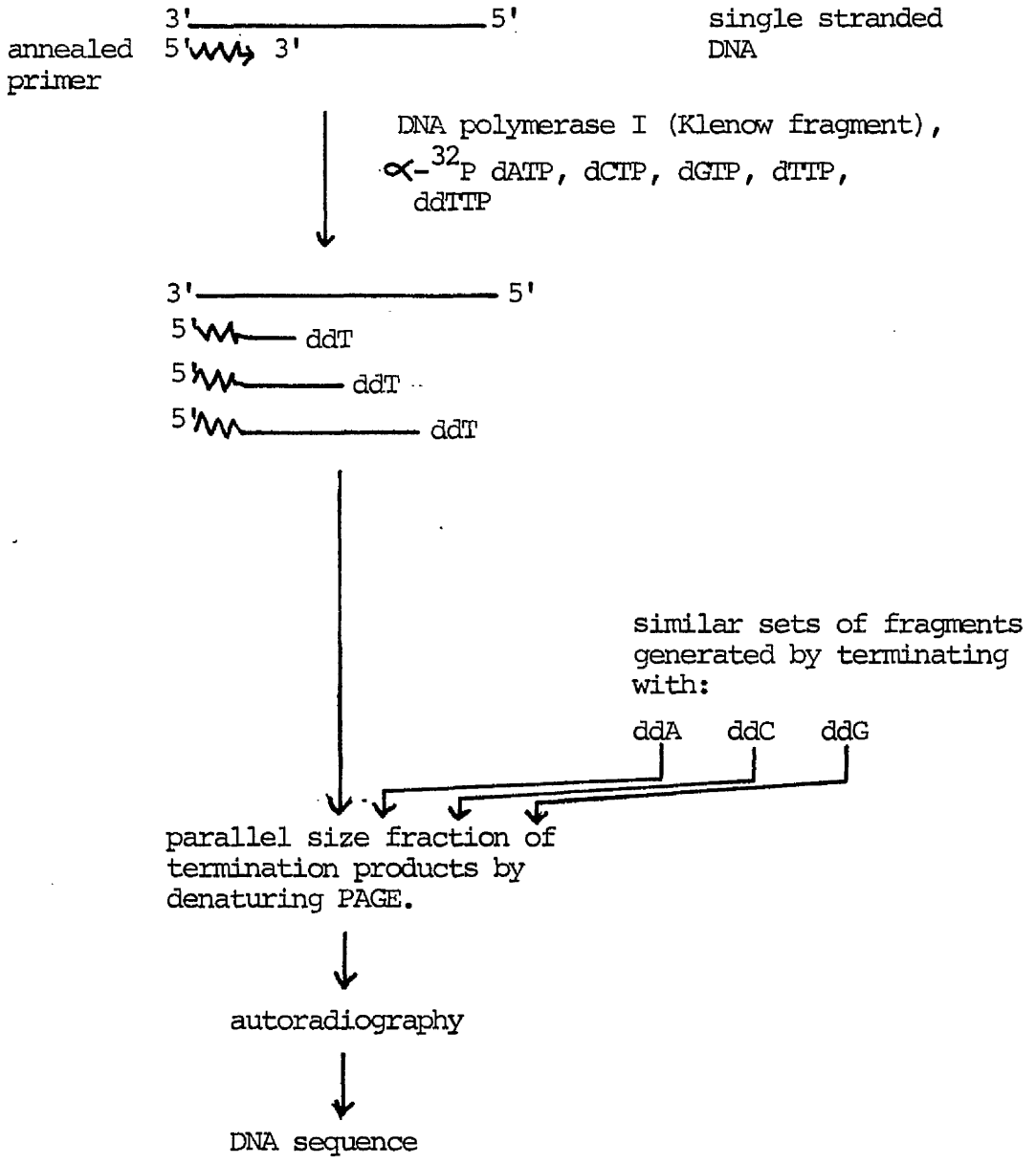


Figure 2.3. Summary of the general steps involved in the "dideoxy" method of DNA sequencing.

2.3.2 Primed synthesis

In the primed synthesis approach a single-stranded template containing the sequence of interest is enzymatically copied to produce the radioactively labelled complementary strand. By using chain-terminating inhibitors e.g. 2', 3'-dideoxynucleoside-5'-triphosphates (ddNTP's) sets of partially elongated molecules are produced which can be fractionated on denaturing gels and the pattern of the labelled bands indicates the sequence (Figure 2.3).

Two disadvantages have limited the use of this method in the past ; the requirement that the DNA to be sequenced is in a single stranded form and that for every 200-300 bases of sequence a specific primer, complementary to the 3' end of the template must be prepared. However, recent genetic manipulations of the bacteriophage M13 into a cloning vehicle (Sanger et al, 1980; Messing et al, 1981) and the construction of a "universal" primer have overcome these problems. In order to understand the relationship between the M13 cloning system and the dideoxy sequencing method it is necessary to understand a little of the phage biology and also how the various M13 cloning vectors were developed.

2.3.3 The biology of M13

M13 is a single stranded DNA-containing, filamentous bacteriophage which is specific for male E. coli. Infected cells contain the double stranded replicative form (RF) of the DNA from which new single stranded DNA is synthesised. This newly synthesised DNA strand (+ strand) is packaged into the protein coat and the M13 phage is extruded out of the cell without lysis. The RF DNA can be purified from infected cells. If foreign DNA is ligated in vitro into the RF and the recombinant DNA is used to transfect E. coli, the phage life cycle provides an easily

obtained source of single stranded DNA which can serve as the template in the primed synthesis method.

2.3.4 The development of M13 vectors

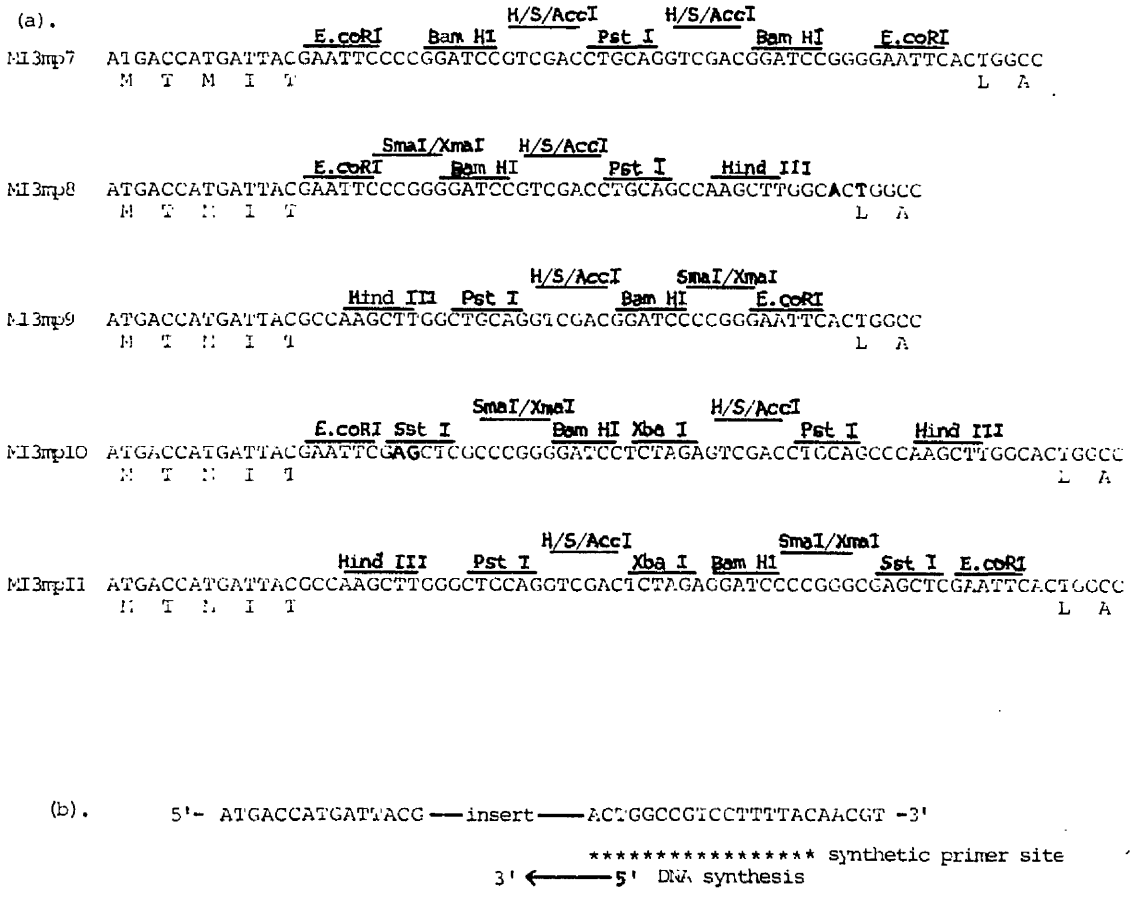
Clever genetic engineering of wild type M13 has converted this phage into a versatile cloning vehicle. M13mp2 was the first engineered M13 cloning vector. It was constructed by the insertion of a Hind II restriction fragment, derived from the lac operon of E. coli, into the intragenic space of wild type M13 phage genome (Messing et al, 1977). The inserted fragment contains the DNA sequence for the promoter, operator and the first 145 amino acid residues of the β-galactosidase gene (the so called α-peptide). Within this gene a single site for the restriction enzyme Eco RI was introduced yielding the recombinant M13mp2 (Gronenborn and Messing, 1978). The α-peptide when induced (by IPTG: isopropyl-beta-D-thio-galactopyranoside) is not a functional β-galactosidase. Moreover the host E. coli JM103 carries a deletion of the lac-promoter region from the chromosome and this region is carried instead on the male episome. The episomal β-galactosidase is also defective, missing the sequence encoding amino acid residues 11-41. When M13mp2 invades the host the two defective β-galactosidase polypeptides associate and complement each other and produce a functional β-galactosidase. Plaques of infected cells therefore produce a functional β-galactosidase which is readily identified by including a lactose analogue BCIG (5-bromo-4-chloro-3-indoyl-β-galactoside) on the plate. This colourless compound is hydrolysed by β-galactosidase to a blue dye, bromo-chloroindole and confers a blue perimeter on plaques of M13mp2. By contrast, recombinant plaques, in which foreign DNA has been inserted at the unique cloning site within the α-peptide of β-galactosidase, yield white plaques. The inserted DNA interrupts the α-peptide and destroys its complementation of the defective episomal β-galactosidase.

The flexibility of the M13 cloning system has been greatly improved by the insertion of synthetic oligonucleotides containing an array of restriction sites into the original Eco RI site of M13mp2 (see Figure 2.4a). The insertion of these additional sequences does not affect the ability of the lac gene product to undergo intra cistronic complementation. The development of M13mp8 and 9 and M13mp 10 and 11 has been especially useful since these two sets of vectors allow asymmetric cloning of restriction fragments.

Sequencing the inserted DNA by the chain terminator method requires both a single stranded template and a primer complementary to the 3' end of the region. M13 biology provides the source of single stranded DNA which can be rapidly prepared from the recombinant molecules (see Chapter 3, Section 5.). Since all fragments which are to be sequenced are cloned into the same specific region of the M13 genome it was possible to develop a "universal" primer for sequencing through the recombinant region. Such chemically synthesised oligonucleotides, complementary to the region 3' to the cloning site, have been developed (Anderson et al, 1980; Heidecker et al, 1980; Messing et al, 1981; Duckworth et al, 1981). For sequencing, the primer is annealed to the complementary region in the single stranded template and this primes the 5'⟶ 3' synthesis of the complementary strand of the insert DNA (Figure 2.4b).

2.4 Application of sequencing methods in the search for transcribed spacer variants

The four major steps involved in the chemical DNA sequencing method (see Figure 2.2) are based on the following. The starting material is double stranded DNA. Generation of the required DNA fragments and separation of the labelled ends is most often achieved by restriction enzyme digestion (followed by gel electrophoresis). Knowledge of



(b). 5'- ATGACCATGATTACG —— insert —— ACTGGCCGTCGTTTTACAACGT -3'

********** synthetic primer site

3' ←——— 5' DNA synthesis

Figure 2.4.

(a). Description of some recently genetically engineered M13 vectors each containing an array of restriction endonuclease sites. The + strand is shown.

(b). Annealing site of 17-mer synthetic primer used in this dideoxy sequencing analysis.

restriction sites in the DNA to be sequenced is therefore pre-requisite. Because of the availability of many rDNA clones with detailed restriction data (e.g. pXlrl01 : Hall and Maden, 1980; Salim and Maden, 1981; Maden et al, 1982a,b. pXl108 and pXl212 : Boseley et al, 1978, 1979) and the fact that this sequencing system was well established in our laboratory, I chose the Maxam and Gilbert method for my initial search of transcribed spacer sequence variants.

The entire 18S-28S intergene region of the clone pXl212 was determined. In order to screen other rDNA clones rapidly only partial sequence analysis was performed. The transcribed spacers of X. laevis rDNA are extremely GC rich so either the G or C cleavage reaction was carried out. The resulting G or C cleavage patterns were compared among the different rDNAs screened. If variation(s) in the patterns arose, a full sequencing experiment was performed on the rDNAs of interest.

My final sequencing objective in this project was to analyse rDNA isolated from a single frog. I wished to study a specific region of the rDNA repeat unit, namely ITS1, by screening as many ITS1 containing clones as possible. It became clear that the most efficient and quickest way of generating such clones and analysing them was by using the M13 cloning system in conjunction with the "dideoxy" sequencing method.

The main advantage of this system over the chemical method is that once clones are constructed, the biology of M13 allows single stranded DNA to be obtained and purified quickly (1 day). Rapid screening of many clones in parallel is achieved by performing one of the dideoxy reactions, and both sequencing and gel electrophoresis can be performed on the same day. Therefore, within 24 hrs of obtaining ssDNA, comparative sequence data on clones is beginning to accumulate. By contrast, in the Maxam and Gilbert method the DNAs to be screened must

first be excised and purified from the parent clone prior to sequencing (steps 1-4, Figure 2.2). Sequencing and gel electrophoresis cannot be performed on the same day. Thus the overall time taken to obtain sequence information is much longer than with the M13 system.

However, use of these two sequencing techniques in succession has provided an ideal way of achieving my sequencing objective.

3. Materials and Methods

Part A : Materials

1. Chemicals

Unless otherwise specified all chemicals were Analar grade supplied by BDH Chemicals Ltd. or Fisons Scientific Apparatus. Where chemicals or equipment were obtained from other sources this is indicated in the text and a list of the names and addresses of the suppliers is given below.

2. Suppliers

Amersham International plc, Amersham, Bucks., England

Aldrich Chemical Co., Gillingham, Dorset, England

BDH Chemicals Ltd., Poole, Dorset, England

A & J Beveridge Ltd., Edinburgh, Scotland

Beckman Instruments Inc., High Wycombe, Bucks., England

Bethesda Research Laboratories (UK) Ltd., Cambridge, England

Bioserv Ltd., Worthing, Sussex, England

The Boehringer Corporation (London) Ltd., Lewes,E. Sussex, England

James Burrough Ltd., Fine Alcohol Division, London, England

Calbiochem-Behring Corp. (UK), Bishops Stortford, Herts., England

Cronex-Lighting, Du-pont (UK), Huntingdon, Cambs., England

Collaborative Research Inc., Universal Scientific Ltd. (UK distr), London, England

Difco Laboratories, West Molesey, Surrey, England

Fisons Scientific Apparatus, Loughborough, Leics., England

Fluka AG., Fluorochem Ltd. (UK distr.), Glossop, Derbyshire, England

Koch-Light Laboratories Ltd. Colnbrook, Bucks., England

Kodak Ltd., Kirby, Liverpool, England

LKB Instruments Ltd., LKB House, South Croydon, Surrey, England

New England Biolabs., CP Labs. Ltd. (UK dist.), Bishops Stortford, Herts., England

PL Biochemicals Inc., Northampton, England

Pharmacia Ltd., Milton Keynes, England

Schleicher and Schuell, Andermann and Co. (UK distr.), E. Molesey, Surrey, England

Sigma London Chemical Co. Ltd., Poole, Dorset, England

Whatman Lab Sales Ltd., Maidstone, Kent, England

Worthington, Flow Labs. Ltd., Irvine, Scotland.

UV Products, Winchester, Hants., England.

3. Media and Antibiotics

3.1 L-Broth

1% Tryptone (Difco),0.5% Yeast extract (Difco),0.5% sodium chloride, 10mM Tris.Cl pH 7.4, 1mM $MgSO_4$

3.2 M9-Medium

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

3.3 Nutrient Agar

15g of "Bacto"-agar (Difco) per litre of L-broth.

Antibiotics (Sigma) were used at the following concentrations : ampicillin 100µg/ml, tetracycline 15µg/ml, kanamycin 50µg/ml.

3.4 YT-Medium

0.8% Tryptone (Difco), 0.5% yeast extract (Difco),0.5% NaCl.

2 x YT is double the above concentrations.

3.5 YT-Agar

15g of "Bacto"-agar (Difco),per litre of YT-medium

3.6 Soft (top) Agar

6g of "Bacto"-agar (Difco),per litre of YT-medium

Part B : Methods

1. General Procedures

During the course of this work, a number of procedures were frequently used. The following section describes these general procedures.

1.1 Description of bacterial strains

Two strains of bacteria have been used during this project :

E. coli HB101 has been used as the host for the growth of all plasmid DNA and has the following genotype :

K12, f^-, pro^-, leu^-, thi^-, lac Y^-, hsd R^-, end A^-, rec A^-, rsp L20, ara-14, gal-K2, xyl-5, mtl-1, Sup E 44 (Bolivar and Backman, 1979).

The second strain, E. coli JM103 was the host used for the growth of the single-stranded bacteriophage M13 (Messing et al, 1981) and M13 recombinants. JM103 has the following genotype :

Δlac pro, sup E, thi, f' pro AB, lac I^q, ZΔM15, f' tra D36, str A, end A, sbc B15, hsd R4.

1.2. Storage of bacteria

Stock cultures of HB101 and HB101 transformed with plasmid DNA and JM103 can be stored under glycerol at -20°C. Add 2.5mls of sterile 80% glycerol to 2.5ml of exponentially growing culture and vortex to

ensure a homogeneous solution is obtained. The bacteria will remain viable for several years if stored at -20°C.

1.3 Storage of plasmid DNA

A list of the plasmids used in this study is given in Table 3.1. A full description of these plasmids is given in Chapter 2 (Section 2.2). Plasmid DNA is stored naked in TE buffer (10mM Tris. Cl pH 8.0, 0.1mM EDTA) over chloroform at 4°C in a tight fitting screw cap tube. DNA stored in this way remains stable for several years providing the chloroform level is maintained. In addition a small aliquot of plasmid DNA in TE buffer is stored at -20°C.

1.4 Restriction Enzymes

Restriction enzymes were purchased from the following companies ; Bethesda Research Laboratories (B.R.L.), New England Biolabs and The Boehringer Corporation (London) Ltd. Enzyme digests were set up according to the manufacturers specifications using the appropriate restriction buffer. The extent of digestion was monitored by electrophoresis of a small aliquot in a 1% agarose mini-gel system (see below).

1.5 Gel electrophoresis

Both agarose and polyacrylamide gel electrophoresis were used for the separation of DNA fragments. In general, agarose gels were used to separate fragments greater than 1kb in size, whereas, fragments of less than 1kb were separated on polyacrylamide gels.

CLONE	VECTOR	ANTIBIOTIC MARKER	REFERENCE
pXlrIOI	pMB9	Colicin EI immunity	R.Reeder (unpublished results).
pXlrIO2	"	"	
pXlrIO3	"	"	
pXlrIOIL	pBR322	Amp. resistance	Subcloned by
pXlrIOIM	"	"	
pXlrIO2L	"	"	L.M.C.Hall,
pXlrIO2M	"	"	
pXlrIO3L	"	"	(I98I).
pXlrIO3M	"	"	
pXlrII	Col EI	Colicin EI immunity	Dawid&Wellauer I976.
pXlrI4	"	"	Botchan et al, I977.
pXl 2I2	pCRI	Kan. resistance	Boseley et al, I978.
pXl IO8	"	"	Boseley et al, 1979.
pXlrIIL	pBR322	Amp. resistance	Subcloned by
pXlrI4A	"	"	B.E.H.Maden.
pXl IO8A	"	"	"
pXl 2I2L	"	"	M.A.Stewart (this wor
pXlcr I	pATI53	Amp. resistance	
pXlcr 2	"	"	
pXlcr 3	"	"	M.A.Stewart,
pXlcr 4	"	"	
pXlcr 5	"	"	(this work).

Table 3.I

Summary of the plasmid-based ribosomal clones used in this study. A full explaination of the rDNA fragment in each clone is given in figure 2.I.

"Xlr" stands for "X.laevis ribosomal"

"Xlcr" stands for "X.laevis chromosomal ribosomal".

1.5.1 Agarose gel electrophoresis

Electrophoresis buffer : 1x agarose running buffer contains 40mM Tris. Cl pH 8.0, 20mM sodium acetate, 2mM EDTA

Loading Buffer : 2M sucrose in 1x agarose running buffer and 0.05% bromophenol blue as marker dye

1% agarose solution : 5g of agarose (Sigma or B.R.L.) in 500ml of 1x agarose running buffer. Store molten at 65°C ready for use. Make up fresh every week

(a) 1% agarose minigels

The development of minigel systems has been invaluable for checking that a preparative restriction digest has gone to completion or for quick qualitative analysis of small quantities of DNA. I have used a BRL minigel system, model H6, gel size 75 x 50 x 2.5mm. The 1% agarose gel is just submerged in 1x agarose running buffer and electrophoresis is carried out at a constant current of 100mA until the bromophenol blue dye has travelled 2/3 way towards the anode. When electrophoresis is complete, remove the gel from the apparatus and stain in a solution of ethidium bromide (10µg/ml) for 15 mins. Visualise the DNA under U.V. light using a U.V. illuminator (Model C-62, U.V. Products) and photograph, if required. (Polaroid Cu-5, hand camera fitted with a yellow filter and polaroid 655 Positive/Negative film). To retain the negative immerse in 1M sodium sulphite for 5 mins then wash in cold water and dry.

(b) Preparative agarose gel electrophoresis

Either vertical 160 x 160 x 3mm agarose gels (B.R.L. Mode V16) or horizontal 250 x 200 x 5mm agarose slab gels (B.R.L. Model H0) were used. For less than 30µg of DNA use a vertical gel. Vertical gels are

normally run at a constant current of 100mA, horizontal gels at 120mA. Electrophoresis, staining and photography of the gel is carried out as described above.

1.5.2 Elution of DNA from agarose gels

There are many methods for recovering DNA from agarose gels. I have obtained highest recovery by electroelution in dialysis bags.

Run the gel and localise the band of interest using ethidium bromide staining and U.V. illumination. Using a sharp scalpel cut out a slice of agarose containing the band. At this point photograph the gel so that a record is kept of which band is eluted. Fill a dialysis bag to overflowing with 1x agarose running buffer and, holding the neck of the bag, pick up the gel slice with forceps and place it inside the fluid-filled dialysis bag. Allow the gel slice to sink to the bottom of the bag then remove most of the buffer leaving just enough to keep the gel slice in constant contact with the electrophoresis buffer. Seal the bag tightly with suitable clips ensuring that no air bubbles are trapped. Immerse the bag in a shallow layer of 1x agarose running buffer (a horizontal gel electrophoresis apparatus is ideal for this) and switch on the power supply. After electrophoresing at 100V for 2-3hrs (or 150V, 1-2hrs) reverse the polarity of the current for 2 mins to release the DNA from the wall of the dialysis bag. Open up the bag and carefully recover all the buffer surrounding the gel slice into a siliconised 15ml corex tube. Using a clean Pasteur pipette wash out the bag

with a small quantity of electrophoresis buffer. Stain the gel slice with ethidium bromide (10µg/ml) then examine under U.V. light to check that all the DNA is eluted. If elution is complete then adjust the DNA solution to 0.3M salt using 3M sodium acetate pH 5.5 and add 2½ volumes of ice cold ethanol. Vortex thoroughly and precipitate the DNA at -20°C overnight. Collect the DNA by centrifugation at 2500g for 45 mins. Dry the pellet in a vacuum dessicator for 5 mins, dissolve in 400µl of 0.3M sodium acetate and transfer to a 1.5ml snap cap polypropylene Eppendorf tube. High speed centrifugation for 5 mins (Eppendorf centrifuge model 5412, 9950g) pellets any contaminating agarose debris and leaves the DNA free in solution. Carefully remove the supernatant into a clean 1.5ml Eppendorf tube and add 1ml of ice cold ethanol. Precipitate the DNA at -70°C in a dry ice/ methylated spirits bath for 15 mins. Centrifuge (9950g) for 5 mins and remove the supernatant to waste taking care not to disturb the DNA precipitate. Wash the pellet in 300µl of 80% ethanol and chill for 5 mins at -70°C. If the DNA is to be used immediately centrifuge as before, remove the supernatant and dry the DNA in a dessicator for 5 mins. Dissolve in the appropriate volume of TE buffer pH 8.0 and store on ice ready for use. Otherwise store the DNA as an ethanol precipitate at -20°C and carry out the final centrifugation and subsequent drying step prior to use.

Addition of salt to a final concentration of 0.3M and ethanol (2½x aqueous volume) is a standard method for precipitating DNA. In subsequent sections this series of steps is referred to as the DNA "work-up".

1.5.3 Polyacrylamide gel electrophoresis (PAGE)

Acrylamide electrophoresis buffer : 1x working solution (1 x TBE)contains 0.1M Tris.borate, 1mM EDTA pH 8.3.

Loading buffer : 50% glycerol in 1 x TBE, 0.05% xylene cyanol and 0.05% bromophenol blue as marker dyes.

Preparative PAGE : Vertical 160 x 160 x 1.5mm 4% polyacrylamide gels were used to separate DNA fragments of less than 1kb. Electrophoresis in 1 x TBE is carried out at 200V for 2-3hrs until the dyes have travelled the required distance. In 4% gels xylene cyanol migrates with fragments 380-400bps, bromophenol blue migrates with fragments of 70-100bp. Visualisation of the DNA is as previously described (1.5.1).

1.5.4 Elution of DNA from polyacrylamide gels

The method used was based on the established method of Maxam and Gilbert (1980) in which the DNA of interest is eluted from the gel piece and collected through siliconised glass wool to retain the acrylamide debris. The DNA is subsequently purified by the "work-up" procedure.

2. Cloning Procedures

The construction of recombinant molecules during this work initially involved subcloning of restriction fragments from parent X. laevis rDNA clones and laterally cloning material from uncloned rDNA. The following section describes the procedures involved in the construction and screening of clones made de novo. The methods described are also suitable for the construction of subclones although a more simple approach for the identification of subclones is adopted (see

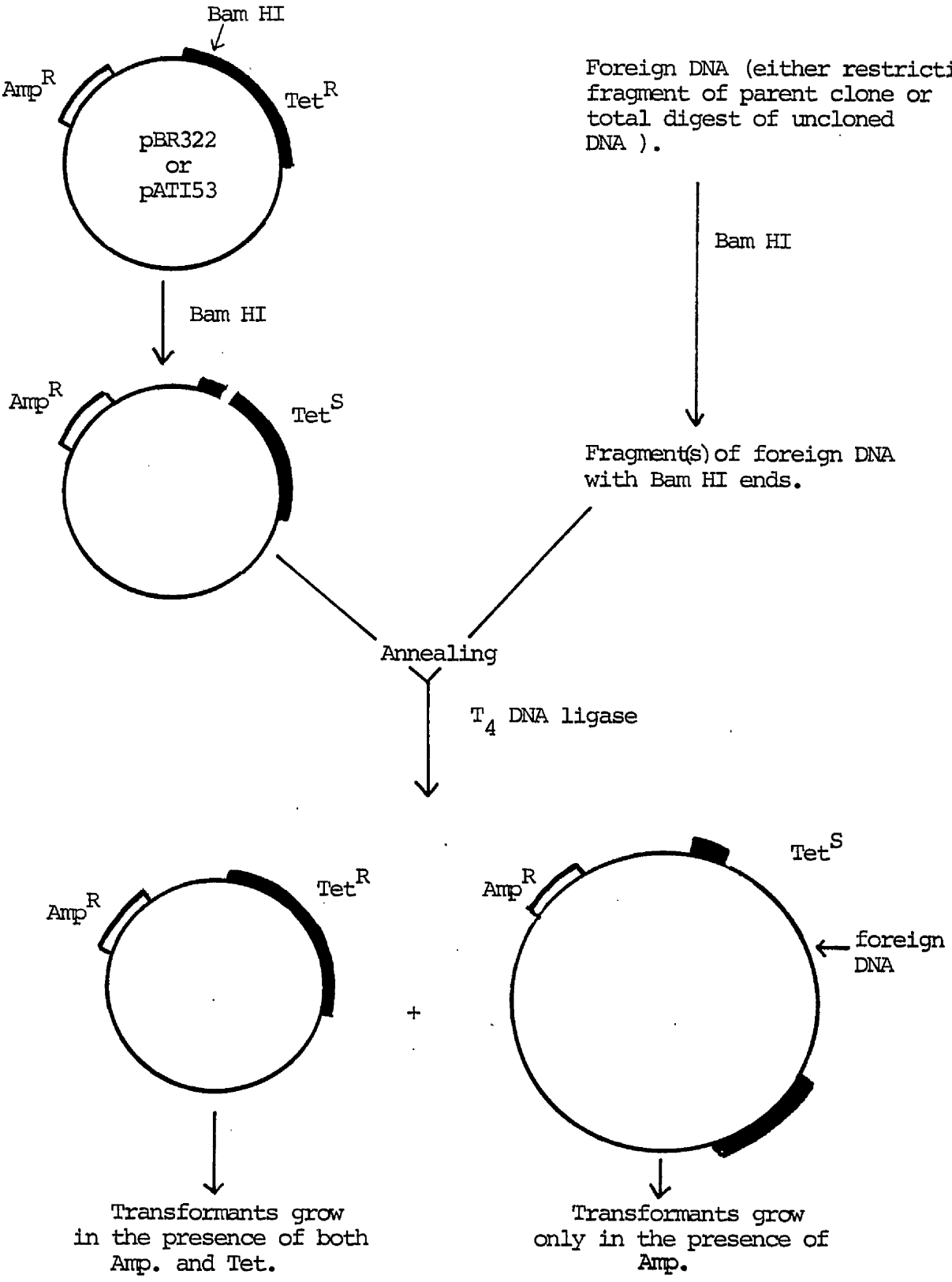


Figure 3.1. General strategy for generation of recombinant molecules. All of the cloning into plasmid vectors performed in this study involved insertion of foreign DNA into the tetracycline gene.

below).

Two related plasmids pBR322 (Sutcliffe, 1978) and pAT153 (Twigg and Sherratt, 1980) have been used in the construction of all the plasmid based recombinants made during this project. The general strategy for producing these hybrid molecules is illustrated in Figure 3.1. In all of the cloning experiments performed, the DNA to be cloned was inserted in the tetracycline gene of pBR322 or pAT153 to produce ampicillin resistant, tetracycline sensitive (Amp^R Tet^S) transformants.

2.1 Ligation reaction

The aim of the ligation reaction is the *in vitro* construction of hybrid DNA molecules in which a foreign piece of DNA is inserted into vector DNA. During this process two competing reactions are taking place (1) the end to end joining of separate molecules (intermolecular collisions) and (2) the joining of opposite ends of the same molecule (intramolecular collisions). The ligation reaction is performed to favour the formation of hybrid molecules. The initial stages of the reaction employ a small volume - high DNA concentrations will favour intermolecular collisions. Increasing the volume will then dilute the DNA and the second competing reaction is favoured (for a theoretical discussion see Dugaiczyk *et al*, 1975). I have found that, if possible, a 5-10 fold molar excess of "insert" DNA over vector DNA favours the formation of recombinant molecules.

If a specific restriction fragment is to be cloned then purify the DNA by gel electrophoresis. If cloning involves a mixture of restriction fragments then purify the mixture of DNAs by phenol extraction to remove any restriction enzymes present. Purify the linear form of the required vector by gel electrophoresis (it is important to eliminate circular forms of the vector as these will transform with a high efficiency and give rise to a high background of colonies). In addition if the vector DNA has been linearised with a single restriction enzyme only, then recircularisation of the linear DNA can be minimised by treating the DNA with bacterial alkaline phosphatase, BAP (Worthington) to remove free 5'-phosphates. The self-ligation reaction is then unable to occur since one of the substrates required by T4 DNA ligase is a 5'-phosphate group.

Set up the ligation reaction in a 1.5ml Eppendorf tube containing 10:1 molar ratio of insert:vector DNA and 50mM Tris, pH 7.6, 8mM $MgCl_2$, 0.1mM ATP and 10mM dithiothreitol (reaction buffer) in a final volume of 10µl. Add 0.5µl of T4 DNA ligase (Bethesda Research Laboratories 1U/µl), vortex, and incubate for 2 hrs at 15°C. Dilute the DNA with 10 volumes of reaction buffer (as above) and add 1 unit of T4 ligase. Incubate overnight at 15°C.

2.2 Transformation of *E. coli* HB101

2.2.1 Preparation of competent cells

Chill a 100ml culture of exponentially growing HB101 on ice for 2 mins then transfer 4 x 25ml aliquots into sterile 30ml corex tubes. Incubate at 0°C for 10 mins then centrifuge the cells at 4000g for 10 mins at 4°C. Discard the supernatant and resuspend the pellets in a total of 50ml of ice cold sterile 50mM $CaCl_2$. Resuspend the pellets gently by tapping the side of the corex tubes and incubate on ice for 20 mins. Centrifuge as before and resuspend the pellets in a total of 4mls of ice cold 50mM $CaCl_2$. Pool the cell suspensions and add 2ml of sterile 80% glycerol. Aliquot 250µl samples into sterile snap cap tubes and freeze at -70°C. These cells can be stored for many months in this way and still retain their competency.

2.2.2 Transformation of HB101 by plasmid DNA

For each transformation, remove an aliquot of competent cells from -70°C and thaw on ice. Add the ligation mix(es) to one (or more) aliquots and incubate on ice for 30 mins. As a control to check the efficiency of transformation add 1ng of uncut pBR322 to an aliquot of competent cells and treat in the same way as the test transformations. Remove the tubes from 0°C, incubate at 37°C for 5 mins then stand at room temperature for 10 mins. Add 500µl of sterile L-broth to each mix, place the tubes inside a large container, such as a 250ml conical flask and incubate (without shaking) for 1hr at 37°C. This allows expression of the antibiotic resistance

markers of the transforming DNA (Cohen et al, 1972). Spread a range of concentrations (1-1/100) of the transformation mixes directly onto detergent-free nitrocellulose filters (Schleicher and Schuell) which are lying on the surface of ampicillin nutrient agar plates. Let the plates stand at room temperature until the liquid has been absorbed then invert and incubate the plates overnight at 37°C. Small colonies (0.1mm in diameter) will appear in 8-10hrs.

2.2.3 Selection of recombinants on the basis of antibiotic resistance

For the test transformations many colonies will be present on the nitrocellulose filters. (If transformation efficiency is high then 10^6 colonies/µg of uncut pBR322 should be obtained). In order to distinguish between recombinant molecules and recircularised plasmid, colonies must be replica plated onto both ampicillin and tetracycline agar plates. This is achieved by pressing a fresh nitrocellulose filter onto the original filter (filter-to-filter contact). A mirror image of the original filter is obtained. This process can be repeated up to five times using the original filter. The replica filters are then placed onto ampicillin and tetracycline nutrient agar plates and incubated overnight at 37°C.

If cloning into the tetracycline gene involved a purified restriction fragment (as in subcloning) then over 90% of the colonies which are Amp^R Tet^S will contain the recombinant DNA of interest. This can be confirmed by limited restriction analysis of a small scale preparation of

the recombinant DNA from a few selected colonies (see below). If however, the cloning has involved a mixture of fragments (as in de novo cloning) then only some of the Amp^R Tet^S colonies will contain the recombinant DNA of interest. Thus, such putative recombinants must be analysed further. A duplicate filter of the nitrocellulose filter containing the Amp^R Tet^S colonies is made by filter-to-filter contact (see above). This duplicate is then used for subsequent analysis of the putative recombinants whereas the original filter serves as a master filter. The master filter should be kept sealed at $4^{\circ}C$ on an ampicillin agar plate.

2.3 Identification of recombinant clones by colony hybridisation

The method used was based on the method of Grunstein and Hogness (1975) in which colonies on nitrocellulose filters are lysed and the liberated DNA fixed to the filter by baking. After hybridisation of a ^{32}P-labelled probe, the filter is monitored by autoradiography. A colony whose DNA gives a positive autoradiographic result may then be easily identified on the master filter, recovered and subsequently analysed.

2.3.1 Solutions

1x Working solution (1 x SSC) contains 0.15M NaCl, 0.015M sodium citrate (Koch Light)

keep 2ℓ of 20 x SSC as a stock solution

Lysis buffer : 0.75M NaOH, 1.5M NaCl

Neutralising buffer : 1M Tris Cl pH 7.2, 1.5M NaCl.

2.3.2 Preparation of DNA from individual colonies

All incubations are at room temperature unless otherwise stated. Place the filters onto 5 layers of Whatman 3MM paper presoaked in lysis buffer and incubate for 10 mins. Transfer each filter onto 0.8ml of neutralising buffer in a petri dish and incubate for 3 mins. Blot excess liquid onto a tissue and repeat the last step using fresh neutralising buffer and a clean petri dish. Place the filters onto a clean tissue, colonies upwards until they assume a dryish appearance.

Proteinase K step : This is optional but I have achieved cleaner hybridisation results when included. Make up 0.7mg/ml Proteinase K (Sigma) in 1 x SSC. Place the filters into separate petri dishes and add 2ml of the solution to each, dropwise onto the colonies. Incubate for 90 mins ensuring that the filters are always covered by the solution during the incubation. Rinse the filters twice with 500mls of 2 x SCC (the washing buffer becomes cloudy due to degraded bacterial debris from the colonies) then lay the filters onto a few layers of tissue to absorb excess liquid. Blot once with a piece of 3MM paper. The liberated DNA is then baked

onto the filter by incubating in a vacuum oven for 2hrs at 80°C.

2.3.3 Preparation of ^{32}P-labelled probe

A high specific activity probe was generated using the method of Maizels (1976) in which HeLa cell 28S rRNA is subjected to mild alkaline hydrolysis to yield fragments of 60-90bp in length, followed by kinase labelling of the available 5' ends. HeLa 28S rRNA was used because large amounts were available in our laboratory and the RNA was shown to hybridise strongly to Xenopus laevis 28S rDNA.

Adjust 5µg of HeLa 28S rRNA to 50mM Tris pH 9.5 and boil for 10 mins at 90°C in a 0.5ml siliconised Eppendorf tube. Centrifuge for 15 secs (9950g) and place tube on ice. Transfer 1µg of the hydrolysed material (usually 2µl) to a 1.5ml siliconised Eppendorf tube and label using T_4 polynucleotide kinase (PL Biochemicals) and high specific activity γ-^{32}P-ATP (> 5000 Ci/mmole in aqueous solution). The method for labelling is the same as used for labelling DNA fragments for Maxam and Gilbert sequencing and is described in Section 3.1.2.

Separation of ^{32}P-28S rRNA from mononucleotides and unincorporated label is achieved by column chromatography using Sephadex G-25 (Pharmacia Ltd.) packed into a Pasteur pipette. Collect approximately 35, 2 drop fractions into 1.5ml Eppendorf tubes and using a hand held mini-monitor determine empirically which fractions contain labelled RNA. Two peaks of radioactivity are obtained. The first peak

corresponds to labelled RNA and is eluted in the void volume of the column. The second peak is due to unincorporated labelled phosphate. Fractions containing labelled RNA should be pooled and the total volume noted.

2.3.4 Determination of specific activity of the probe

In a 5ml glass tube add 1ml of 0.2% SDS, 1ml of 25% trichloroacetic acid (TCA), 50µl 0.2% bovine serum albumin and 1µl of the labelled RNA sample. Vortex and incubate on ice for 30 mins. Collect the precipitate by filtering the solution through a prewashed nitrocellulose disc (0.45µm). Wash the filter three times with 5ml of ice cold 5% TCA. Dry the filter under a heat lamp, place inside a glass scintillation vial, add 5ml of liquid scintillant (5g/ℓ PPO (2,5-diphenyloxazole) in toluene) and monitor counts using a scintillation counter. This gives a measure of the radioactivity incorporated into 28S rRNA. Specific activities of 1-5 x 10^7 cpm/µg RNA were obtained. Store the probe, frozen, ready for use.

2.3.5 The hybridisation reaction

The pre-hybridisation and hybridisation reactions are performed in a polythene bag slightly larger than the filters. The bag is sealed with the expulsion of air.

Pre-hybridise the nitrocellulose filters in 10mls of pre-hybridisation mix (4 x SSC, 50% formamide (Fluka), 10µg/ml sonicated calf thymus DNA) for 4-6hrs at 42°C. Remove the pre-hybridisation mix and add 10mls of hybridisation mix which contains 4 x SSC, 50% formamide and the ^{32}P-

labelled 28S rRNA probe. Reseal the hybridisation bag and incubate at 42°C overnight. Remove the hybridisation mix (it can be re-used if stored at -20°C) and transfer the filters to a 2ℓ beaker. Wash the filters extensively first, in 1ℓ of 5 x SSC with fairly vigorous agitation followed by incubation in 200mls 4 x SSC, 50% formamide for 8 hrs at 42°C. Replace the wash solution and incubate overnight at 42°C.

2.3.6 Autoradiography

Blot the filters using 3MM paper, air-dry at room temperature for 10 mins then bake in vacuum oven at 80°C for 30 mins. Tape the filters by their extreme edges onto a clean piece of 3MM paper attached to a glass plate and using radioactive ink arrow several locations on the 3MM paper around the filters. These markers serve to align the autoradiograph with the filters. Then cover with a layer of Saran wrap.

In the darkroom cover the filters with a sheet of Kodak-X-Omat H-film using an intensifying screen (Cronex-lighting, Dupont, U.K.) and wrap up to prevent any light penetration. Expose overnight at -70°C. Develop the autoradiograph and align the film with the filters using the marks left by the radioactive ink. Locate any positive hybridisation signals and identify the positive clones by referring back to the master filter stored at 4°C on an ampicillin agar plate. Small scale preparations of the plasmid DNA from the individual colonies which gave positive signals can now be made.

2.4 Small Scale Isolation of Recombinant DNA

A number of rapid procedures are available for isolating plasmid DNA grown from individual bacterial colonies. These methods yield DNA in sufficient quantity and purity for analysis by restriction enzyme digestion and gel electrophoresis. Therefore, colonies that have been identified by their ability to hybridise to a given probe can be rapidly characterised in some detail. Recombinant molecules containing the DNA of interest can then be prepared on a larger scale. The rapid method I have used (Birnboim and Doly, 1979) relies on the selective alkaline denaturation of high molecular weight chromosomal DNA while covalently closed circular (ccc) DNA remains double-stranded.

2.4.1 Solutions

I Lysozyme solution : 2mg/ml lysozyme (Sigma), 50mM glucose, 10mM EDTA, 25mM Tris. Cl pH 8.0.
Prepare a 10x stock excluding lysozyme which is added prior to use. (store $4^{\circ}C$).

II Alkaline SDS solution : 0.2M NaOH, 1% SDS (store room temperature).

III High salt solution : 3M sodium acetate pH 4.8 (store $4^{\circ}C$).

IV 0.1M sodium acetate, 0.5M Tris. Cl pH 8.0 (store $4^{\circ}C$).

2.4.2 Extraction of plasmid DNA

Prepare 2.5ml overnight cultures of selected colonies from the hybridisation experiment. Transfer 0.5ml of each culture to a 1.5ml Eppendorf tube and store the remaining culture at -20°C after addition of sterile glycerol to 40%. Centrifuge the 0.5ml culture (9950g) for 1 min and resuspend the cell pellet in 100µl solution I. Incubate for 30 mins at 0°C then add 200µl of solution II and gently vortex the tubes. After 5 mins at 0°C add 150µl of solution III and mix the contents by inversion (most of the protein, high mol. wt. RNA and chromosomal DNA precipitates). Centrifuge (9950g) for 5 mins and remove 400µl of supernatant from each tube into clean 1.5ml Eppendorf tubes. Add 1ml of ice cold ethanol and precipitate the DNA (and low mol. wt. RNA) by freezing at -70°C for 15 min. Centrifuge as before and dissolve the pellet in 100µl solution IV. Add 2 volumes of ice cold ethanol and after 15 mins at -70°C collect the precipitate by centrifugation. Wash the pellet twice in 80% ethanol prior to enzyme digestion.

Limited restriction analysis is then carried out using enzymes with known recognition sites for the recombinant of interest.

2.5 Large Scale Isolation of Plasmid DNA

2.5.1 Growth and harvesting of bacteria

The volumes given below are for a 500ml culture but can be adapted for the preparation of plasmid from larger

cultures or for smaller preparations of different plasmids.

Prepare an overnight culture of the transformed bacteria (Section 2.4.2) and inoculate 500ml of M9 medium in a 2ℓ flask with 5ml of the overnight culture. Incubate at 37°C with vigorous shaking until the culture reaches logarithmic growth stage (OD_{660} 0.4-0.5 units) and then add 2.5ml chloramphenicol (Sigma) solution (34mg/ml in ethanol). Continue the incubation for 12-16hrs during which time plasmid DNA amplification occurs while chromosomal DNA synthesis is inhibited. Harvest the cells by centrifugation at 5000g for 10 mins at 4°C and wash the cell pellets in a total of 100ml of ice cold STE (0.1M NaCl, 10mM Tris. Cl pH 7.8, 1mM EDTA). Centrifuge as before.

2.5.2 Lysis of cells by treatment with SDS

Resuspend the bacterial pellet from the 500ml culture in 10ml of ice cold 10% sucrose in 50mM Tris. Cl pH 8.0. Transfer the suspension to a 50ml screw cap glass tube and add 2ml of a freshly made lysozyme solution (10mg/ml in 0.25M Tris.Cl pH 8.0). Add 8ml of 0.25M EDTA and mix by inversion several times. Incubate on ice for 10 mins then add 4ml of 10% SDS. Mix quickly with a glass rod to disperse the SDS evenly through the bacterial suspension and gently so as not to sheer the liberated bacterial DNA. Immediately add 6mls of 5M NaCl, mix gently by inversion and incubate on ice for at least 1hr. Centrifuge to remove high molecular weight DNA and bacterial debris in a Beckman Type 50 Ti rotor at 30,000 rpm for 30 mins at 4°C. Pour off and

save the supernatant. Discard the pellet which should be firm and tight. Nucleic acid is then purified from residual protein by extraction twice with an equal volume of 1:1 phenol:chloroform and once with an equal volume of chloroform. After each extraction, transfer the aqueous layer to a clean tube. Deproteinisation is more efficient when two types of organic solvent are used and the final extraction with chloroform removes any lingering traces of phenol from the preparation. Recover the DNA by ethanol precipitation followed by two washes in 80% ethanol to remove residual salt. After the final centrifugation dry the pellet and dissolve in 15ml of TE buffer, pH 8.0.

2.5.3 Purification of closed circular DNA

Closed circular plasmid DNA is purified from linear plasmid DNA and any remaining chromosomal DNA by centrifugation to equilibrium in caesium chloride (CsCl) gradients containing saturating quantities of ethidium bromide. cccDNA binds much less ethidium bromide than linear DNA and therefore bands at a higher density in such gradients. Any RNA that is present will travel through the gradient and pellet.
Add 1g of solid CsCl to each ml of DNA solution and mix gently to dissolve all the salt. Add 0.8ml of ethidium bromide (10mg/ml) for every 10ml of CsCl/DNA solution. Mix well. Check that the final density of the solution is 1.55g/ml by weighing 1ml of the solution. Centrifuge to equilibrium at 20°C either in a Beckman Ti 60 rotor for 36 hrs at 45,000 rpm (if volumes are less than 30mls) or in a Beckman VTi 50 rotor for 16 hrs at 49,000 rpm (if volumes are larger than

30ml). Two bands of DNA are visible in ordinary light. The upper band corresponds to linear bacterial DNA and nicked plasmid DNA, the lower band consists of cccDNA. To enhance resolution for removal of the cccDNA visualise the bands under U.V. light. Collect the lower band of DNA into a glass tube.

2.5.4 Removal of Ethidium bromide

Extract the ethidium bromide/DNA solution with an equal volume of saturated isopropanol solution (isopropanol, 0.05M Tris. Cl pH 8.0, 0.05M NaCl, 0.01M EDTA, saturated with solid CsCl) until all the pink colour disappears from the aqueous phase.

Dialyse the aqueous phase against several changes of TE pH 8.0 at $4^{\circ}C$. When complete, measure the absorbance of 1ml of the DNA solution at 260nm (1OD unit is equivalent to 50µg/ml DNA). Store the plasmid DNA at $4^{\circ}C$ as described (Section 1.3).

I have obtained yields of 2mg of plasmid DNA from 500ml starter cultures using this method.

3. DNA Sequencing by the Maxam and Gilbert Chemical Method

Plasmid DNA was prepared for, and sequenced by, the method of Maxam and Gilbert. This procedure was well established in our laboratory and is described in great detail by Hall (1981). For a detailed analysis of the procedure consult Maxam and Gilbert (1980).

3.1 5'end labelling

3.1.1 Phosphatase reaction

Dissolve the DNA fragment in 90µl of TE pH 8.0 and add 10µl of bacterial alkaline phosphatase (Worthington, further purified according to Hall (1981)). Incubate at 37°C for 60-90 mins then extract with redistilled phenol. Reextract the phenol phase with an equal volume of TE pH 8.0, pool the aqueous phases and ethanol precipitate after the addition of salt to 0.3M. "Work-up" the DNA as described (1.5.2).

3.1.2 Polynucleotide kinase reaction

If the DNA fragment to be sequenced has 5' protruding ends then the reaction conditions are as follows. Dissolve the phosphatased DNA in 2µl of H_2O and add 1µl of 10x kinase buffer pH 8.0 (1 x kinase buffer is 0.05M Tris, 0.01M $MgCl_2$), 1µl 50mM dithiotheitol, 6µl aqueous γ-^{32}P-ATP (Amersham International plc, 5000 Ci/mmole) and 5 units of T_4 polynucleotide kinase (PL Biochemicals). Mix, then incubate at 37°C for 30 mins (If the DNA to be sequenced has flush or 3' ends then labelling is favoured if, prior to the addition of the γ-^{32}P-ATP and kinase, 1µl of 10mM spermidine is added to the DNA sample and the mixture heated to 75°C for 5 mins). Add 40µl of 2.5M ammonium acetate and 160µl of ice cold ethanol. Precipitate the DNA at -70°C for 15 mins, centrifuge for 5 mins and "work-up" the DNA as described (1.5.2).

3.1.3 Separation of labelled ends

The 5'-labelled ends of a piece of double-stranded DNA are separated by cleavage of the fragment into two or more subfractions with a restriction enzyme which is known to cut the DNA fragment, followed by PAGE. The products are visualised either by ethidium bromide staining or if only very small amounts are present then by autoradiography (expose for 20 mins only using Kodak nonscreen 2T film). Identify the bands in the gel and elute as described (1.5.4).

3.2 Base Specific Chemical Cleavage Reactions

3.2.1 Solutions

1. Pyridinium formate : 4% v/v formic acid adjusted to pH2.0 with pyridine (using 0.005M H_2SO_4 as pH 2.0 std.)
2. DMS buffer : 50mM sodium cacodylate, 10mM $MgCl_2$, 0.1mM EDTA pH 8.0.
3. "DMS stop": 1.5M sodium acetate, 1M β-mercaptoethanol (Koch-Light),100µg/ml yeast RNA
4. "Hydrazine stop" : 0.3M sodium acetate, 0.1mM EDTA, 50µg/ml yeast RNA

3.2.2 Additional reagents

1. Dimethylsulphate - DMS (Gold Label, Aldrich Chemical Co. Ltd.)
2. Hydrazine -Hz (Kodak Ltd.)
3. Piperidine (Koch-Light)
4. AnalaR ethanol (James Burrough Ltd.)

3.2.3 Base modification reactions and chain cleavage

The four reactions I used for full sequence determination were specific for guanine (G), guanine and adenine (G + A), cytosine (C) and cytosine and thymine (C + T). When only partial sequencing was performed then only the G or C specific reactions were utilised. Chain cleavage was achieved using 1M piperidine. Table 3.2 describes the precise protocol followed.

3.3 Gel electrophoresis

High resolution thin (0.4mm) sequencing gels according to Sanger and Coulson (1978) were used. Routinely these are 8% or 6% polyacrylamide gels containing 7M urea (Aristar grade, BDH) and electrophoresis is in 1 x TBE buffer. Pre-electrophorese the gel for 1-2hrs at 25-30mA (LKB 2103 power pack). This current initially draws a voltage of 1.2-1.3kV but as the gel heats up the voltage rises to 2kV then drops back to about 1.5-1.7kV and remains constant at this value throughout electrophoresis.

While the gel is pre-electrophoresing dissolve the samples in sequencing loading dye (99% deionised formamide, 0.05% xylene cyanol). For an overnight autoradiographic exposure 10,000 cpm per loading is ideal so, if possible, dissolve the DNA to give 10,000 cpm/µl. Boil the samples for 2 mins then chill quickly on ice. Wash out the wells of the gel prior to loading to remove residual urea which lies in the bottom of the wells. Normally three consecutive loadings are carried out per gel, 6% (8%), to allow up to 180 (140) nucleotides to be read from the 5' labelled end of the DNA fragment.

	G	G + A	C + T	C
I.	Iμl carrier DNA Iμl ^{32}P-DNA 98μl DMS buffer	Iμl carrier DNA Iμl ^{32}P-DNA IIμl H_2O	Iμl carrier DNA Iμl ^{32}P-DNA 6μl H_2O	Iμl carrier DNA Iμl ^{32}P-DNA 8μl sat. NaCl.
2.	0.5μl DMS	2.5μl pyr. form.	I5μl Hz.	I5μl Hz.
3.	20°C, 4mins.	30°C, 70mins.	20°C, 6mins.	20°C, 8mins.
4.	24μl DMS stop 400μl Ar. ETOH	freeze -70°C lyophilise. add IOμl H_2O freeze -70°C lyophilise.	60μl Hz. stop 250μl Ar. ETOH	
5.	-70°C, I5mins. c'fuge (9950g) 5mins.		-70°C, I5mins. c'fuge (9950g) 5mins.	
6.	60μl 0.3M Na.Ac. 200μl Ar. ETOH		60μl 0.3M Na.Ac. 200μl Ar. ETOH	
7.	repeat step 5		repeat step 5	
8.	200μl 70% ETOH		200μl 70% ETOH	
9.	repeat step 5		repeat step 5	
IO.	200μl 70% ETOH		200μl 70% ETOH	
II.	repeat step 5		repeat step 5	
I2.	dry pellet		dry pellet	
I3.	IOOμl IM Piperidine			
I4.	90°C, 30mins.			
I5.	c'fuge (9950g) 30secs.			
I6.	freeze -70°C, lyophilise, add 20μl H_2O, freeze -70°C, lyophilise, add 20μl H_2O, freeze -70°C, lyophilise.			

Table 3.2. Protocol used for DNA sequencing by the Maxam and Gilbert chemical method.

DMS-dimethylsulphate; Hz-hydrazine; sat. NaCl-saturated NaCl; pyr. form.-pyridine formate,pH2; Ar. ETOH-analar ethanol; Na.Ac.-sodium acetate.

All the reactions were performed in 1.5ml siliconised Eppendorf tubes. DMS and Hz wastes were disposed of into 5M NaCl and 3M ferric chloride respectively and steps involving handling of these chemicals were carried out in the fume hood. After the final lyophilisation in step 16, the radioactivity of each sample was determined by Cerenkov counting.

3.4 Autoradiography

Remove one of the glass plates to expose the gel and cover with cling film. Identify the lanes with radioactive ink then wrap up using an intensifying screen and store as previously described (2.3.6). Expose for the desired length of time depending on the number of counts loaded.

4. Isolation and Purification of rDNA from X.laevis

4.1 Maintenance of animals

Three mature female X. laevis frogs were purchased from Bioserv Ltd. The frogs (3-4 yrs old) ranged from 9-10cm in length (Figure 3.2). They were kept in a large aquarium containing copper free water kept at a constant temperature of 25°C. Feeding was once per week, each animal receiving 30-50g of sheep heart. The tank was cleaned and water changed one day after feeding.

4.2 Dissection of animals

4.2.1 Anaesthesia

Place a single frog in a large beaker containing a 0.4% solution of MS222 (Sigma : ethyl-m-aminobenzoate methane sulphonic acid salt). This chemical is a powerful anaesthetic and is absorbed through the foot and hand pads of the frog. Within 20 mins. the frog is fully anaesthetised and will show no sign of movement when held ventral side upwards.

4.2.2 Cardiac puncture

Place the frog, dorsal side down onto a couple of pieces of 3MM paper. Wash the ventral side with ethanol. Place pads of cotton wool soaked in MS222 onto the foot and hand pads to ensure that the level of anaesthesia is maintained during the dissection. Using sterile forceps and sharp

scissors make a small slit 1cm below the xiphisternum (Figure 3.2). Insert one blade of the scissors into the slit and cut the skin forwards to the level of the lower jaw. Cut transversely at the level of the arm as far as the elbows. Loosen the skin from the underlying muscle so that it can be turned back freely. With clean scissors make a small incision in the abdominal wall on the right side of the mid-ventral line about 1cm below the xiphisternum, and cut the muscle forwards to the level of the lower jaw. Be careful not to sever the anterior abdominal vein which runs along the mid ventral line from the abdomen to the heart. In addition try not to puncture any internal organs with the tip of the scissors.

Lift the xiphisternum using forceps and cut across the centre of it - this exposes the heart which lies directly beneath the xiphisternum in a small cavity. Clean the scissors and forceps with ethanol. Using the clean forceps pull the heart out of the cavity (it is quite robust and will stand up to quite rough handling) and inject 0.5ml of heparin 100 U/ml (Sigma) into the heart using a hypodermic needle (25mm gauge). Allow the heparin to pump through the circulatory system for approximately 1min then gently start to pull back the plunger of the syringe drawing blood into the heparinised hypodermic. If the needle slips out or no more blood is being drawn out then transfer the blood into a sterile 50ml Falcon centrifuge tube (supplied by A and J. Beveridge Ltd.) on ice. To ensure that the blood does not clot add 0.5ml of the heparin solution to the blood and mix briefly. Puncture the heart again with the hypodermic and repeat the extraction procedure. Once no more blood can be extracted in this way nick the apex of the heart using sharp scissors and collect the blood that is pumped out using a Pasteur pipette. Pool all the blood into the 50ml Falcon tube.

Figure 3.2 Schematic representation of the ventral side of a frog.

The figure illustrates how access is gained to the body cavity by an incision along the mid-ventral line. The heart lies directly beneath the xiphisternum in a small cavity.

Figure 3.3 Schematic representation of the position of some of the internal organs in X. laevis

The position of the heart relative to the ovaries is illustrated. In a mature X. laevis the ovaries cover the entire body cavity. Note the positions of the gall bladder and stomach - care must be taken not to damage these organs during the dissection.

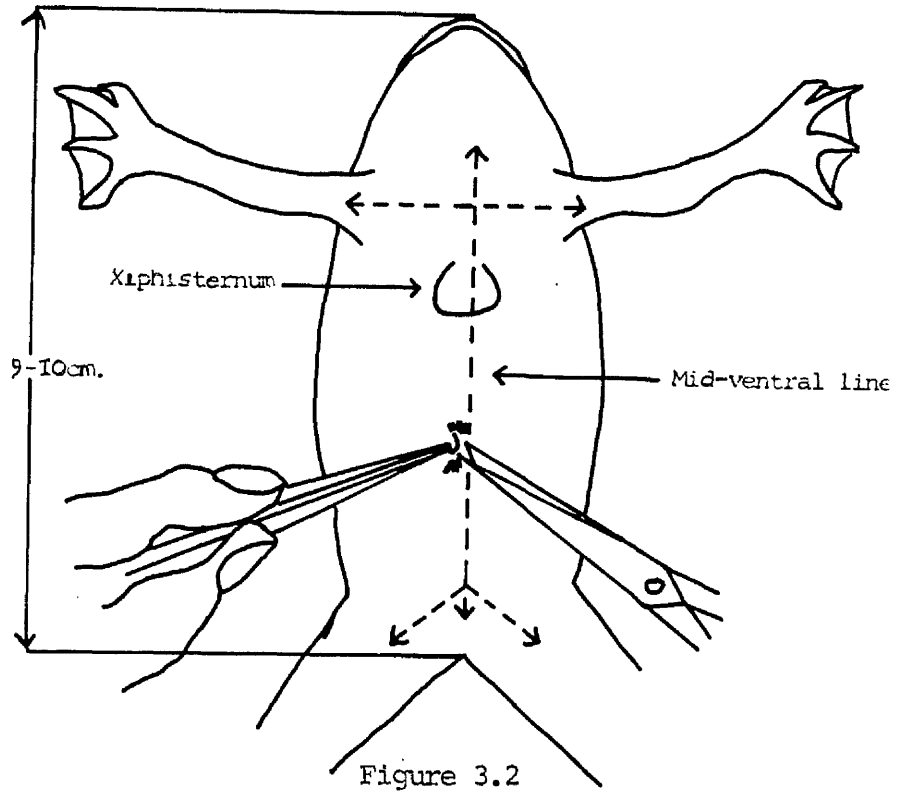


Figure 3.2

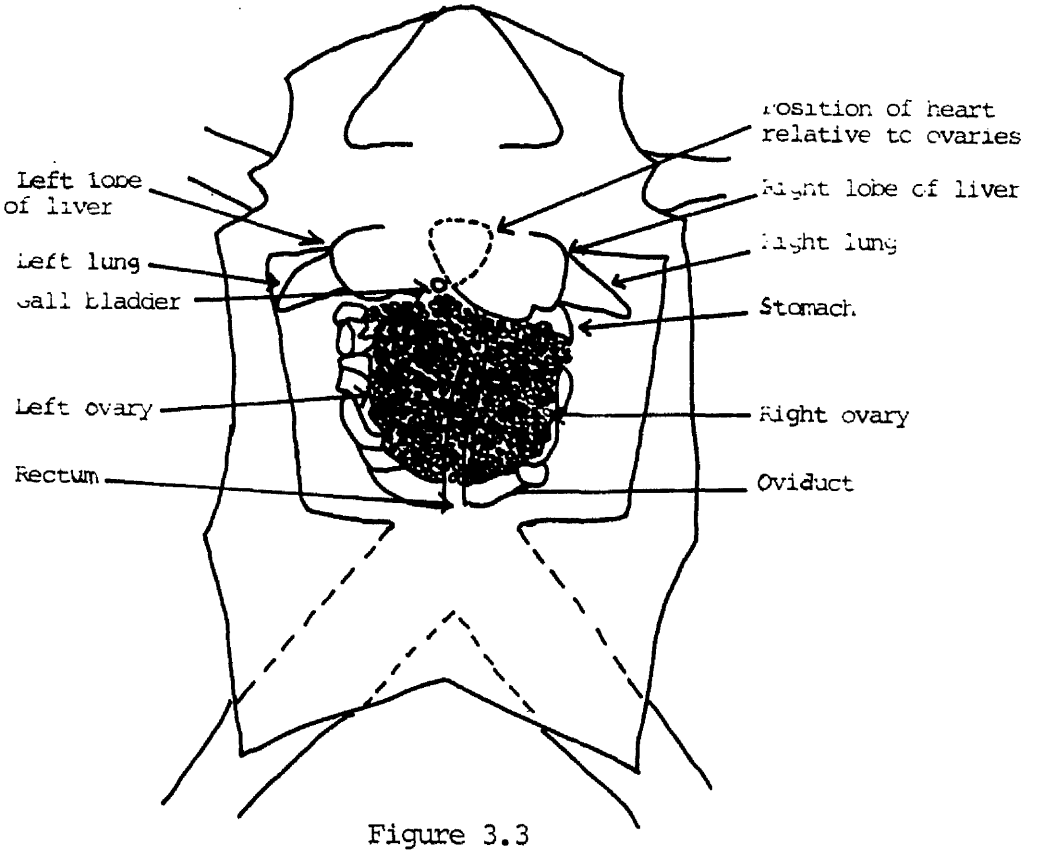


Figure 3.3

4.2.3 Removal of ovaries

Extend the skin cut down towards the level of the pelvic girdle and down each thigh to the level of the knee (Figure 3.2). Loosen the skin away from the underlying muscle and turn back. Repeat this for the muscle layer taking care not to damage any underlying organs especially the gall bladder and stomach (both rich in strong digestive enzymes).

In a mature X. laevis, the ovaries cover the entire abdominal cavity (Figure 3.3) and are very easily removed from the joining connective tissue. With sterile forceps and scissors pick up one of the ovaries and carefully snip away the thin layers of connective tissue. To remove the ovary completely cut through the join of the ovary with the oviduct and place it in a petri dish on ice containing 3-4mls of 1 x SSC. Repeat the process for the second ovary.

4.3 Isolation and Purification of Total DNA from Erythrocytes

The protocol devised, is based on the methods of Marmur (1961) Dawid et al (1970) and Bird and Southern (1978). It was also used for extraction of DNA from X. laevis oocytes.

Make the volume of the red cell/heparin mix up to 15ml with 1 x SSC. Lyse the cells by adjusting the solution to 1% SDS and add 3mls of 0.5M EDTA pH 8.0 (to inhibit any digestive action of DNases). Incubate at 60°C for 15 mins then phenol extract, gently, overnight at 4°C. Add 10ml of 1 x SSC to increase the volume of the aqueous phase. The solution is quite lumpy due to the mass of denatured protein. Subsequent phenol extractions compact this proteinaceous material. Centrifuge (1500g, 15 mins, 15°C) and remove the lower phenol phase. Add 20mls of phenol/ chloroform solution (1:1 v/v) and mix the contents of the tube by inversion for 10 mins. Repeat centrifugation and remove the organic phase. Reextract

the aqueous phase and protein mass with an equal volume of chloroform by inversion for 5 mins. Centrifuge as before. By this stage the aqueous phase is becoming clear but is very sticky due to the high molecular weight chromosomal DNA. Remove the aqueous phase to a clean 50ml Falcon tube leaving behind the now tightly packed protein layer and the organic phase. Extract the aqueous phase once more with chloroform and remove the aqueous phase into two sterile 30ml corex tubes on ice. A wide bore glass pipette is used to avoid shearing the high molecular weight material. Add two volumes of ice cold ethanol to each, invert the tubes and a stringy mass of nucleic acid precipitates immediately. Transfer the nucleic acid to a sterile 30ml corex tube on ice using a glass rod. Dissolve in 10ml of 1 x SSC by gently shaking the solution until dispersion is complete (lumps of nucleic acid can be recognised by adhering air bubbles when the solution is shaken).

Add DNase free pancreatic ribonuclease (Sigma) and RNase T1 (Calbiochem) to final concentrations of 50µg/ml and 50U/ml respectively. Incubate at 37°C for 30 mins then add 500µl of a solution of predigested Pronase (100mg/ml in 10% SDS incubated for 3 hrs at 37°C) and incubate for 1hr at 37°C. Remove proteins by a phenol/chloroform extraction followed by chloroform extraction. Precipitate the nucleic acid (high molecular weight DNA and small fragments of RNA) by adding 2 volumes of ice cold ethanol and chilling to -20°C for 30 mins. Centrifuge (2500g, 15 mins 0°C) and wash the pellet twice in 80% ethanol. Dry briefly under vacuum and dissolve the nucleic acid in a total volume of 10ml TE buffer pH 8.0. Store at 4°C prior to ultracentrifugation in CsCl gradients.

4.4 Isolation and purification of total DNA from ovaries

Pre-cool a sterile glass homogeniser tube (40ml volume) and a loosely fitting teflon pestle. Homogenise the two freshly dissected ovaries in 10ml of 1 x SSC for four strokes of the pestle at full speed. Keep the homogeniser tube cool by swirling on ice between strokes. Transfer the homogenate into a sterile 50ml Falcon tube on ice, and wash the homogeniser tube and pestle with 5ml of 1 x SSC. Pool this rinse with the homogenate. Lyse the cells as described (4.3) and phenol extract overnight. After centrifugation (1500g, 15 mins, 15°C) three distinct layers are clearly visible - a lower organic phase, a tightly packed protein layer and a cloudy upper aqueous phase. Subsequent extractions are much more easily achieved than during the erythrocyte preparation because of the tightly packed nature of the protein obtained at this early stage. Extract the aqueous phase and protein layer with an equal volume of phenol/chloroform by inversion for 10 mins. Centrifuge as before and extract the aqueous phase, with chloroform, as above. Aliquot the final aqueous phase into two sterile 30ml corex tubes on ice and precipitate the nucleic acid by the addition of 2 volumes of ice cold ethanol. The solution becomes immediately cloudy and nucleic acid flocculates. In contrast to the chromosomal preparation which comprises mostly high molecular weight DNA, the bulk of the DNA obtained from oocytes is amplified rDNA of a few repeat units in length (Chaper 1, 1.4.2). Centrifuge (2500g, 15 mins, 0°C) and dissolve the pellets in a total of 10mls of 1 x SSC. RNase and pronase digestion is performed as for erythrocyte DNA purification (4.3).

4.5 Purification of rDNA from total DNA

rDNA has a higher G + C content (63%) than bulk DNA (40%) and therefore has a higher bouyant density in caesium chloride (Wallace and Birnstiel, 1966). When total DNA is centrifuged to equilibrium in CsCl gradients, the rDNA component bands as a satellite on the heavy side of the bulk DNA band. I have purified rDNA from total DNA using a protocol based on the methods of Wellauer et al (1976b) and Buorngiorno-Nardelli et al (1977).

Measure the A_{260nm} of both oocyte and erythrocyte preparations. Increase the volume of the samples to give a DNA concentration of no more than 50µg/ml. Add solid CsCl to give a final density of $1.7g/cm^3$ then centrifuge to equilibrium using a Beckman Ti 60 rotor (33,000 rpm, 18°C, 62 hrs).

Pump the gradients through a Guilford 240 recording spectrophotometer, monitoring continuously at 260nm and collect about 30 fractions per gradient. Locate the rDNA component by hybridisation.

4.5.1 Hybridisation

The basic method used to detect rDNA was to bind aliquots of the fractions to nitrocellulose and detect rDNA by hybridisation to a HeLa ^{32}P-28SrRNA probe. However, because of the very small amounts of rDNA to be detected certain modifications to the standard procedure, previously described, were made.

Add 7.5µl of each fraction to be screened to 10µl of 0.1M NaOH in a titre plate. Mix. After 10 mins at room temperature, add 2µl of 0.5M Tris, 0.5M HCl. Draw a grid on 4 x 4cm detergent-free nitrocellulose filters and apply each sample to a single position on the grid. Because

of the large volume (20µl) apply the sample in 5µl aliquots allowing it to dry between applications. The DNA is then baked onto the filter in a vacuum oven for 2hrs at 80°C.

Hybridisation reaction : The filters are prehybridised as previously described. However the hybridisation volume is reduced to 2-3mls since the filters are much smaller than previously. A small hybridisation volume and high specific activity probe will aid in the detection of the nanogram amounts of rDNA bound to the filters. Washing and autoradiography of the filters is the same as before. Pool the fractions which give positive signals from the hybridisation, rerun the gradients (8ml/gradient) in a Beckman Ti 50 rotor at 33,000 rpm for 60 hrs at 18°C and collect fractions as before.

Although much of the total DNA is removed by the first centrifugation there is some carry through to the second gradient. This is especially true of the chromosomal rDNA preparation. Locate the rDNA after the second gradient by filter hybridisation reducing the volumes by half because of the reduced volume/fraction. Pool the fractions which give strong positive signals and remove CsCl by extensive dialysis against TE buffer pH 8.0 at 4°C (4 x 1ℓ changes). Store the DNA at -20°C as an ethanol precipitate after washing in 80% ethanol.

5. Analysis of rDNA by M13 Cloning and "Dideoxy" Sequencing

Restriction fragments of rDNA obtained from a single frog were analysed by using the versatile M13 cloning system and "dideoxy" sequencing method. The rationale behind this decision is described in Chapter 2.

5.1 Construction of M13 recombinant molecules

Preparation of vector and "insert" DNA and the ligation reaction are performed according to the protocol described in Section 2.1. A three times excess of insert to M13 RF was utilised to favour recombinant formation.

5.2 Transfection

5.2.1 Preparation of Competent Cells

Streak out stock JM103 onto a glucose minimal agar plate. Allow this plate to grow overnight at 37°C. This plate is used as a stock for single colonies for at least one month when stored at 4°C. Cultures of JM103 are always started by inoculation of 2 x YT medium with a single bacterial colony and grown overnight at 37°C.

Inoculate 25ml of 2 x YT medium with 2.5ml of overnight JM103 and grow the cells at 37°C until the OD_{660} is 0.3-0.4 units. Collect the cells by centrifugation at 4000g for 10 mins at 4°C. To obtain exponentially growing JM103 for later use (see below) inoculate 5ml of fresh 2 x YT medium with 50µl of the primary culture and incubate at 37°C. During the time the cells are being made competent, this new culture will grow to the density required for plating the transfection mix.

Resuspend the pelleted cells in 12.5ml of 50mM $CaCl_2$ and incubate on ice for 20 mins. Centrifuge as before and resuspend the cells in 2.5ml of 50mM $CaCl_2$. Keep on ice ready for use. Unlike competent HB101, JM103 lose their competency if frozen at -70°C. Fresh cells must be

prepared for each transfection experiment.

5.2.2 Plating of Recombinants

Add each ligation mix to 0.3ml of freshly prepared competent JM103. In addition add 1ng of uncut vector DNA (i.e. M13 RF) to a 0.3ml aliquot to check the transfection efficiency. Incubate all the mixes on ice for 40 mins. Heat shock for 2 mins at 42°C. While the cells are being shocked, to 3ml of soft agar at 45°C add 25µl 2% BCIG (Sigma) (25mg/ml in dimethylformamide (Sigma)), 25µl 2.5% IPTG (Sigma) (25mg/ml in water) and 0.2ml exponentially growing JM103 to provide a lawn of cells. Add this agar mix to the transformed cells and plate out on YT plates. Invert the plates and incubate overnight at 37°C. Recombinant molecules can be tentatively identified as white plaques on the plates whereas vector DNA containing no insert will give blue plaques.

5.3 Preparation of single stranded template for sequencing

It is convenient to prepare 24 templates at the same time. Add a drop of overnight JM103 to 25ml of 2 x YT. Aliquot the 2 x YT into 24 snap cap culture tubes. Individually toothpick a single white plaque into the 24 x 1ml aliquots and shake at 37°C for $4\frac{1}{2}$ - $5\frac{1}{2}$ hours. Pour into 1.5ml Eppendorf tubes and centrifuge (9950g) for 5 mins. Pour the supernatant into another Eppendorf making no effort to transfer completely (~0.8ml). The cell pellet is stored at 4°C and provides the starting material for

growth of recombinant RF if needed. Add 200µl 2.5M NaCl, 20% polyethyleneglycol (PEG) to the supernatant and leave at room temperature for 15 mins. Centrifuge (9950g) as before, pour off supernatant to waste, recentrifuge for 30 secs. to bring liquid off the walls of the tube, and carefully remove all the remaining traces of the PEG supernatant. The PEG pellet should be visible (although very small). Add 200µl of TE buffer pH 8.0 and phenol extract with 100µl of redistilled phenol. Add 20µl of 3M Na acetate pH 5.5, 500µl of ice cold ethanol and store overnight at -20°C. Centrifuge (9950g) for 5 mins then wash the pellet in 80% ethanol. Finally dissolve the pellet in 50µl TE pH 8.0 and store at -20°C.

As a quick check on whether the virus has grown or whether cloned DNA has been lost during the work up, run 5µl of DNA on a 1% agarose minigel and visualise the DNA.

5.4 DNA sequencing with chain terminating inhibitors

A primer annealed to a single stranded template can be extended in the 5'→3' direction by DNA polymerase I (Klenow fragment) in the presence of deoxynucleotide triphosphates (dNTPs). Klenow polymerase will also specifically incorporate ddNTPs in place of their dNTP counterparts, but the lack of a 3'-hydroxyl group on the sugar moiety of the ddNTP prevents further polymerisation. Thus the ddNTPs act as specific chain terminators. In four separate sequencing reactions, template and specific primer are incubated with Klenow fragment in the presence of a different ddNTP (ddTTP, ddATP, ddGTP, ddCTP), as well as the four dNTPs,

one of which is labelled with α^{32}-phosphate. Partial incorporation of the ddNTP results and the four reactions give products with a fixed 5' end (the primer) and terminating at either T or C or G or A.

In order to screen as many clones as possible in the time available, I performed only one of the four sequencing reactions. The clones constructed contained the entire ITS1 sequence (see Chapter 7) so the G sequencing reaction was chosen. The following sections describe the procedures involved in carrying out this rapid screening objective.

5.4.1 Sequencing solutions and buffers

(i) 10x annealing buffer : 100mM Tris pH 8.5, 100mM $MgCl_2$.

(ii) Stock dNTP/ddNTP solutions:

dNTPs were purchased from the Boehringer Corporation Ltd. and ddNTPs from P.L. Biochemicals.
Stock 20mM dNTPs in TE buffer pH 8.0, and 10mM ddGTP in TE pH 8.0 were kept frozen at -20°C.

(iii) Working "G reaction" mix

The degree of primer extension in each of the four reactions depends on the ratio of each ddNTP to its dNTP counterpart. Too much ddNTP will cause termination at too short a chain length whereas excess dNTP will dilute out the effect of the competing ddNTP and will not allow sufficient termination to occur. I found the following mixture of dGTP/ddGTP gave suitable chain termination : 500µl 0.5mM dTTP, 500µl 0.5mM dCTP, 25µl 0.5mM dGTP, 16µl 10mM ddGTP,

1000µl TE pH 8.0. This solution was stored at -20°C and reused for each screening experiment. dATP is added during the actual sequencing reaction in the form of high specific activity α-^{32}P-dATP (> 400 Ci/mmole, obtained from Amersham International plc).

(iv) Specific M13 primer :

The primer I used was a 17-mer described in Figure 2.4. The primer was kindly donated by A.T. Bankier but is now commercially available from Collaborative Research Inc.

(v) Klenow fragment :

DNA polymerase I was purchased from The Boehringer Corporation (London) Ltd. at a concentration of 1unit/µl.

5.4.2 The annealing reaction

The following volumes are for a single screening of 12 clones in parallel. Add 2µl aliquots of the clones to be sequenced to 12, 1.5ml Eppendorf tubes. To each add 2µl of primer mix (5µl 17-mer, 5µl 10x annealing buffer, 15µl H_2O, freshly prepared and stored at 4°C) and incubate at 60°C for 1hr.

5.4.3 The polymerisation reaction

Add 12µl of α-^{32}P-dATP to a siliconised glass tube and dry under vacuum. Redissolve thoroughly in 25µl of working "G reaction" mix. This is now called G* solution. Place

12 capless 1.5ml Eppendorf tubes into a plastic Eppendorf rack tilted at an angle of 60° from the horizontal. Add 2µl of each hybridisation mix to the rim of each of the 12 tubes. Add 2µl of G* solution ensuring the two drops of liquid do not mix. Finally add 2µl of freshly diluted Klenow fragment (0.125U/µl in H_2O) and centrifuge (9950g) for 5 secs to mix and start the polymerisation. Incubate at 30°C for 15 mins.

"Pausing" of chain extension due to the low molar concentration of dATP can also occur, so after the initial synthesis period with α-^{32}P-dATP the incomplete chains are extended by chasing with cold dATP (if other α-^{32}P-dNTPs are used then the chase is made with the corresponding cold dNTP). Thus the only short chains after a chase are those which underwent true termination due to incorporation of dideoxynucleotide.

Add 2µl of 0.25mM dATP to the rim of each tube, centrifuge as above and incubate at 30°C for 15 mins. The polymerisation is stopped by adding 4µl of formamide sequencing dye (3.4) and placing the tubes on ice prior to gel electrophoresis.

5.5 Gel Electrophoresis and Autoradiography

The same gel system was utilised as previously described for separation of Maxam and Gilbert sequencing products (3.3).

Prepare the samples for electrophoresis by boiling for 4 mins then chilling immediately on ice for 5 mins (replace

caps to prevent evaporation during boiling step). Load the 12 samples in parallel onto the gel (2µl/loading) and electrophorese until the xylene cyanol dye has travelled 2/3 way towards the anode. For a longer run, electrophorese until the dye reaches the bottom of the gel.

To improve resolution of the bands for autoradiography, fix the gel in 10% acetic acid/10% methanol for 10-15 mins. Transfer to Whatman 3MM paper by placing a wet piece of 3MM onto the gel and carefully peeling away - the gel adheres to the 3MM paper. Cover the gel with a layer of cling film and dry for 15-30 mins on a gel drier at 80°C. Remove the cling film and expose overnight using Kodak-X-Omat H-film (2.3.6).

4. Analysis of the 18S-28S intergene nucleotide sequence of pX1212

4.1 Subcloning region L

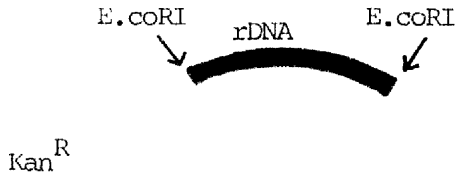
pX1212 is a large recombinant plasmid (17.5kb) containing a 4.5kb rDNA fragment cloned into the single Eco RI site of the plasmid pCRI (Figure 4.1). To simplify preparation of restriction fragments for sequence determination, I decided to subclone region L into pBR322.

Digestion of pX1212 with restriction enzymes Eco RI and Bam HI excises region L from the parent plasmid in addition to regions M and R (see Figure 4.1). Region L is bound by Eco RI and Bam HI restriction sites at opposite ends, allowing it to be cloned into pBR322 after removal of the small Eco RI/Bam HI fragment from pBR322 (Sutcliffe, 1978). Ligation was carried out as described (Chapter 3, Methods 2.1) and recombinant molecules selected on the basis of antibiotic resistance. Insertion of foreign DNA into this region of pBR322 inactivates the tetracycline gene, therefore bacteria harbouring recombinant molecules are Amp^R Tet^S (see Figure 3.1).

Small scale plasmid DNA preparations were made on four Amp^R Tet^S colonies and limited restriction analysis revealed all four plasmid DNAs to contain region L (Figure 4.2). A large scale DNA preparation was made from colony 1 and the plasmid denoted pX1212L.

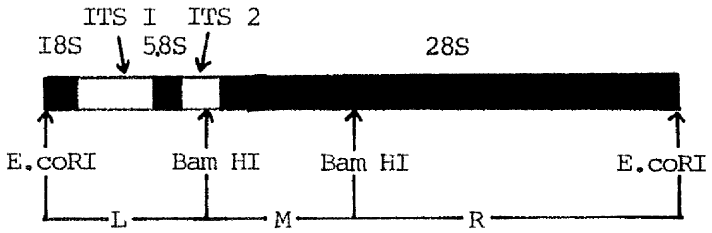
Figure 4.1 Preparation of region L from pX1212.

10µg of pX1212 were digested with Bam HI and Eco RI and the digestion products resolved on a vertical 1% agarose gel. The gel was stained with ethidium bromide and the DNA located by U.V. illumination. The band containing region L was carefully excised from the gel and eluted in preparation for subcloning.



pX1 2I2

I7.5Kb.



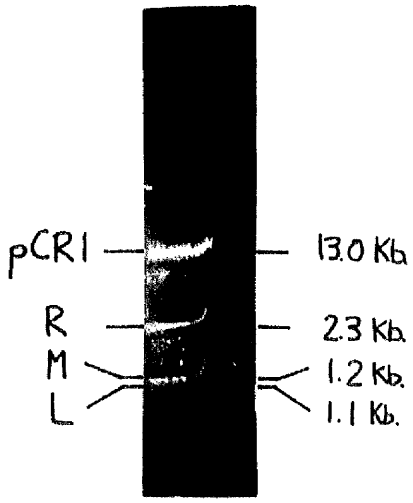


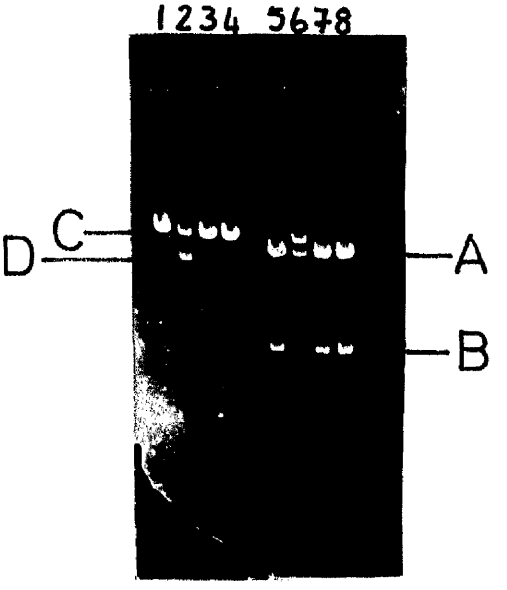
Figure 4.1

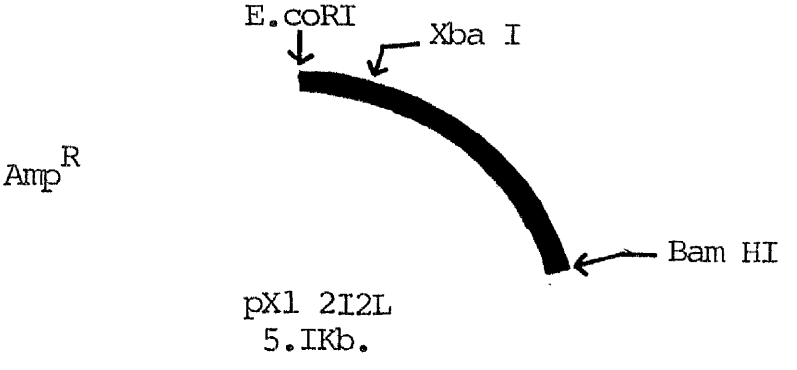
Figure 4.2 Identification of recombinant molecules.

The DNA from 4 Amp^R Tet^S colonies was digested with Xba I (lanes 1-4) or with Eco RI and Bam HI (lanes 5-8) and the products resolved on a vertical 1% agarose gel.

Digestion with Eco RI and Bam HI releases a large DNA fragment containing the parent plasmid pBR322 (A) and a smaller fragment (B) corresponding to region L. Xba I has one recognition site in region L and no sites in pBR322. Digestion with Xba I linearises a recombinant molecule, the linear form (C) running more slowly than the uncut circular form (D).

The extra bands in the lanes 2 and 5 represent incomplete digestion of the DNA in these lanes.





pX1 2I2L
5.IKb.

Figure 4.2

4.2 Sequence determination

The complete sequence from the Eco RI site in the 18S gene through to the start of the 28S gene was derived from a large number of individual determinations on overlapping restriction fragments. This region had already been fully sequenced in a different rDNA clone - pXlrl01 (Hall and Maden, 1980) and a detailed restriction map was available. Knowledge of the sequence revealed previously unmapped restriction sites which simplified this present sequence analysis e.g. Bst NI. Figure 4.3 illustrates the restriction enzyme sites used in this sequence determination and the direction of sequencing runs extending from those sites.

pXl212L provided the primary source of DNA for the restriction fragments extending from the Eco RI site in the 18S gene to the Bam HI site in ITS 2. The parent plasmid, pXl212, was used to obtain restriction fragments extending from the Bam HI site to the start of the 28S gene. In order to read through the Bam HI site in ITS2, a 420bp Hinf I fragment from a complete Hinf I digest of pXl212 was isolated and sequenced (Figure 4.3, gel no. 12). Isolation and sequencing of this fragment from pXlrl01 had been previously achieved (Hall, 1981).

All of the fragments in this sequence analysis were prepared by digesting a pure 5'-end labelled DNA fragment asymmetrically. After secondary restriction the fragments can again be separated on a 4% acrylamide gel and the bands located directly by staining with ethidium bromide.

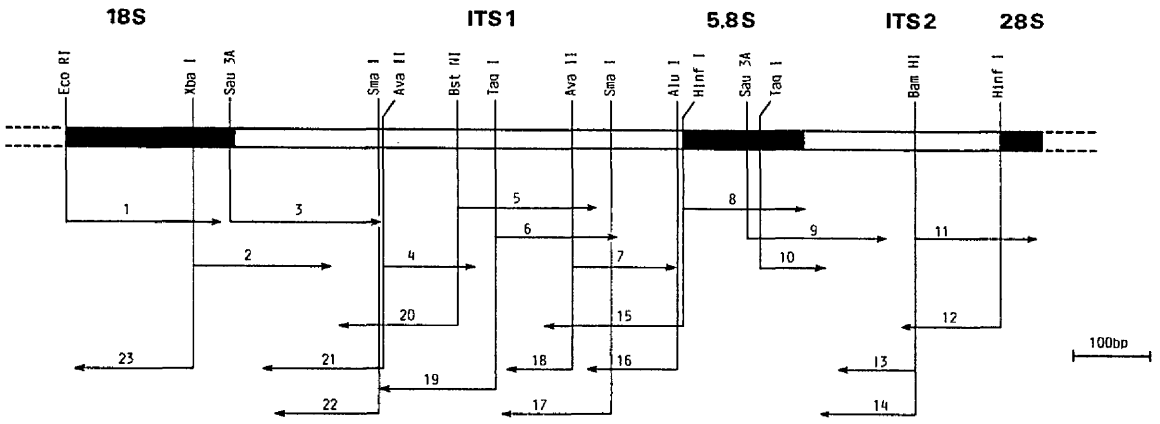


Figure 4.3. Sequencing strategy for the 18S-28S intergene region of pX1212.

Only those sites used for labelling following primary restriction are shown. The beginning of each arrow denotes the restriction site at which the fragment was labelled at the 5'end. The arrow tip denotes the longest reading of all the gels starting from the indicated restriction site. The arrows are numbered sequentially along the rightwards and leftwards DNA strands and serve as reference numbers for the text. 75% of the sequence was determined on both strands, the remaining 25% unambiguously on one or other strand. The procedure for sequencing end-labelled DNA is as described (Methods 3.2.). The length of sequence which can be read from each gel depends on the resolution of individual gels and the extent of attack in each chemical modification.

Figures 4.4 and 4.5 show respectively, the series of steps involved, and the sequence information obtained from a typical sequencing experiment.

Region L was digested with Bst NI or Ava II and the products separated on a 4% acrylamide gel (Figure 4.4(i)). Each of the required fragments was end labelled, digested with an appropriate restriction enzyme and then separated on a second gel (Figure 4.4(ii)). Two of the sequencing gels obtained are shown in Figure 4.5. The line diagram at the bottom of Figure 4.5 illustrates the total amount of information obtained from this complete sequencing experiment. Any gaps in the sequence are filled in by sequence determinations on overlapping restriction fragments.

4.3 Preliminary sequence comparison between pX1r101 and pX1212.

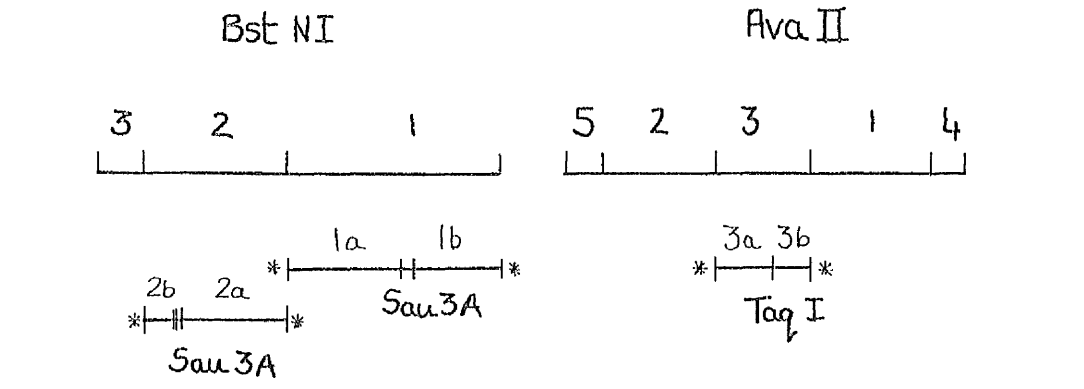
As mentioned previously several specific sites of sequence variation were thought to exist between pX1r101 and pX1212. The regions in question covered the last 80 nucleotides of ITS1 and the first 36 nucleotides of ITS 2.

In order to study both these regions in pX1212, a Hinf I digest of L was performed. Hinf I has only one recognition site in region L, located at the start of the 5.8S gene (Figure 4.6). Sequencing the ITS1-containing Hinf I fragment leftwards from the Hinf I site leads directly into the 3' end of ITS1. In addition, by sequencing the 5.8S-containing fragment, rightwards from the Hinf I site, I hoped to be able to read through the 5.8S gene to the start of ITS2. This however proved to be rather overambitious

Figure 4.4 Preparation of restriction fragments for sequencing - an example.

10µg of region L weredigested with Bst NI and 5µg of region L digested with Ava II and the products separated on a 4% polyacrylamide gel (i). The background bands in both digests are due to incomplete digestion of region L. In both cases a Taq I digest of pBR322 was run in parallel to aid in identification of required restriction fragments.

The appropriate bands were eluted, 5'end labelled and digested with a second enzyme as illustrated. The products were then separated on a second 4% polyacrylamide gel (ii). (The secondary restriction products of the Ava II digest are very faint but could be clearly seen on the original gel).


Bst NI
Ava II
3
2
1
5
2
3
1
4
1a
1b
2b
2a
Sau3A
Sau3A
3a
3b
Taq I

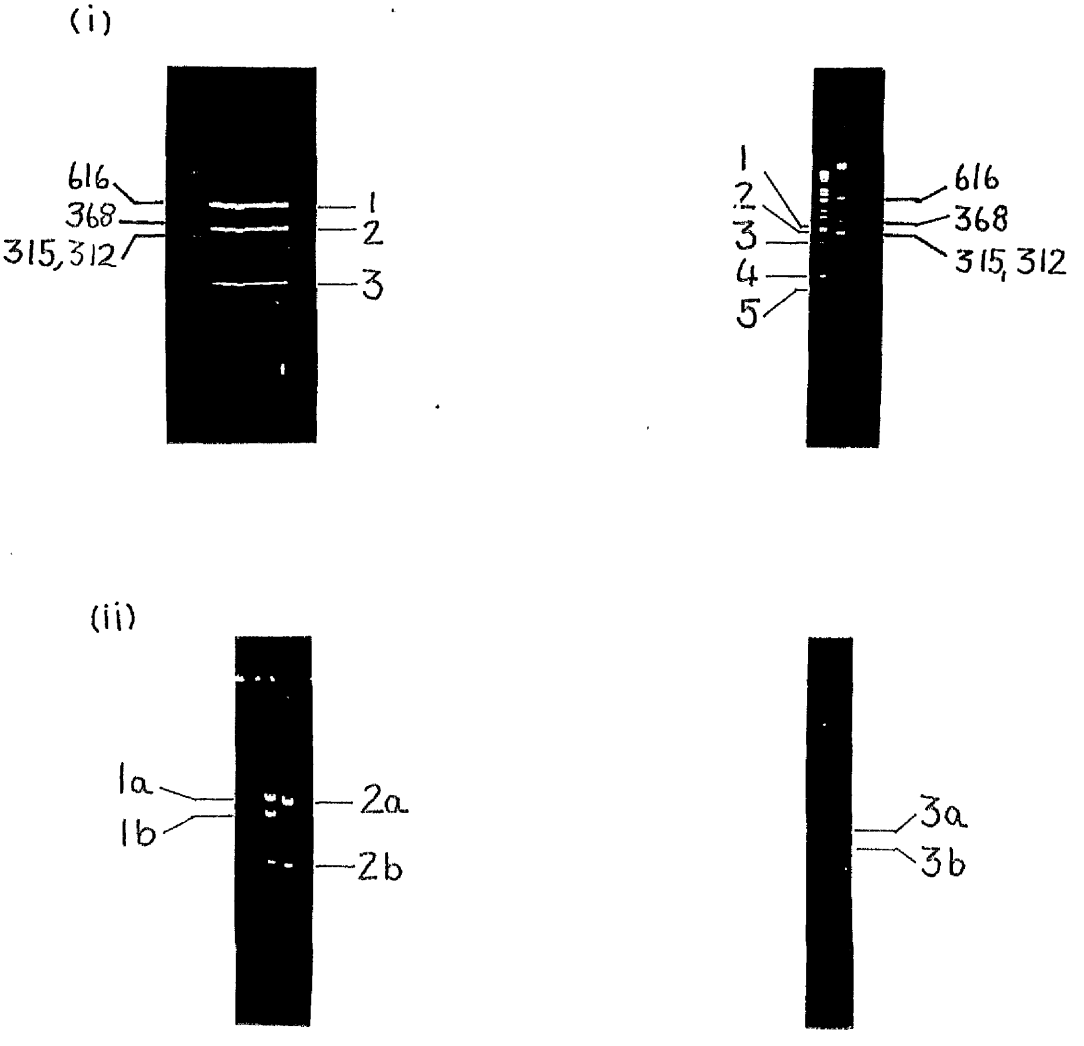

(i)
616
368
315,312
1
2
3
1
2
3
4
5
616
368
315,312
(ii)
1a
1b
2a
2b
3a
3b

Figure 4.4

Figure 4.5 Sequencing gels for *Ava II/Taq I and *Bst NI/Sau 3A restriction fragments

This figure shows the three overlapping loadings from two overlapping sequencing gels spanning the central region of ITS1. The Ava II/Taq I fragment corresponds to fragment 3a in Figure 4.4, the Bst NI/Sau 3A fragment corresponds to fragment 1a in Figure 4.4. After 5' end labelling and secondary restriction, each fragment was subjected to full sequencing chemistry using the chemical reactions specific for G, G+A, C+T and C (Methods 3.2). After chain cleavage with piperidine, the sequencing products were resolved on 8% denaturing polyacrylamide gels.

Nucleotides are numbered according to the individual numbering system of pXlr101 (see Figure 4.8). Uncertain nucleotides are described according to Staden (1979). "D" denotes more than one C, "H" denotes more than one G, ⌒ denotes order of nucleotides is unclear. In problem regions, the sequence was determined unambiguously on the opposite strand. The line diagram at the bottom of the Figure illustrates the total overlapping sequence information obtained from fragments 1a and 3a (above) and fragments 2a and 3b (Figure 4.4).

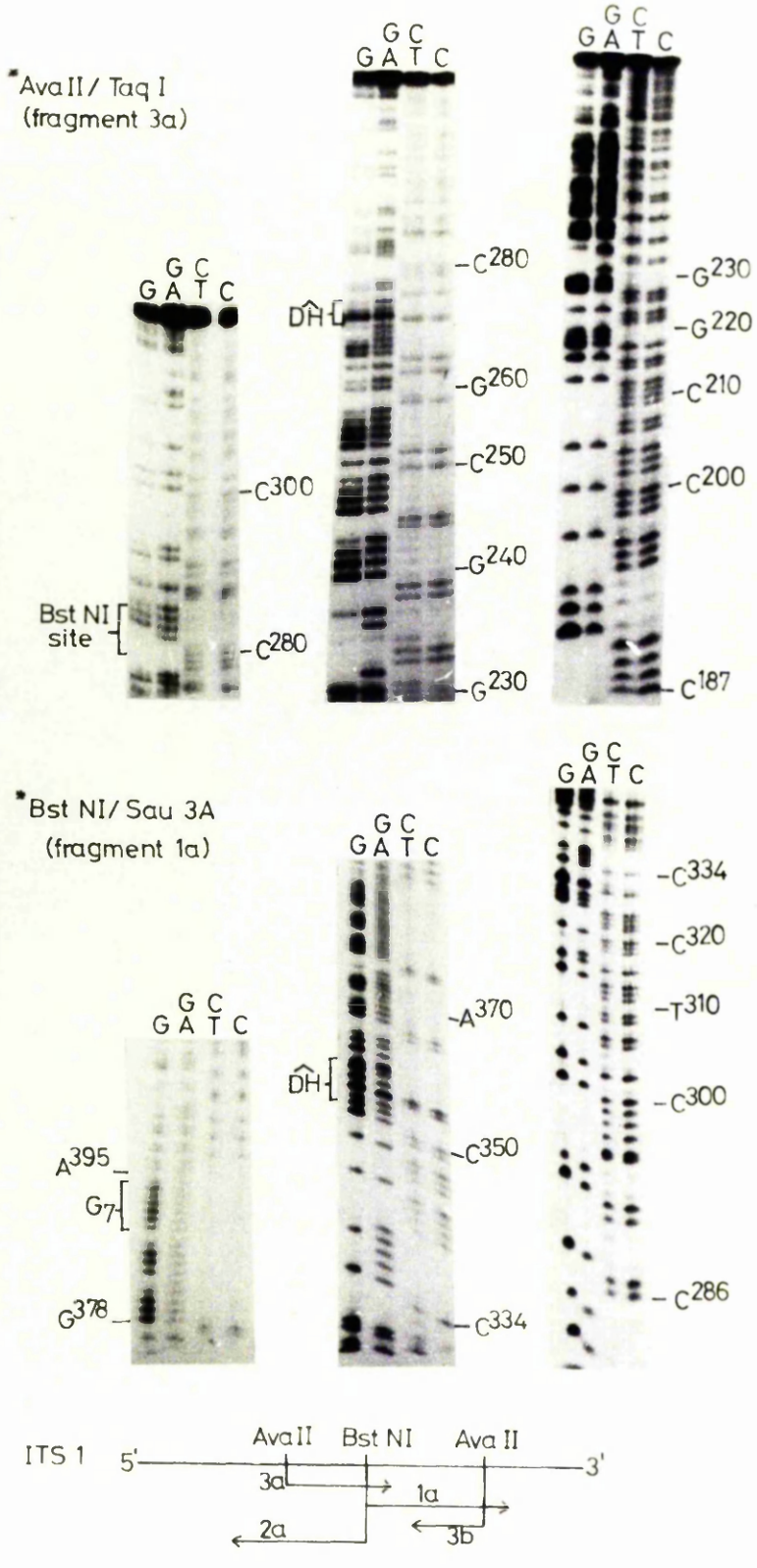

*Ava II / Taq I
(fragment 3a)
G C
G A T C
C280
DH
G260
C250
G240
G230
C300
Bst NI
site
C280
G230
G220
C210
C200
C187
*Bst NI / Sau 3A
(fragment 1a)
A370
C350
C334
A395
G7
G378
C334
C320
T310
C300
C286
ITS 1
5'
3'
Ava II
Bst NI
Ava II
3a
1a
2a
3b

Figure 4.5

(see below).

The sequencing gel obtained from the leftwards sequencing determination is shown in Figure 4.6. Comparison of the sequence with the respective sequence in pXlrl01 revealed 100% homology in the regions of suspected sequence variation. Repetition of the sequence determination on the opposite strand (Figure 4.3, gel no. 7) confirmed this. The use of thin gels for electrophoresis has greatly improved resolution of sequencing products. Since there are no apparent ambiguities in my data I feel confident that the sequence determined for pX1212 in this project is correct.

However, analysis of the pX1212 sequence further upstream revealed a definite alteration of the sequence when compared to the same region in pXlrl01. Careful comparison showed that pX1212 contains a segment of DNA that is not present in pXlrl01. Determination of the nucleotide sequence of this DNA was hampered because of clustering of bands in the gel at this region. This problem was also encountered in other sections of both ITS1 and ITS2. Such compression of bands and uneven spacing of adjacent bands makes it difficult to tell the number and order of the nucleotides. The distortion is due to strong local secondary structure which is not fully eliminated even under the denaturing conditions of the gel. However, in most cases the precise sequence can be elucidated if sequence determinations are carried out on both strands of the "problem" DNA.

Figure 4.6 Identification of an "extra" block of nucleotides in pX1212.

Full DNA sequencing was performed on a labelled Taq I/ Hinf I fragment containing the 3' portion of ITS1. The figure shows the three overlapping loadings from the same sequencing gel.

ITS1 and ITS2 have been designated individual numbering systems in the reference clone pX1r101 and the nucleotides numbered in the figure correspond to this system. Nucleotide 557 is the last nucleotide in ITS1.

* Indicates positions in the sequence which were originally thought to be sites of variation between pX1r101 and pX1212. Comparison of the full sequences for these two clones in this region revealed 100% homology.

The DNA sequence shown at the bottom of the figure represents the sequence determined in pX1212, extending leftwards from the Hinf I site in the 5.8S gene. Homologous sequences in pX1r101 are indicated by a continuous line. The precise number and nucleotide sequence of the extra DNA in pX1212 cannot be elucidated from this gel alone. Uncertain nucleotides are described according to Staden (1979). 'D' denotes more than one C, 'H' denotes more than one G, and ^ denotes order of nucleotides is unclear.

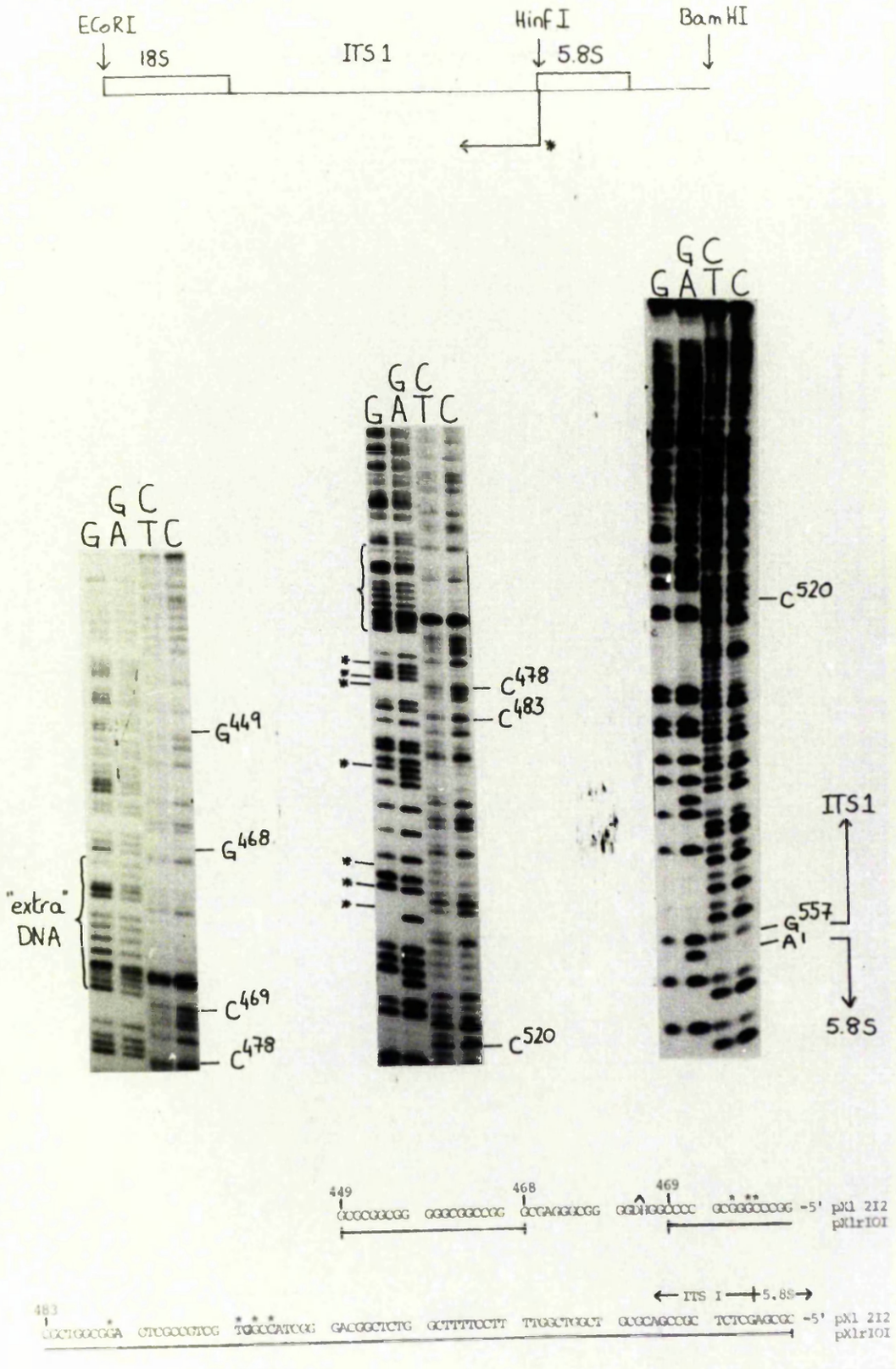

EcoRI
18S
ITS 1
HinfI
5.8S
BamHI
G C
G A T C
G449
G468
"extra" DNA
C469
C478
C478
C483
C520
C520
ITS1
G557
A
5.8S
449
468
469
-5' pX1 2I2
pX1rIOI
ITS I
5.8S
483
-5' pX1 2I2
pX1rIOI

Figure 4.6

Figure 4.7 Determination of "extra" DNA sequence in pX1212.

The figure illustrates the appropriate regions of three sequencing gels which together allow the unambiguous determination of the extra DNA sequence identified in pX1212. Gel (i) shows sequencing of the insert on the rightwards strand extending from the Ava II site shown. Gels (ii) and (iii) are overlapping sequencing runs extending from the Alu I site at the extreme 3' end of ITS1 and an Sma I site at the extreme 3' end of the extra sequence, respectively. The code for uncertain nucleotides is the same as described for Figure 4.6. Nucleotides are numbered according to the pX1r101 numbering system (see Figure 4.8) and the extra block of 20 nucleotides designated 468^{1}-468^{20}.

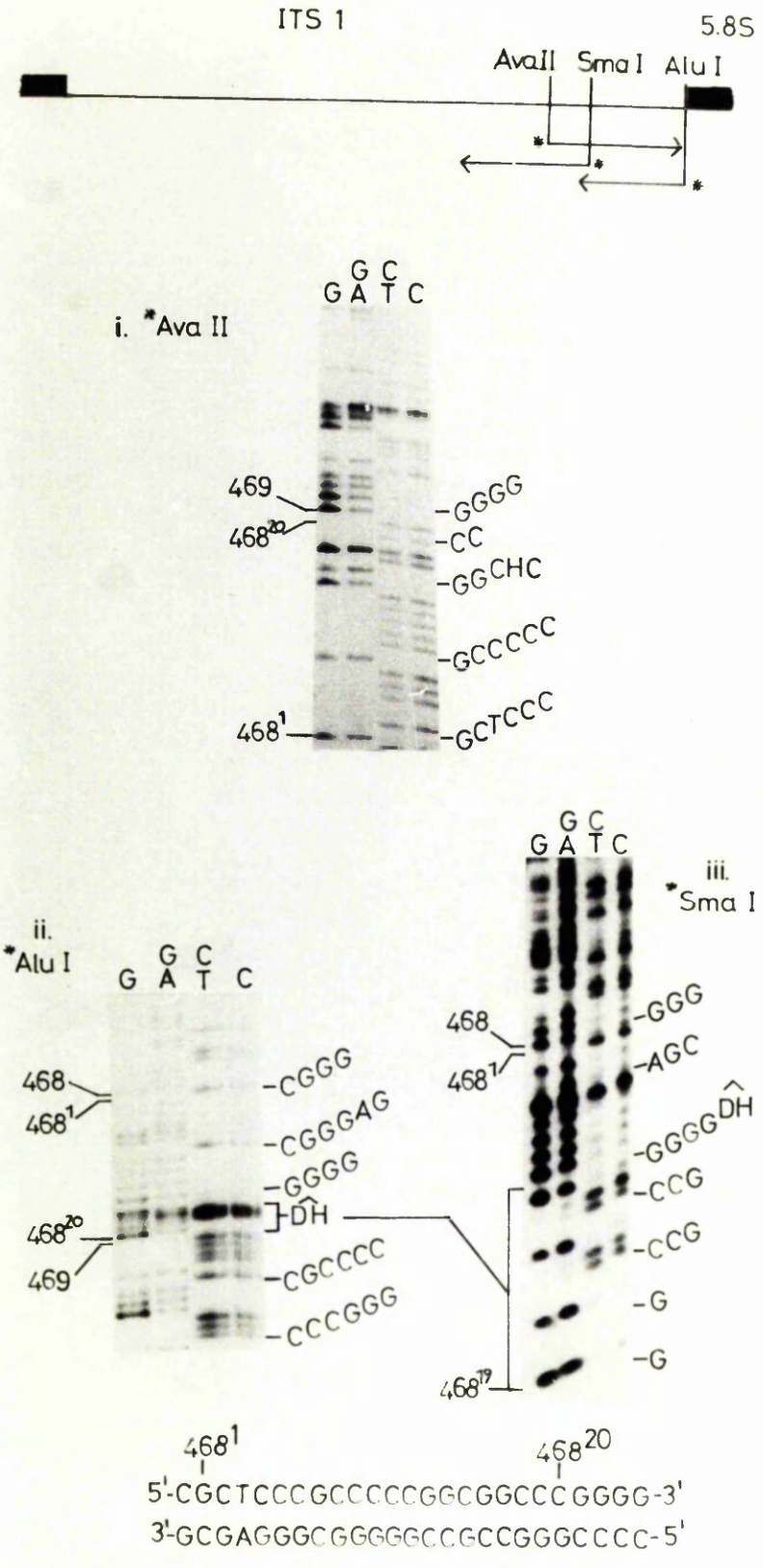

ITS 1
5.8S
AvaII
Sma I
Alu I
i. *Ava II
G A T C
469
468^20
468^1
-GGGG
-CC
-GGCHC
-GCCCCC
-GCTCCC
ii.
*Alu I
G A T C
468
468^1
468^20
469
-CGGG
-CGGGAG
-GGGG
-DH
-CGCCCC
-CCCGGG
iii.
*Sma I
G A T C
468
468^1
468^19
-GGG
-AGC
-GGGGDH
-CCG
-CCG
-G
-G
468^1
468^20
5'-CGCTCCCGCCCCCGGCGGCCCGGGG-3'
3'-GCGAGGGCGGGGGCCGCCGGGCCCC-5'

Figure 4.7

Figure 4.7 illustrates the sequencing strategy and gels required to determine the sequence of the extra DNA found in pX1212. The segment of DNA is 20bp in length and like the rest of ITS 1 comprises largely of guanine and cytosine. It is located at nucleotide 468 in ITS 1 and increases the length of ITS1 from 557 (pX1r101) to 577 nucleotides.

Sequencing the start of ITS2 using the 5.8S-Hinf I fragment (Figure 4.6) was unsuccessful due to the length of run required to reach ITS2 - the 162 nucleotides of the 5.8S gene must be read through first. However, a different sequencing strategy was employed and restriction fragments were generated which allowed an unambiguous sequence determination of the start of ITS2 to be achieved (Figure 4.3, gel nos. 9 + 10). Comparison with the same region in pX1r101 again revealed 100% homology of the sequences.

4.4 Comparison of the entire 18S-28S nucleotide sequence of pX1212 with the reference sequence from pX1r101

4.4.1 rRNA coding sequences

In total, 419 nucleotides of gene sequence was analysed in pX1212 : the 3' end of the 18S gene extending from the Eco RI site, the entire 5.8S gene and the first 20 nucleotides of the 28S gene. Comparison of the sequence data with the reference sequence revealed 100% homology in the rRNA coding sequences analysed.

Figure 4.8 Summary of comparative data in the internal transcribed spacers of pXlrl01 and pXl212.

The figure shows the sequence of the RNA-like strand of ITS 1 and ITS 2 in the reference clone pXlrl01 and pXl212. A continuous line represents complete sequence homology of pXl212 with the reference sequence. Dashes indicate deletions with respect to the longer sequence examined.

Individual numbering systems have been adopted for ITS 1 and ITS 2. Two sites of variation occur in ITS 1: at position 29 and immediately following nucleotide 468 in the reference sequence. Variation also exists in ITS 2: at position 193 and positions 225 and 257. The latter two sites occur in the homopolymeric C tracts just upstream from the start of the 28S gene.

INTERNAL TRANSCRIBED SPACER 1.

```
pXlr101 ACGAGACCCC CCTCACCCGG AGAGAGGGAA GGCGCCCGCC GCACCCTCCC CGCGGAGAGA      60
pXl 212 |----------------------------C|----------------------------|

pXlr101 GAGAGAGACG CCCGCCCCGG AGCGGAGACC GCCCCCCCCC CACGGGGGGG GGGCGGCCGC     120
pXl 212 |----------------------------|----------------------------|

pXlr101 CCCGAAAGGG ACGACGAGGA ACCCCCAGAC GGCCCCGGCG AGGGGGCGGC GGCGGCCCCG     180
pXl 212 |----------------------------|----------------------------|

pXlr101 GGTCCACCCC GGGCCCGCCC GCCCGCCTCC CCGCCGCGGG CCCGCCCGGG TACCTAGCCG     240
pXl 212 |----------------------------|----------------------------|

pXlr101 GGGCCGGGGC GCGGGGGCTG GCGCGGGAGC GGGGCGGCCC CAGGGCCGTC CGGCCTCCCC     300
pXl 212 |----------------------------|----------------------------|

pXlr101 GCGTCCGCCT CCCGCGACCC GCCCCGGGCG GTTCGAAGAC CCCGCCCGCC GGGCGGCGGG     360
pXl 212 |----------------------------|----------------------------|

pXlr101 AGGGCCGGGA GGGAGCCGGG GAGGGGAGGG GGGGAGGCGG CGGCGAGCCC GGCCGGGCGC     420
pXl 212 |----------------------------|----------------------------|

                                                  468                469
pXlr101 CGCCGCCGCA GGACCCCCGT CCCCGTCCCG CGCCGCCCCC GCCGGCCC- ---------- ----------GG GGCGCCCGGG
pXl 212 |----------------------------|----------------------------|GC TCCCGCCCCC GGCGGCCC----------|

pXlr101 CCGCGACCGC CTCAGCGGCA GCACCGGTAG CCCTGCCGAG ACCGAAAAGG AAAACCGACC     540
pXl 212 |----------------------------|----------------------------|

pXlr101 GACGCGTCGG CGAGAGC                                                  557
pXl 212 |----------|
```

INTERNAL TRANSCRIBED SPACER 2.

```
pXlr101 GACGTCCATC GCCCCCGCCG GGTCCCGTCC CGGCGCGGAG GCGCGGCTGG GGCCGTCGCA      60
pXl 212 |----------------------------|----------------------------|

pXlr101 GGGGCGCGCC GCTCCCCTTC GTCCCCCCAA GGCCAGACCC CCGGCCCGGC GCCCGGGCCC     120
pXl 212 |----------------------------|----------------------------|

pXlr101 CGGCCCGGCC GGCGGCGGCT GTCTGTGGAT CCCTTCACGG CTGCCCCCCC GGCCGGCCCC     180
pXl 212 |----------------------------|----------------------------|

pXlr101 CCGGGGCCCC GGCCCGCCGG CGGGAGCGGG CCCGGCCCCC CCCCCCGGGC CGCGGCCCCG     240
pXl 212 |----------|G|----------------------------|--|----------|

pXlr101 CGCCCCCCCC CCCCCCCACG AC                                           262
pXl 212 |----------|-|----|
```

Figure 4.8

4.4.2 ITS1

Two sites of variation exist in ITS1 of pX1r101 and pX1212 (Figure 4.8(i)). At nucleotide 29, pX1212 contains C whereas pX1r101 contains A. This change occurs in a purine rich tract which shows no obvious secondary structure. The second site of variation occurs immediately following nucleotide 468 in the reference sequence. pX1212 contains an extra block of 20 nucleotides (denoted 468^{1}-468^{20}) as mentioned above.

4.4.3 ITS2

Between positions 1 and 192 the sequence of pX1r101 and pX1212 correspond exactly. At nucleotide 193, pX1212 contains C whereas pX1r101 contains G. This base change generates a new Sma I recognition site in pX1212. Hall and Maden (1980) reported the presence of two long homopolymeric tracts of C residues in ITS2. These unusual sequences both occur in pX1212 although the number of residues in each tract differs from the corresponding numbers in pX1r101 (Figure 4.8(ii)).

4.5 Conclusions

Determination of the nucleotide sequence of ITS1 and ITS2 from this second rDNA clone, pX1212, has revealed sequence variation in the form of single base changes and length variation. This finding of spacer variation immediately raises three questions :

1) Could this variation have arisen during the cloning of the rDNA ?

2) If not, is transcribed spacer variation a common phenomenon in X. laevis rDNA ?

3) Do variations in amplified rDNA truly represent variations in chromosomal rDNA?

The following two chapters describe the experiments performed to answer these questions.

5. Comparative sequence analysis of cloned and uncloned amplified rDNA.

My search for further sequence variants covered part of the external transcribed spacer and the whole of both internal transcribed spacers.

5.1 Isolation of restriction fragments from cloned rDNA

A description of the clones used in this comparative study is given in Figure 5.1. To simplify preparation of restriction fragments for sequencing, subclones were used when available. Selected control experiments showed that no variation between subclones and parent clones existed. Restriction digests of subclones were resolved on 4% polyacrylamide gels and digests of parent clones on 1% agarose gels. The appropriate fragments were eluted from the gels (Chapter 2, Materials 1.5.2 and 1.5.4) and prepared for sequencing by 5'-end labelling and secondary restriction.

5.2 Isolation of restriction fragments from uncloned rDNA

Approximately 20µg of uncloned amplified rDNA were donated by D.D. Brown. This sample was part of a larger preparation isolated from the ovaries of 50-100 small frogs and purified by banding in caesium chloride according to Dawid et al., (1970). After dialysis (by B.E.H. Maden) and rather prolonged storage, the DNA was at first difficult to restrict. However subsequent phenol extraction and ethanol precipitation rendered the rDNA susceptible to restriction. The purity of the rDNA was confirmed by analytical restriction of small aliquots with

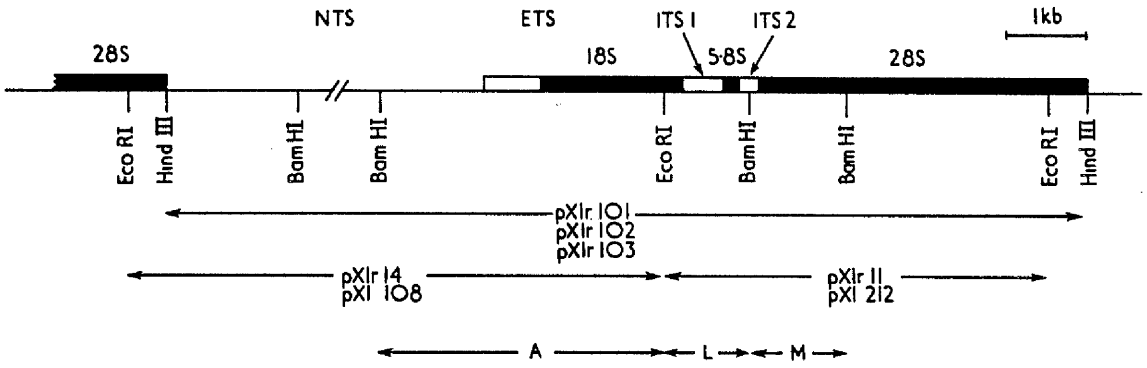


Figure 5.1. Description of amplified rDNA clones in comparative sequence study.

Slightly more than one repeating unit of rDNA from X.laevis is shown. The restriction sites used for cloning and the rDNA clones used are indicated.

pXlr101, 102 and 103 contain a complete repeat unit bound by Hind III sites. They therefore each contain the ETS and both ITS1 and ITS2 from a single transcription unit. pX1212 and pXlr11 contain only ITS1 and ITS2. pX1108 and pXlr14 contain the ETS.

To simplify preparation of restriction fragments, subclones containing regions A,L or M were used in preference to the parent clones.

enzymes that cut rDNA at known sites e.g. Eco RI. The expected bands were seen with no significant contamination.

The amount of uncloned rDNA was quite small so, to make maximum use of the material, I wished to restrict the DNA so as to produce several resolvable fragments containing appropriate regions (i.e. ITS1, ITS2, ETS) of the rDNA repeating structure. A survey of numerous restriction sites which had been mapped in one or more rDNA clones (Botchan et al., 1977; Boseley et al., 1978, 1979) or inferred from sequence data (Hall and Maden, 1980) showed that combined digestion with Bam HI and Xba I would yield a suitable array of products. After a successful trial digest on a small quantity of material, a large scale digest was carried out. The design of the experiment and the results obtained are shown in Figure 5.2.

After their recovery from the gel, the appropriate fragments or aliquots therefrom were used directly for full or, partial sequencing runs or after further restriction and fragment separation, to give access to additional sites for sequencing. Secondary restriction of end-labelled fragments was carried out on the basis of known restriction sites in clones, yielding the expected fragments usually in just sufficient quantity for visualisation in polyacrylamide gels after ethidium bromide staining.

Figure 5.2 Restriction of uncloned amplified rDNA.

Upper section : Map of Bam HI and Xba I sites in rDNA, initially predicted from data from clones (see text). Lines with arrow heads represent fragments obtained from complete digestion of uncloned rDNA with Bam HI and Xba I. Fragments 1-6 are numbered according to size. Fragments 2a and 2b and fragments 6a, 6b and 6c are all variants due to heterogeneity in the location of Bam HI sites in the NTS.

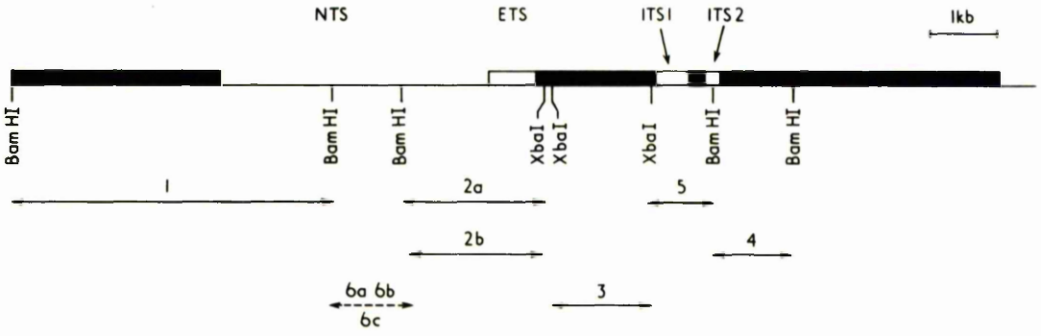
Lower section : Restriction digests of rDNA. Restriction products were separated on a 1% vertical agarose gel and visualised, after ethidium bromide staining, by U.V. illumination. The wide central section contains the preparative Bam HI plus Xba I digest of uncloned rDNA. The fragment numbers correspond to those in the upper section : the relationship of the fragments to the rDNA repeating structure was inferred from the known distances between restriction sites in cloned rDNA and confirmed by markers in Lanes 1-3.

Lane 1: complete digestion of pX1r101 with Bam HI + Xba I.

Lane 2: complete digestion of pX1r101 with Xba I.

Lane 3 : complete digestion of pX1212 with Bam HI.

The top two fragments in lane 1 and the top fragment in each of lanes 2 and 3, contain some material from the respective vector linked to rDNA. The more rapidly migrating fragments contain only rDNA. The sizes of these fragments are indicated on the left (including a very small Xba I fragment 7 from the 18S gene not numbered in the upper section).


NTS
ETS
ITS1
ITS2
1kb
Bam HI
Bam HI
Bam HI
Xba I
Xba I
Xba I
Bam HI
Bam HI
1
2a
5
2b
4
6a 6b
6c
3

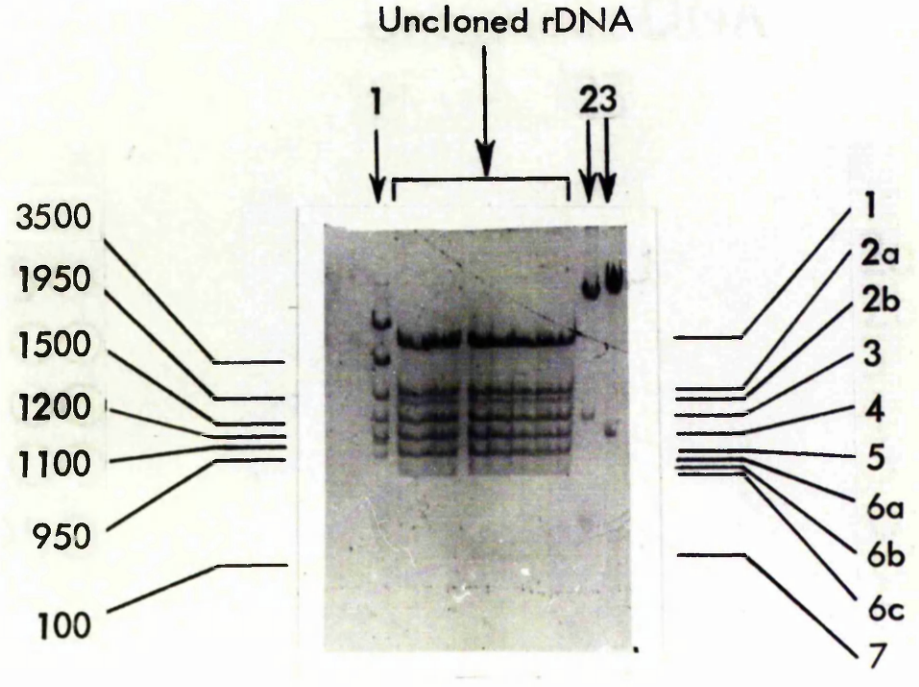

Uncloned rDNA
1
23
3500
1950
1500
1200
1100
950
100
1
2a
2b
3
4
5
6a
6b
6c
7

Figure 5.2

5.3 Comparative partial chemical cleavage

In order to screen rapidly for sites of variation between clones and heterogeneity in uncloned rDNA, corresponding restriction fragments from the various rDNA sources were examined in parallel. After 5'-end labelling and secondary restriction the fragments were subjected to partial G modification or partial C modification. Subsequent manipulations were the same as used for full sequencing runs. The choice between the G or C reactions was made so as to maximise the amount of useful information that could be obtained, on the basis of data from the reference clone and on any preliminary data from other clones. Where suitable, G cleavage was preferred because hydrazine (C specific) can cross react weakly with G giving weak extra bands in the cleavage pattern. If sequence variation was thought to exist on the basis of the single cleavage experiment then, whenever possible, a full sequencing run was performed on the DNA in question.

5.4 Comparative sequence analysis

Comparative sequencing studies were carried out first on both internal transcribed spacers because my preliminary studies had already revealed several sites of variation between two rDNA clones. Figure 5.3 depicts the analyses that were performed on the 18S-28S region of five clones and uncloned amplified rDNA.

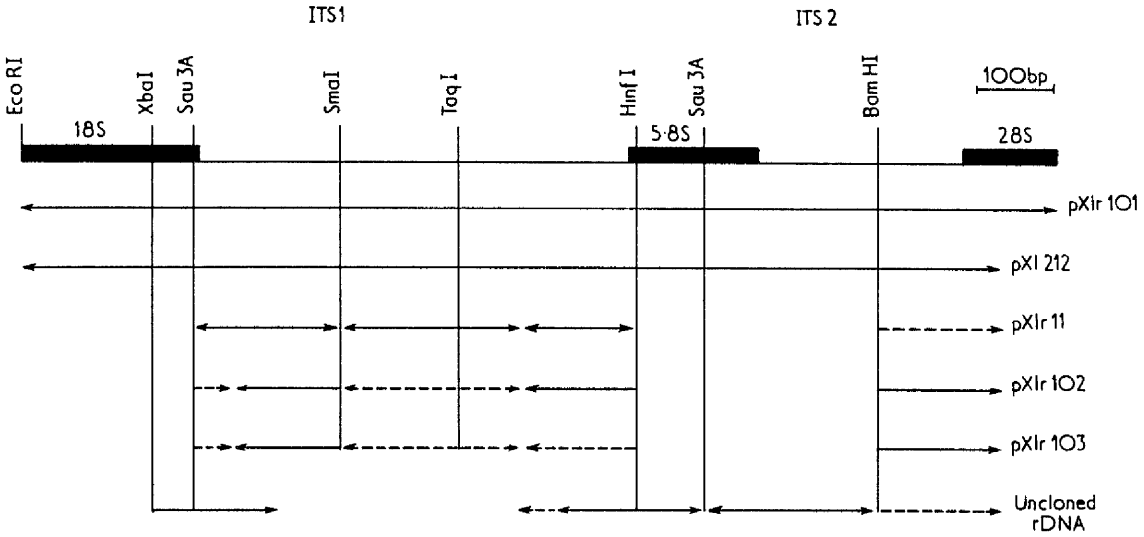


Figure 5.3. Summary of sequencing runs on the 18S-28S intergene region of five cloned rDNAs and uncloned amplified rDNA. pXlr101 was fully sequenced in this region as described by Hall and Maden, 1980. pXl212 was fully sequenced in this region during the course of this work. pXlr102, pXlr103, pXlr11 and uncloned rDNA were subjected either to full sequencing runs (continuous arrows) or to G or C cleavage (interrupted arrows) from the restriction sites shown.

5.4.1 ITS1

Initial studies on ITS1 concentrated on the 5' and 3' ends and were then extended into the internal section whenever possible (Figure 5.3). Several sites of variability were found. At nucleotide 29, pX1212 contains C whereas all the other clones contain A. Next, in the vicinity of nucleotide 100, four out of the five clones contain a tract of 10 cytosines followed after two nucleotides by 10 guanines. pX1r11 contains 11 of each of these nucleotides (Figure 5.4). The number of guanines was determined by sequencing the complementary strand because the G bands are compressed in sequencing gels of the rightwards strand. Beyond the compression effect, sequencing gels of cloned rDNA become readable again. pX1r11 also contains an extra C following C366 of the reference sequence (Figure 5.4). This gives rise to an extra Sma I site which was experimentally demonstrated (Hall, 1981).

Uncloned amplified rDNA (band 5, Figure 5.2) was analysed rightwards from the Xba I site at the 3' end of the 18S gene by a full sequencing run. Despite the small amount of uncloned material available, full sequencing experiments were very successful (with one exception, see rRNA coding regions) and gave clear sequence readings. Part of the Xba I full sequencing gel is shown in Figure 5.5. At position 29, A shows less strongly than neighbouring adenosines and is partly replaced by C. Normally A and C do not cross react in such a way (see elsewhere in the gel) and the unique occurrence of this dual reaction corresponds to the site of

Figure 5.4 Summary of comparative data from amplified rDNA for ITS1 (rightwards RNA-like strand).

The figure shows the complete sequence in pX1r101 and the variations between rDNA sources. Dashes indicate deletions with respect to the longest sequence examined. Plus signs for uncloned rDNA means that part of the population contains extra bases at this point. A shaded box signifies that the indicated rDNA sources were subject to complete sequencing runs, with identical findings to pX1r101 in the boxed region. An open box signifies that G cleavage runs were carried out, with no differences evident from pX1r101 in the boxed region. The zig-zag lines for uncloned rDNA indicate partly unresolved cleavage patterns due to length heterogeneity at the origins of the lines. Unambiguous sequence determination of the block of 20 nucleotides was achieved only in pX1212. "H" in the remaining clones and uncloned rDNA signifies more than one G and the order of nucleotides between position 468^{13}-468^{17} was not resolved due to compression of nucleotides in the gels obtained. "X" in uncloned rDNA indicates heterogeneities for which no corresponding variants were found in the amplified rDNA clones.

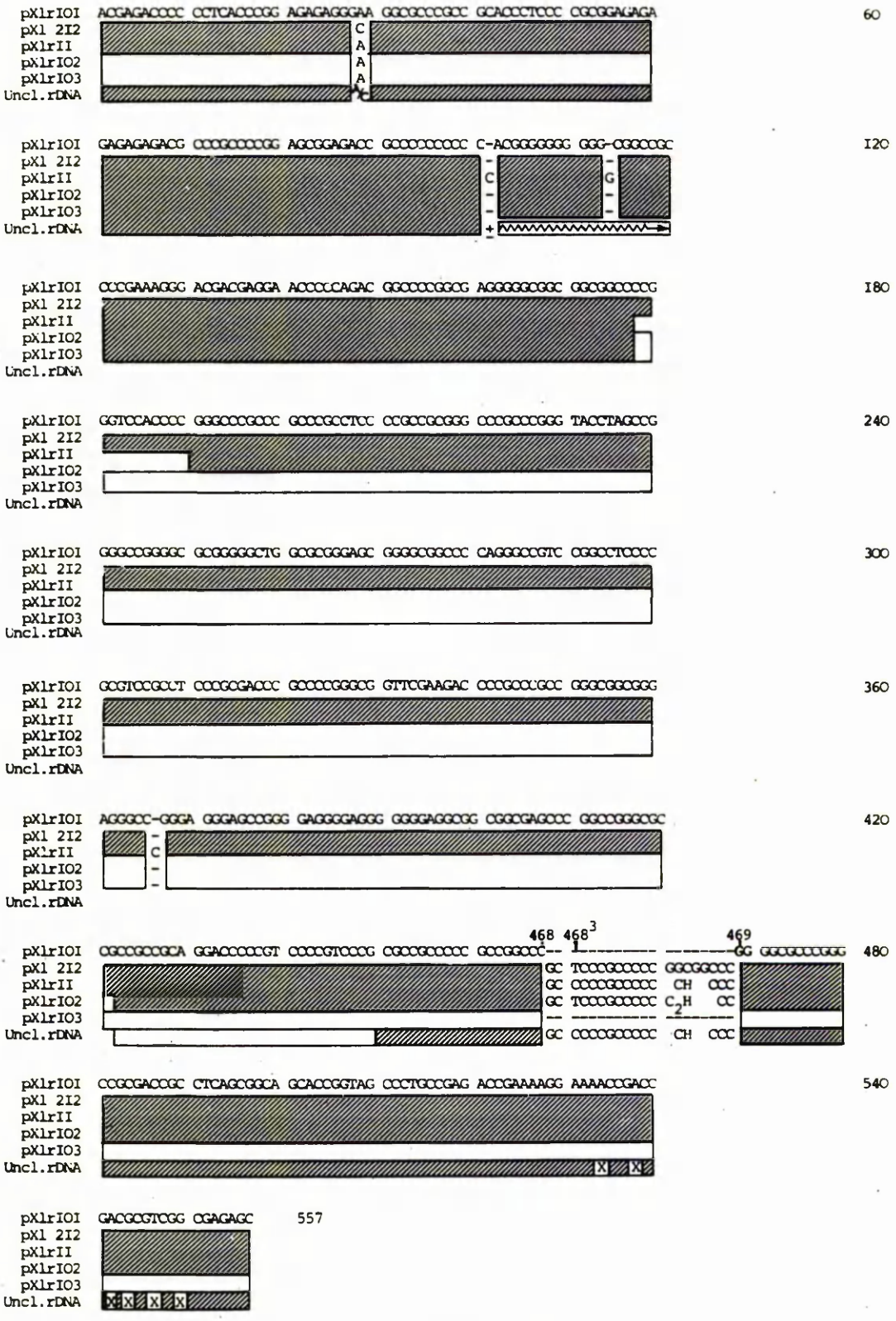

pXlrIO1
pXl 2I2
pXlrII
pXlrIO2
pXlrIO3
Uncl.rDNA
60
120
180
240
300
360
420
480
540
557
468
468³
469

Figure 5.4

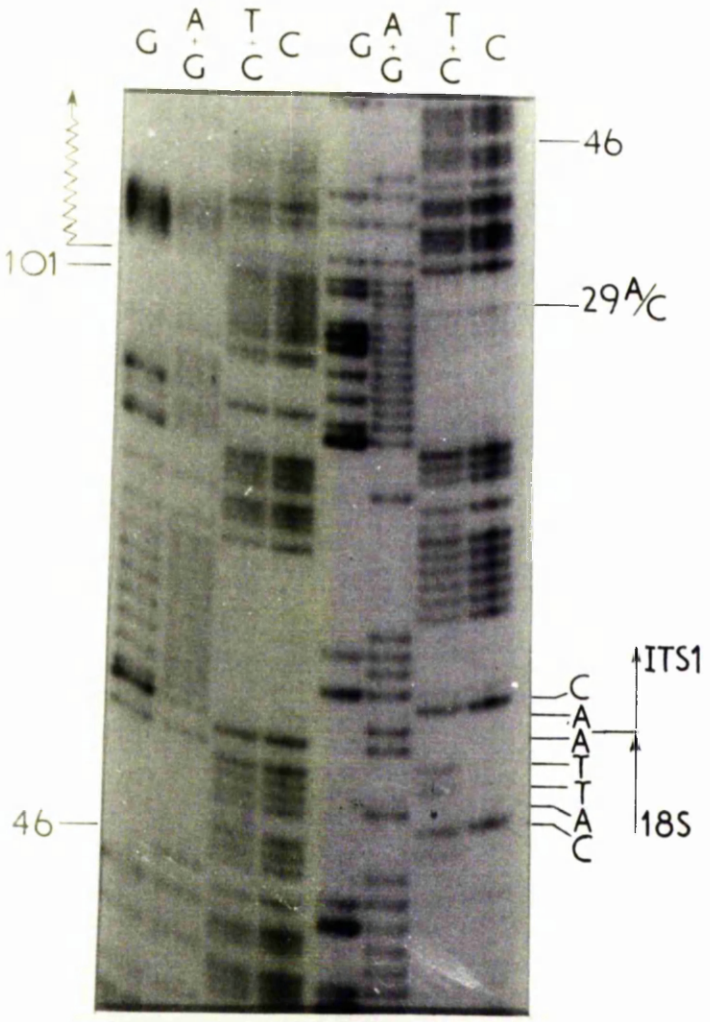


Figure 5.5. Sequencing run on uncloned amplified rDNA.

The figure shows part of a sequencing gel obtained from a complete sequencing run on uncloned rDNA. The 3'end of the 18S gene and the first part of ITS1 are illustrated. An aliquot of fragment 5 from uncloned rDNA was 5'end labelled and secondary restricted with TaqI. The labelled *XbaI/TaqI fragment was then subjected to full sequencing chemistry. Three loadings were applied to the gel; the two longer loadings are shown here. Nucleotides are numbered according to the pXlr101 numbering system. Despite the small amount of material the sequencing gel obtained yields good sequence information: at position 29, the dual A/C reaction corresponds to the A/C heterogeneity found at this position in the clones. The zig-zag line following nucleotide 101 indicates loss of readability due to length heterogeneity in this region of the clones.

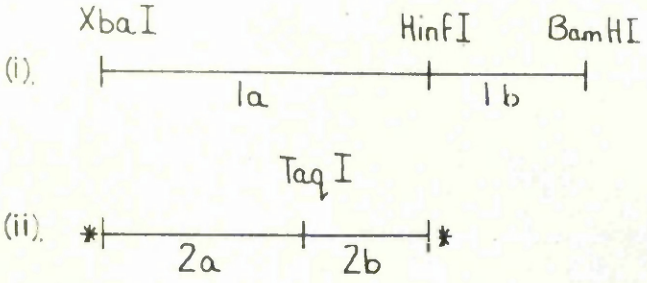
A/C variation between clones. Further downstream at nucleotide 100 the sequence was readable in the oligo-C tract but became unreadable after the compressed G tract (Figure 5.5). This was in marked contrast to the clones. The same feature was confirmed in the sequencing gel starting from the Sau 3A site just before the 3' end of the 18S gene (Figure 5.3). This sudden loss of readability can be attributed to length heterogeneity in the oligo-C/oligo-G tract corresponding to length variants at this location found in clones. Thus the sites of variation between clones correspond to sites of heterogeneity in uncloned rDNA and are therefore not artefacts of cloning.

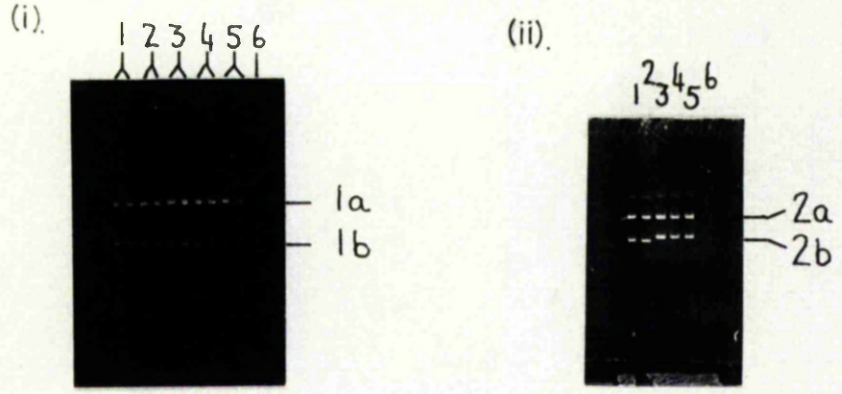
The most exciting site of variation in ITS1 occurs immediately following nucleotide 468 in the reference sequence (Figure 5.4). A G cleavage experiment was carried out on the five clones and on an aliquot of uncloned amplified rDNA band 5, extending leftwards from the Hinf I site at the start of the 5.8S gene (Figure 5.6). Full sequencing of pX1212 from this site had revealed an extra block of 20 nucleotides at nucleotide 468. Comparison of the G cleavage patterns revealed that two of the five clones showed the same pattern as pX1212. Thus pX1212, pX1r102 and pX1r11 contain the extra block of 20 nucleotides not found in pX1r101 and pX1r103. Complete sequencing runs also showed variation within the 20bp segment at position 468^3, two clones possess T and a third possesses C (Figure 5.4). The preparation of uncloned rDNA contains predominantly the sequence with the extra 20 nucleotides (Figure 5.6). In addition, a complete sequencing run showed a relatively homogeneous sequence in the region con-

Figure 5.6 Preparation and screening of the 3' region of ITS1.

The XbaI/BamHI ITS1-containing fragment of five clones and an aliquot of uncloned rDNA, band 5, was digested with Hinf I and the products resolved on a 4% polyacrylamide gel (i). Lanes 1-6 represent pX1r101, 103, 102, pX1r11, pX1212 and uncloned rDNA respectively. Fragment 1a contains the entire ITS1 and was eluted from the gel for each of the clones and uncloned rDNA. After 5' end labelling and secondary restriction with Taq I the products were separated on a second 4% polyacrylamide gel (ii) (Lanes 1-6 as before). Fragment 2b contains the 3' region of ITS1 and migrates more slowly in pX1r102, pX1r11 and pX1212 than the equivalent band in pX1r101 and 103. Uncloned rDNA fragment 2b was very faint. G cleavage was performed on fragment 2b from each of the rDNA sources extending leftwards from the Hinf I site at the start of the 5.8S gene. The relevant areas from two overlapping loadings from the same sequencing gel are shown (iii). Guanine bands correspond to cytosine on the rightwards strand in Figure 5.4. Nucleotides are numbered according to the pX1r101 numbering system (Figure 4.8).

"X" in uncloned rDNA is as previously described (Figure 5.4) and in the text (5.3.1).


XbaI
HinfI
BamHI
(i).
1a
1b
TaqI
(ii).
2a
2b


(i).
1 2 3 4 5 6
1a
1b
(ii).
1 2 3 4 5 6
2a
2b

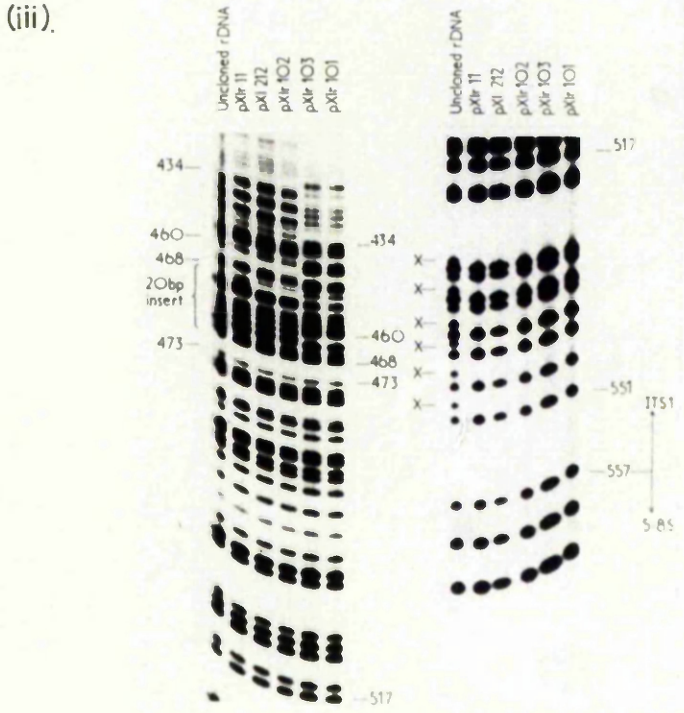

(iii).
Unclonedr DNA
pXlr 11
pXl 212
pXlr 102
pXlr 103
pXlr 101
434
460
468
20bp insert
473
517
551
557
ITS1
5.8S

Figure 5.6

Figure 5.7 Analysis of uncloned rDNA in the 3' region of ITS1.

(a) Complete sequencing run on uncloned rDNA extending leftwards from the Hinf I site at the start of 5.8S into ITS1. Three loadings were applied to the sequencing gel : the two longer loadings are shown. Nucleotides are numbered according to the pX1r101 system in Figure 4.8. It was not possible to determine the precise sequence of the block of 20 nucleotides at position 468 due to compression of cytosines (denoted by D) on the leftwards strand.

(b) This illustrates the two size classes of rDNA due to the variable presence of an extra 20bp segment following nucleotide 468 in ITS1. The righthand side of (b) shows fragments labelled at the Hinf I site in 5.8S and extending to the Taq I site in ITS1 (fragment 2b, Figure 5.6). Uncloned rDNA contains predominantly the larger fragment with a small quantity (arrowed) of the smaller fragment. The lefthand side of (b) shows the same fragments after partial G cleavage (as in Figure 5.6) to provide size calibration. pX1r11 contains an extra nucleotide following position 366 and this seems to cause slight expansion of spacing in the last few nucleotides of the fragment from this clone.

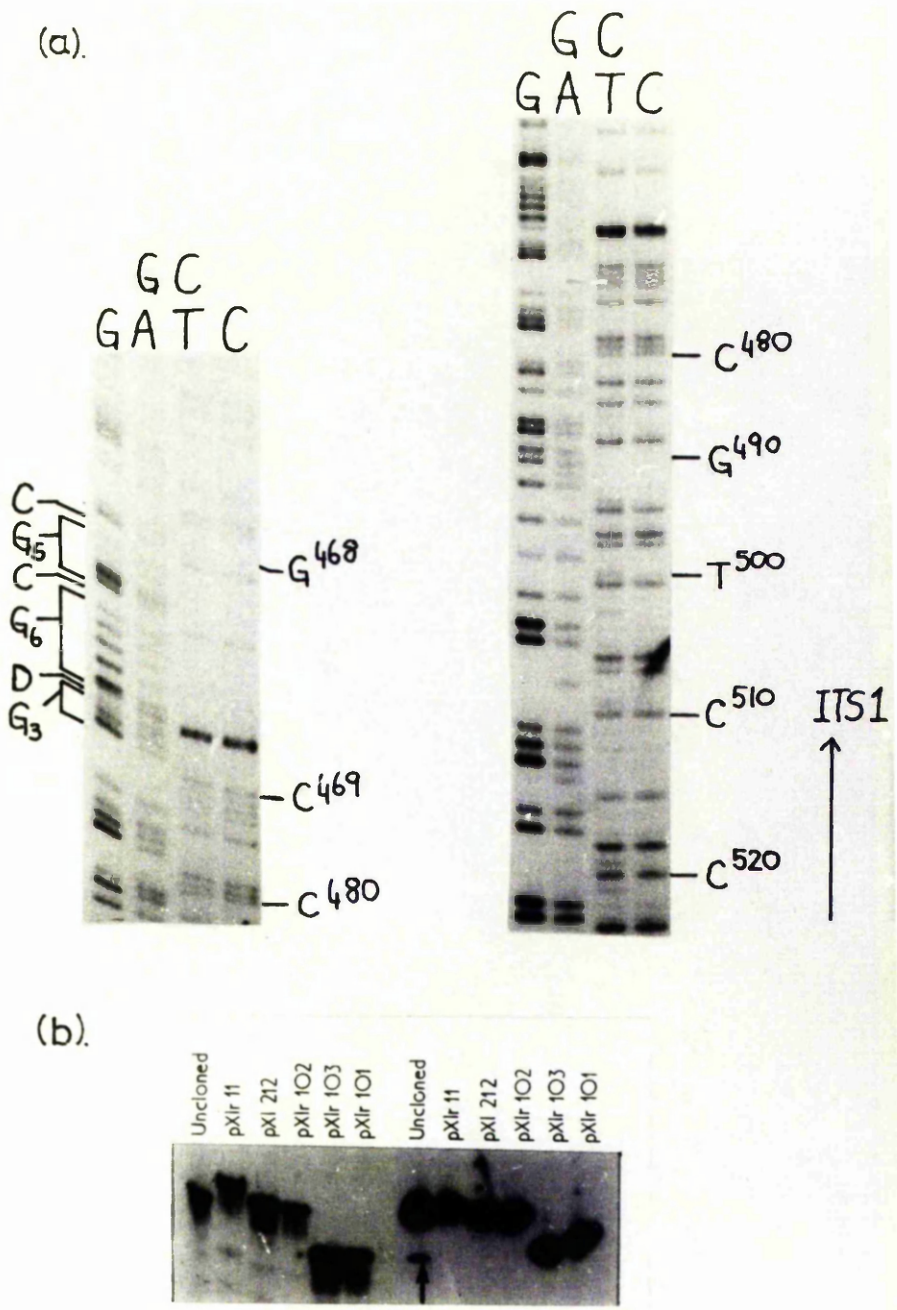

(a).
G C
G A T C
G C
G A T C
C
G5
C
G6
D
G3
G468
C469
C480
C480
G490
T500
C510
C520
ITS1
(b).
Uncloned
pXlr 11
pXl 212
pXlr 102
pXlr 103
pXlr 101
Uncloned
pXlr 11
pXl 212
pXlr 102
pXlr 103
pXlr 101

Figure 5.7

taining the extra 20 nucleotides (Figure 5.7(a)). A sizing experiment with an aliquot of the labelled Hinf I restriction fragment encompassing this region showed that only 2-5% of the material lacked this 20bp segment (Figure 5.7(b)). Therefore, although this preparation of uncloned rDNA contains predominantly the longer sequence version at this point, a small but detectable quantity of shorter version is also present.

The last region of variability in ITS1 was found only in uncloned rDNA and occurred in the region flanking the 5.8S sequence. In Figure 5.6, X marks the six sites at which G shows in uncloned rDNA but not in any of the clones. The complete sequencing run on uncloned rDNA showed heterogeneity at each of these sites. A possible explanation for the complex pattern of heterogeneity might be the presence, in uncloned rDNA, of a sequence variant that is "frame-shifted" one position to the left in the sequence of the RNA-like strand. A single base deletion immediately preceding the affected region and a single base insertion immediately afterwards would produce this effect. Subsequent analysis of further clones (see under chromosomal clones, Chapter 6) revealed this to be the case.

5.4.2 ITS2

The single Bam HI site near the middle of ITS2 (Figure 5.3) serves as a useful dividing point for describing ITS2 sequence variants. The initial sequence comparison between the reference clone and pX1212 revealed a homogeneous

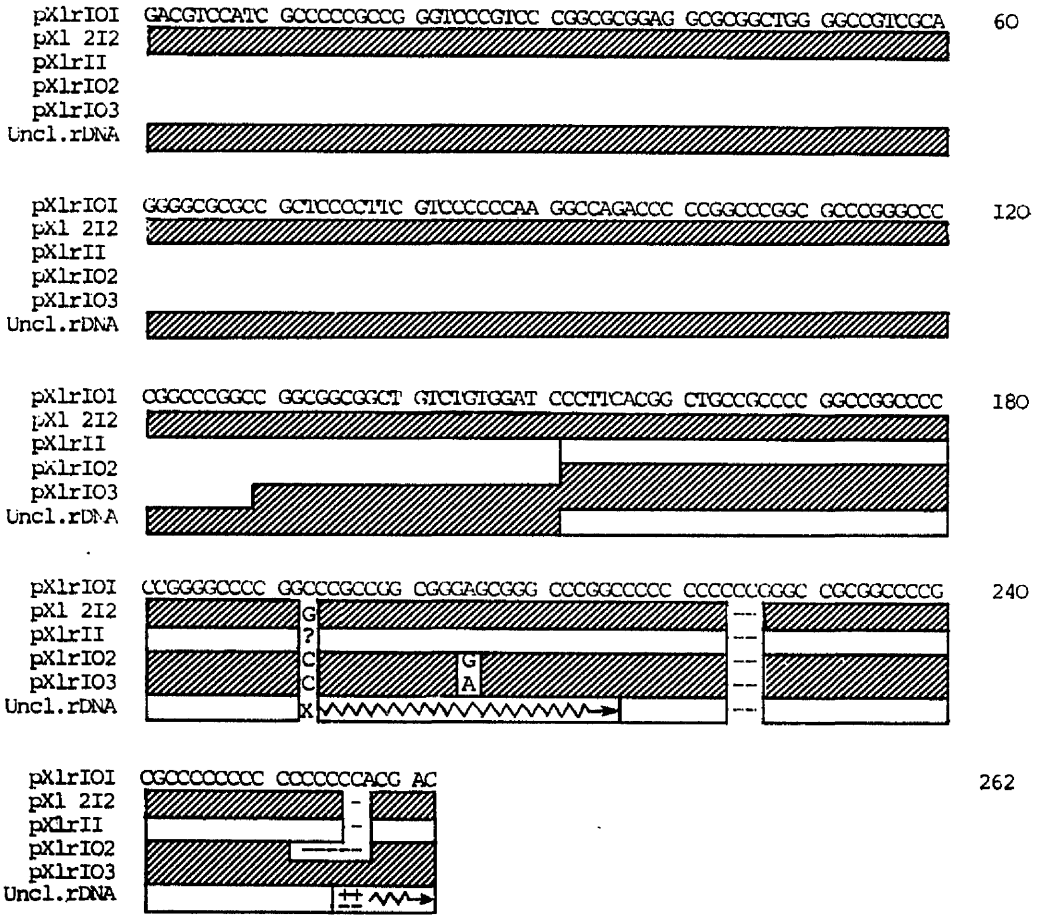


Figure 5.8. Summary of comparative data for ITS 2 (rightwards, RNA-like strand).

The figure shows the complete sequence in pXlrlOl and the variation between rDNA sources. "Boxing" is as explained in figure 5.4 with one exception; an open box signifies that C cleavage runs were carried out, with no difference evident from pXlrlOl in the boxed region. The question mark at position 193 in pXlrll means that there was no C in the C cleavage pattern. Therefore G was inferred by analogy with the full sequence determination for pXlrlOl.

sequence through to this Bam HI site (Figure 5.8). A full sequencing run on uncloned amplified rDNA (band 5), extending leftwards from the Bam HI site (Figure 5.3) showed no detectable heterogeneities in this region.

The latter half of ITS2 harbours all the ITS2 sequence variants. A C cleavage experiment was performed on the five clones and uncloned rDNA (band 4, Figure 5.2) extending rightwards from the Bam HI site. In the cloned DNAs, two sites of base substitution at positions 193 and 205 were found (Figure 5.8). Uncloned rDNA in this region showed a complicated pattern for about 20 nucleotides following the first site of sequence variation between the clones but the pattern became normal again at the start of the first oligo-C tract (Figure 5.9, lefthand section). Evidently variant sequences are present in this region but it is not possible to say whether the variants are the same as those identified in the clones.

In pX1r101, the first oligo-C tract contains 15 cytosines. In the other clones and uncloned rDNA that tract contains only 11 cytosines (Figure 5.9, lefthand section). In contrast, the number of cytosines in the second tract is more generally variable, ranging from 9-15 C in the clones (Figure 5.9, righthand section). It is not possible to distinguish the number of cytosines in uncloned rDNA. The C cleavage pattern after the tract becomes unclear as a result of length heterogeneity in this tract (Figure 5.9, righthand section).

Figure 5.9 Comparative C cleavage of the 3' region of ITS2 in amplified rDNA.

The Figure shows the C cleavage patterns of cloned and uncloned rDNA in the region of ITS2 immediately following the first site of variation at position 193. Nucleotides are numbered according to the individual numbering system of pX1r101 (Figure 4.8).

Following the first site of variation of position 193, uncloned rDNA shows a complicated C cleavage pattern for about 20 nucleotides ("X" in left-hand section of Figure) : tightly spaced doublets of C bands occur in which the separation is approximately half that for the normal interval between single nucleotides. At position 216 the spacing reverts to normal.

The second oligo-C tract is shown in the right-hand section. It varies from 9-15 in cloned rDNAs. The zig-zag line for uncloned rDNA indicates partly unresolved cleavage patterns due to length heterogeneity corresponding to length variation at this point in the clones.

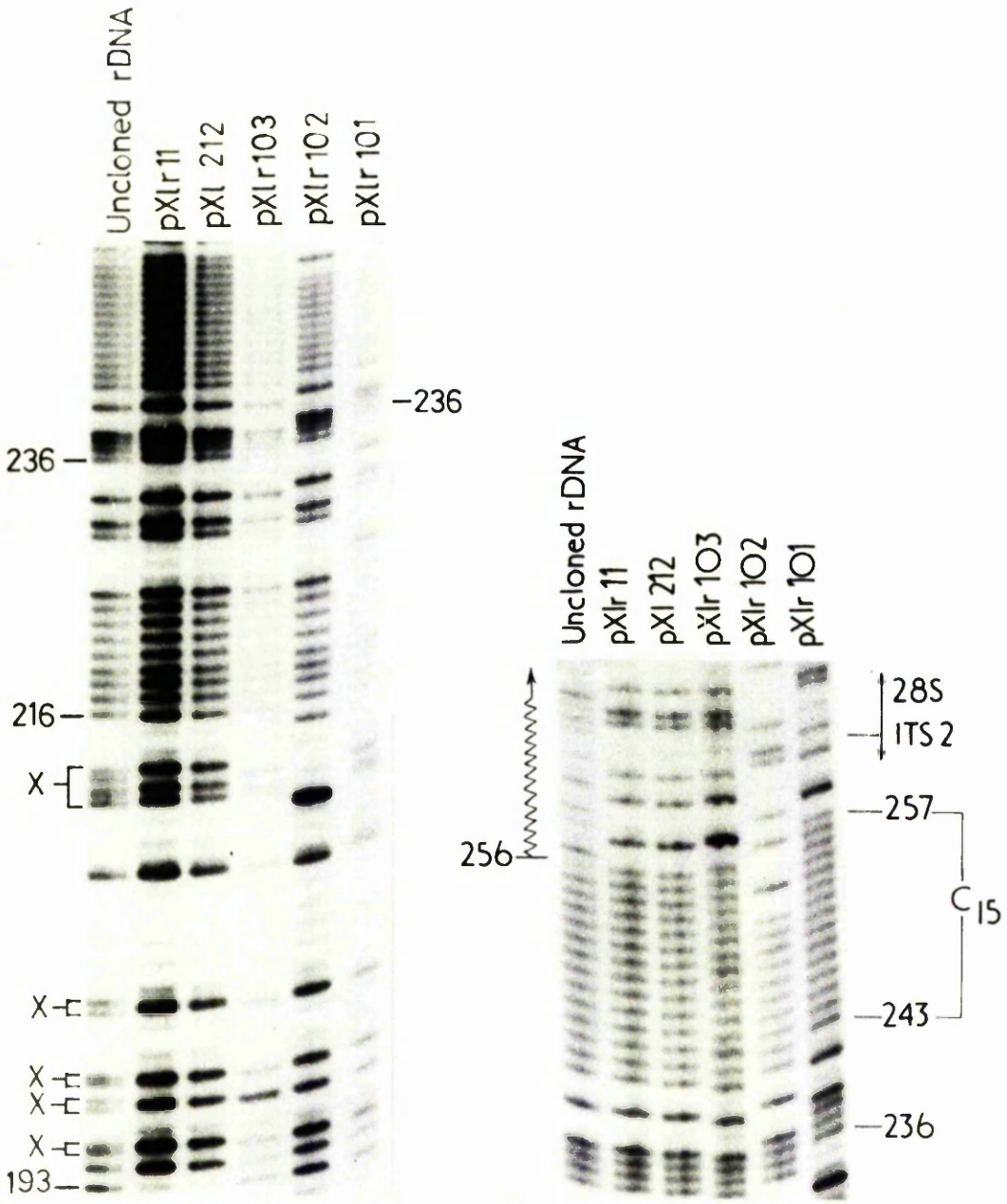

Uncloned rDNA
pXlr11
pXl 212
pXlr103
pXlr102
pXlr101
236
216
X
193
−236
Uncloned rDNA
pXlr 11
pXl 212
pXlr 103
pXlr 102
pXlr 101
256
28S
ITS 2
257
15
243
236

Figure 5.9

5.4.3 ETS

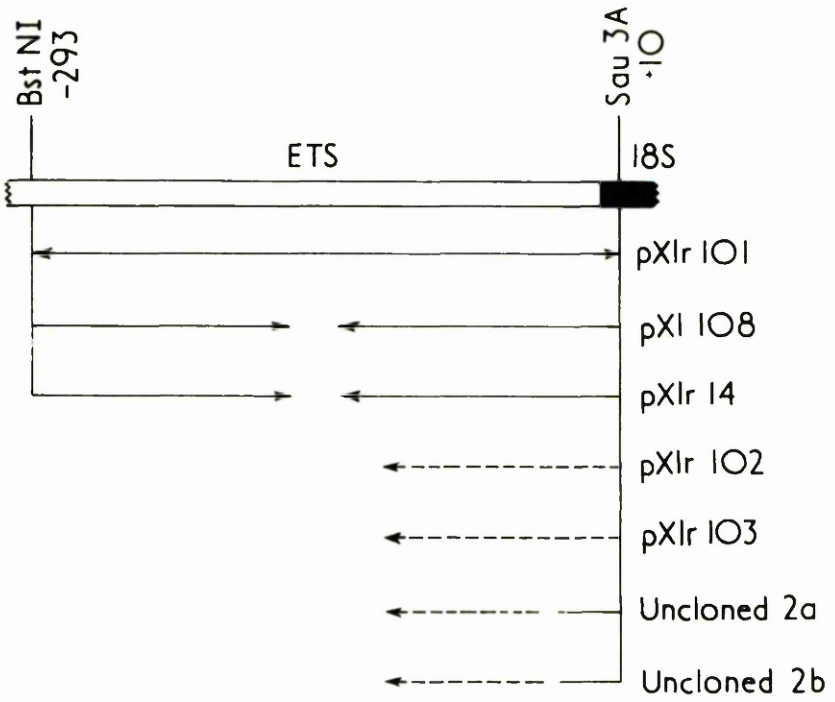
This is the longest of the three transcribed spacers in X. laevis rDNA and has been fully sequenced in the reference clone pX1r101 (Maden et al., 1982a). In preliminary comparative studies the region near the 18S gene revealed some sites of variation between clones (B.E.H. Maden, pers. comm.). This region was therefore chosen for a more detailed study.

Five ETS-containing clones, pX1r101, 102, 103, pX1108 and pX1r14, and fragments 2a and 2b from uncloned amplified rDNA were used for this comparative study. Fragments 2a and 2b each contain part of the NTS, the entire ETS and part of the 18S gene (Figure 5.2). The two fragments differ in length from each other by about 150 nucleotides. The main source of this length difference is in the NTS. The reasons for inferring this are as follows. Restriction data on the NTS in rDNA clones show a general pattern in which two or more Bam HI sites are flanked on the right by variable length repetitious regions containing multiple Alu I restriction sites (see Introduction). After the last Alu I repetitious region in the NTS there are no further Alu I sites before the first Xba I site in the 18S gene shown in Figure 5.2. (Boseley et al., 1979; Maden et al., 1982(b)). Restriction tests on small aliquots of fragments 2a and 2b showed that the distance from the last Alu I site in the NTS to the Xba I site was similar to within a few nucleotides in the two fragments (B.E.H. Maden, pers. comm.). This indicates that the major site of length difference must be, as expected, within the Alu I repetitious region immediately following the last Bam HI site in the NTS.

Figure 5.10 Analysis of the ETS in amplified rDNA.

Upper section: Diagram of sequencing runs in the ETS. pXlr101 was subjected to complete sequence analysis throughout the ETS (Maden et al, 1982a). pX1108 and pXlr14 were subjected to full sequencing runs (continuous arrows) from the restriction sites shown. pXlr102 and pXlr103 were subjected to G cleavage runs (interrupted arrows) from the Sau 3A site indicated. Uncloned rDNA fragments 2a and 2b were subjected to complete sequencing runs from the Xba 1 site at nucleotide +158 in the 18S gene, permitting reading into the first 25 nucleotides of the ETS (Maden et al, 1982b), and to G cleavage runs (interrupted arrows) from the indicated Sau 3A site.

Lower section: G cleavage patterns of fragments from rDNA clones and uncloned rDNA extending leftwards from the Sau 3A site at nucleotide +10 in the 18S gene. The relevant areas of two overlapping loadings from the sequencing gel are shown. Guanine bands in this figure correspond to cytosine on the rightwards strand in Figure 5.11. Nucleotides are numbered according to the pXlr101 numbering system for the ETS (Maden et al, 1982a). The pXlr101-type sequence gave some spacing anomalies between nucleotides -20 and -33 on the leftwards strand. The full sequence was established for both strands in pXlr101. It was thereby possible to identify the pXlr101-type sequence unambiguously in other sources from data on the leftwards strand.


Bst NI
-293
Sau 3A
+10
ETS
18S
pXlr 101
pXl 108
pXlr 14
pXlr 102
pXlr 103
Uncloned 2a
Uncloned 2b

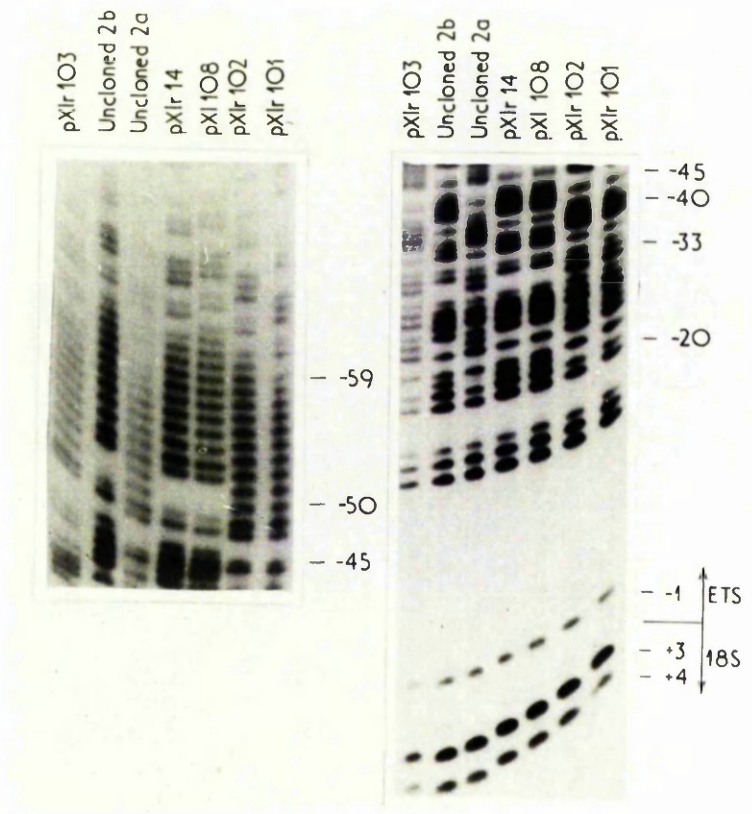

pXlr 103
Uncloned 2b
Uncloned 2a
pXlr 14
pXl 108
pXlr 102
pXlr 101
-59
-50
-45
-40
-33
-20
-1
+3
+4
ETS
18S

Figure 5.10

A Sau 3A restriction fragment containing the bulk of the ETS and the first 10 nucleotides of the 18S coding region was purified from the five ETS-clones (see above) and uncloned rDNA fragments 2a and 2b. After 5'-end labelling and secondary restriction, G cleavage was carried out, extending leftwards from the 18S Sau 3A site into the ETS. In addition, full sequencing runs were performed on two of the clones, pX1108 and pX1r14, in this region (Figure 5.10).

Comparison of the G cleavage patterns obtained (Figure 5.10) showed several interesting features. In the region between nucleotides -20 and -33, the clones fall into two groups with different cleavage patterns: pX1r101, 102 and 103 show one pattern whereas pX1r14 and pX1108 show another pattern. Full sequencing runs showed that pX1r14 and pX1108 differ from pX1r101 by the presence of two mini-insertions, each comprising two nucleotides, and a base substitution (Figure 5.11). At nucleotide -33, the two types of pattern become similar again but are out of register by four nucleotides due to the extra nucleotides in the pX1r14-type sequence. Uncloned fragments 2a and 2b show a very interesting relationship to the patterns for the cloned rDNA in this region. The longer fragment 2a contains predominantly the pX1r101-type sequence whereas the shorter fragment 2b, contains predominantly the pX1r14-type sequence. Had these sequence types been randomly distributed among fragments 2a and 2b, then the G cleavage patterns of both fragments would have become obscure immediately upstream due to superposition of "out-of-register" sequences. This is clearly not the case (Figure 5.10).

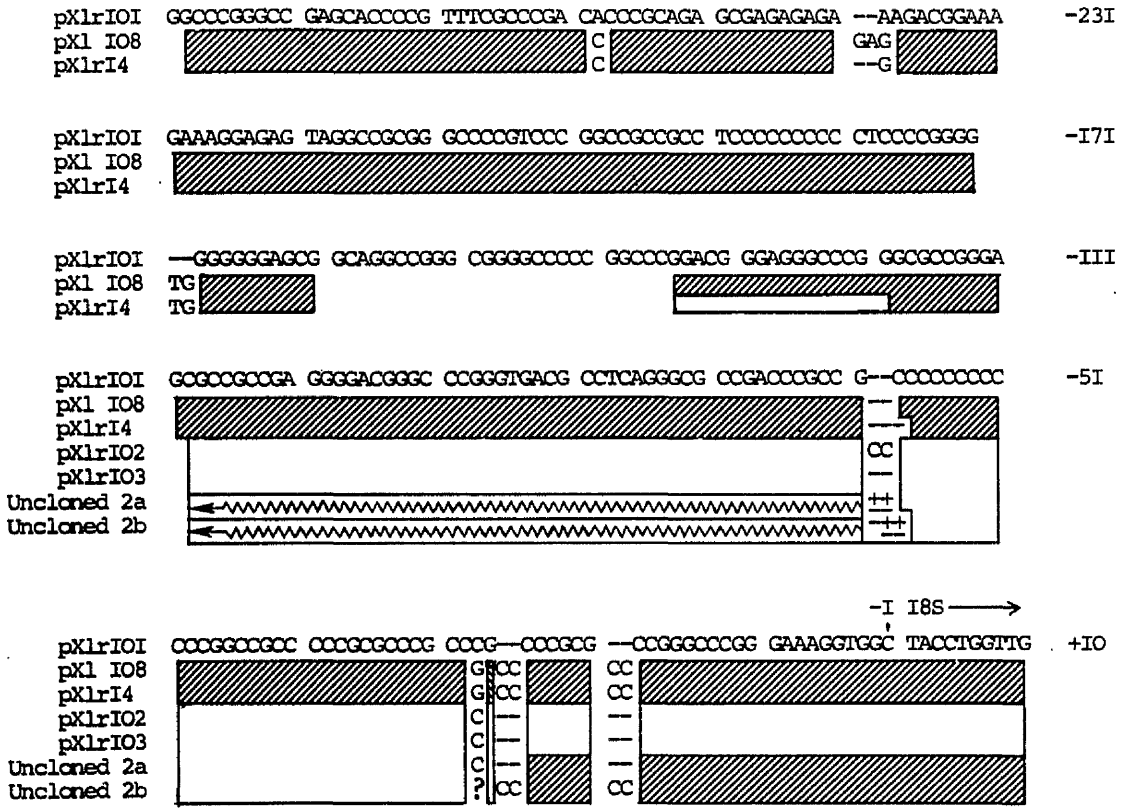


Figure 5.11. Summary of comparative data in the ETS (rightwards, RNA-like strand).

The figure shows the complete sequence in pXlrlOl and the variations between rDNA sources. "Boxing" is as explained in figure 5.4. The numbering system is according to Maden et al, (1982a). The question mark at position -28 in uncloned fragment 2b means that there was no G in the leftwards strand at this point (figure 5.10). Thus the actual nucleotide cannot be specified with certainty. However, by analogy with sequence data from pXllO8 and pXlrl4 a cytosine is almost certainly present, and hence guanine in the rightwards strand.

Continuing upstream the cleavage patterns of both cloned and uncloned rDNA remain distinct until a long run of guanines (Figure 5.10, lefthand section of G cleavage gel) corresponding to the oligo-C tract at nucleotide -50 in the rightwards strands (Figure 5.11). As is the case for ITS2, the homopolymeric tract exhibits length variation of 11-14 residues in the different clones. In uncloned rDNA, the G cleavage pattern beyond this region shows loss of internal register. Although a dominant pattern can still be recognised, there are contributions from "out-of-phase" patterns indicating that length heterogeneity occurs at this point in both fragments. This is in correspondence with the length variants at this site in the clones.

Analysis of three clones in the region extending rightwards from nucleotide -293 performed by B.E.H. Maden, revealed some additional sites of variation as shown in Figure 5.11.

5.4.4 rRNA Coding Regions

As mentioned previously (4.4.1) no sites of variation were found in the rRNA coding regions studied in the reference clone and pX1212. Maden *et al* (1982b) reported that no sites of sequence variation were found in an extensive survey of the 18S gene from several cloned rDNAs and uncloned amplified rDNA. I have sequenced the 5.8S coding region from this uncloned amplified rDNA preparation. Two overlapping sequencing gels through the 5.8S gene were obtained (Figure 5.3).

The first gel, extending rightwards from the Hinf I site, yielded poor A and T reactions and prevented elucidation of the precise nucleotide sequence. The second gel, extending from the Sau 3A site illustrated in Figure 5.3, yielded good chemistry and the data from the two gels together excluded the possibility of any gross heterogeneities, such as length heterogeneity, in the 5.8S gene.

During analysis of two short regions of 28S rDNA, one site of variation was found between clones; pX1r102 contains an extra cytosine approximately 40 nucleotides in from the start of the sequence. Because of length heterogeneity in the oligo-C tract, upstream in ITS2, I was unable to determine the sequence in this region of the 28S gene in uncloned rDNA.

Comparative G cleavage runs, leftwards from the Bam HI site in the 28S gene (Figure 5.2) revealed an identical pattern in both cloned and uncloned rDNA in this region.

5.5 Conclusions

I started this chapter by posing three questions. The experiments just described have answered the first two questions. Sequence microheterogeneities are a common phenomenon among the transcribed spacers of *X. laevis* rDNA. In addition, analysis of uncloned rDNA has shown that these heterogeneities are not artefacts introduced during cloning but are representative of sequence variants occurring in a large rDNA population.

However, one question remains unanswered. Do the sequence variants found in amplified rDNA also occur in chromosomal rDNA ? It is important to answer this question because the possibility remains that sequence variation may arise in amplified rDNA because of aberrant DNA replication of chromosomal rDNA during amplification. By the same token, identification of chromosomal rDNA sequence variants would indicate that the primary reservoir of variant sequences is genomic rDNA.

The following chapter describes the construction and subsequent analysis of five X. laevis chromosomal rDNA clones.

6. Analysis of X. laevis Chromosomal rDNA Clones

6.1 Initial characterisation of chromosomal rDNA

3µg of purified X. laevis chromosomal rDNA were donated by A.P. Bird. The rDNA was isolated from the erythrocytes of several frogs and was purified by banding in caesium chloride. The frogs had been purchased from Xenopus Ltd. and originated from the wild.

In order to check the purity of the rDNA, small aliquots (0.1µg) were restricted with Eco RI or Hind III. The expected bands were obtained. Another small aliquot was shown to be highly resistant to digestion with Hpa II (Figure 6.1). This resistance is characteristic of erythrocyte chromosomal rDNA in contrast to either amplified or cloned rDNA, and is due to methylation of the many Hpa II sites (Bird and Southern, 1978).

6.2 Construction and identification of chromosomal rDNA clones

My objective was to perform sequence analysis in the transcribed spacers of chromosomal rDNA. In order to achieve this aim, I decided to construct chromosomal rDNA clones. My previous results made me confident that cloning would not introduce any sequence artefacts. I wished to construct clones homologous in configuration to the amplified rDNA clones pXlr101-103 (Figure 5.1). For this reason therefore, chromosomal rDNA (0.8µg) was digested to completion with Hind III and the restriction products (containing complete repeat units) cloned into the Hind III site of pAT153. To aid in the formation of recombinant molecules, Hind III cut pAT153

Figure 6.1 Characterisation of uncloned chromosomal rDNA.

The sequence CpG is contained within the recognition site for the restriction enzyme Hpa II. Hpa II cleaves only at unmethylated CpG sequences.

Small aliquots of uncloned chromosomal rDNA (methylated) and amplified cloned rDNA (unmethylated) were digested to completion with Hpa II or Hae III and the products resolved on a 1% horizontal agarose gel. The products were transferred to nitrocellulose paper according to the method of Southern (1979) and hybridised to a HeLa ^{32}P-28S + 18S rRNA probe. The figure shows the resulting autoradiograph.

Lane 1 : 2µg pXlrl01 digested with Hpa II

Lane 2 : 0.75µg chromosomal rDNA digested with Hpa II

Lane 3 : 2µg pXlrl01 digested with Hae III

Lane 4 : 0.25µg chromosomal rDNA digested with Hae III.

Chromosomal rDNA gave only high molecular weight products after digestion with Hpa II due to methylation of CpG sites. Digestion of chromosomal rDNA with Hae III gave the characteristic Hae III-rDNA restriction pattern (compare lanes 3 and 4) and confirmed the susceptibility of this rDNA preparation to restriction.

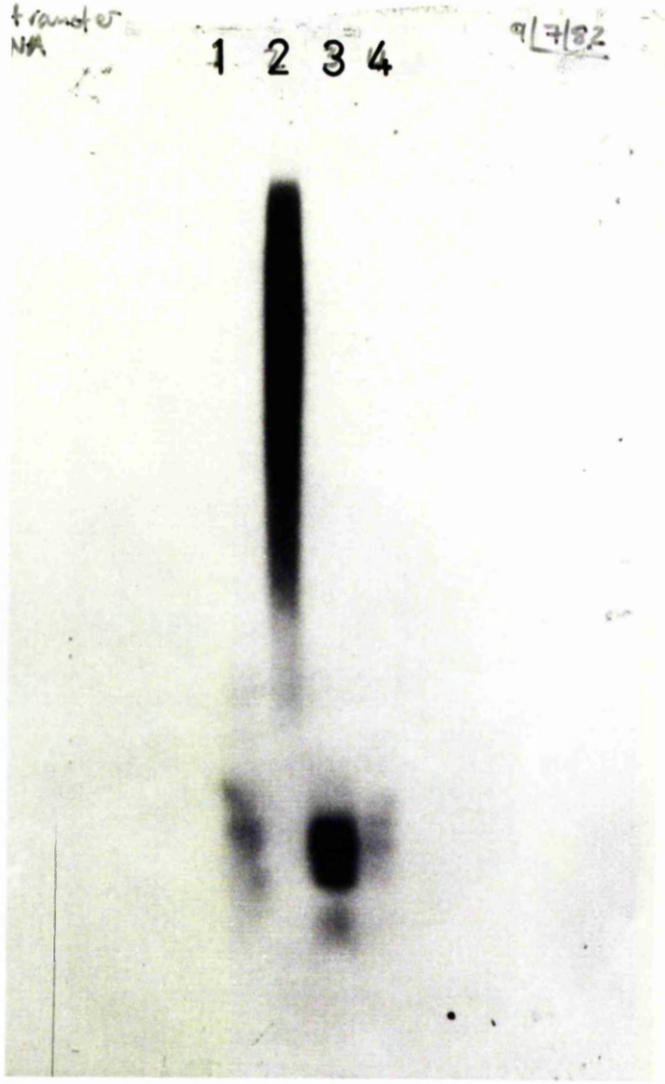


Figure 6.1

was treated with bacterial alkaline phosphatase to prevent recircularisation of the vector DNA. Ligation was performed as described (Chapter 2, Methods 2.1) and recombinant molecules were initially selected on the basis of antibiotic resistance. The single Hind III site in pAT153 is located in the promoter region of the tetracycline coding sequence and cloning into this site prevents expression of the tetracycline gene. Recombinant molecules are therefore Amp^R Tet^S. Of the 108 colonies screened, 35 were Amp^R Tet^S and were designated as putative recombinants.

In order to identify rDNA containing clones, both colony hybridisation and small scale plasmid DNA preparations were performed on the 35 putative recombinants. Examination of the autoradiograph after colony hybridisation revealed five strong positive signals with the ^{32}P-labelled HeLa 28S rRNA probe used for the hybridisation. The positive colonies were identified as colonies 4, 7, 13, 28 and 31. Restriction analysis with Bam HI on the DNA prepared from the 35 colonies (Figure 6.2) revealed that the DNA from the same five colonies gave bands characteristic of rDNA digested with Bam HI. Large scale DNA preparations were then made from these five colonies and the plasmids denoted pXlcrl, 2, 3, 4 and 5 where "Xlcr" strands for "X. laevis chromosomal ribosomal".

Thus of the 35 Amp^R Tet^S colonies identified initially, only five contained recombinant rDNA molecules. I think that this low efficiency of recombinant formation is due largely to the large size of the insert DNA (11kb) in relation to

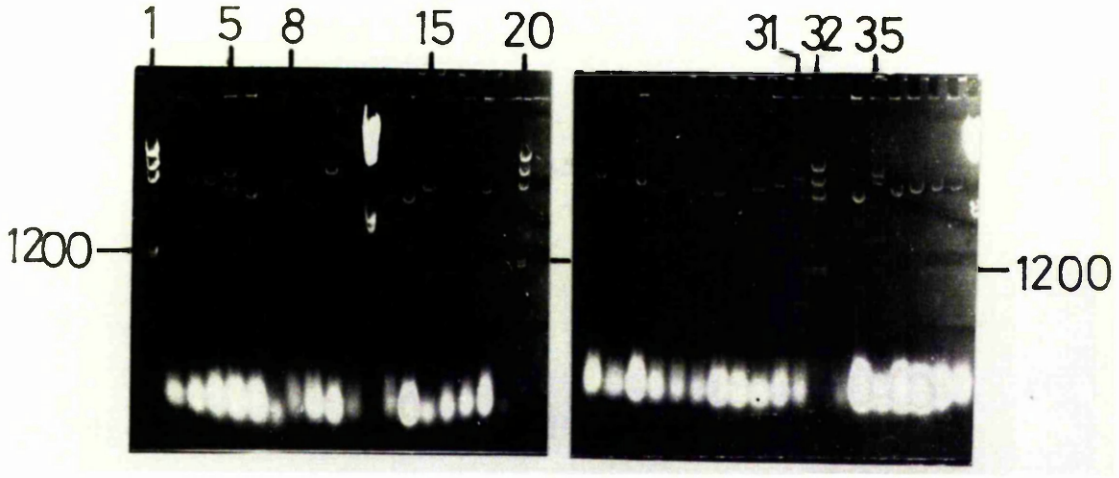


Figure 6.2 Identification of chromosomal rDNA clones by limited restriction analysis.

DNA prepared from the 35 putative recombinants (see text) was digested with Bam HI and the products resolved on 1% vertical agarose gels. Identification of rDNA clones was aided by comparison of the Bam HI restriction products with those obtained for a known rDNA clone - pX1r101 (lanes 1, 20 and 32). An invariant 1200bp fragment containing the 3'end of ITS 2 and the start of the 28S gene (seeabove) was identified in the DNA purified from colonies 4, 7, 13, 28 and 31 (lanes 5, 8, 15, 31 and 35 respectively).
The strong bands at the bottom of both gels represents low molecular weight RNA present in the small scale preparation of each DNA.

vector DNA size (3kb). Subsequent control experiments carried out in our laboratory revealed a dramatic fall in the number of recombinant molecules obtained as insert size was increased to over 6kb. With hindsight, a higher cloning efficiency would surely have been obtained if I had used the bacteriophage lamba cloning system for constructing full length chromosomal rDNA clones.

6.3 Restriction and sequence analysis of pX1cr1-5

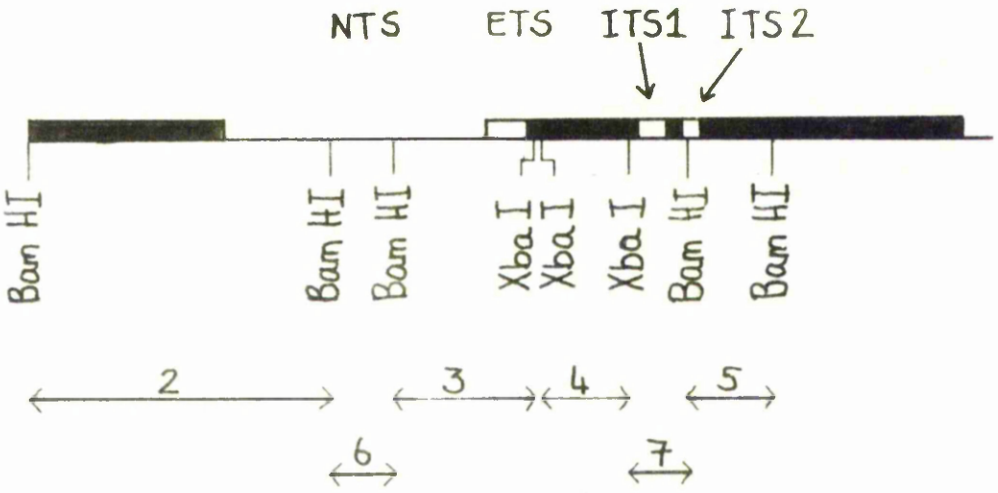
Digestion of pX1cr1-5 with Bam HI and Xba I yields a series of fragments similar to those obtained for amplified rDNA (Figure 5.2). Figure 6.3 shows the restriction products of two chromosomal clones, pX1cr 4 and pX1cr 5, and pX1r101 after combined digestion with Bam HI and Xba I. Certain fragments are common to both pX1r101 and pX1cr 4 and 5. Bands 4, 7 and 5 in pX1r101 correspond to fragments containing part of the 18S gene, 18S-28S intergene region and 28S gene, respectively. Similar bands were found in all five chromosomal clones and the equivalent bands in pX1cr1-5 were subsequently used for sequence analysis of ITS1 and ITS2. Because of heterogeneity in the location of Bam HI sites in the NTS, identification of the ETS-containing fragment in the chromosomal clones was more difficult. A different restriction strategy was devised, using Sau 3A, to obtain an appropriate ETS fragment.

Analysis of all the known Sau 3A sites in pX1r101 led to the prediction that the entire length of ITS1 and three-quarters of the ETS should be contained in separate Sau 3A

Figure 6.3 Restriction analysis of pXlcr4 and pXlcr5

Upper section : Map of Bam HI and Xba I sites in the reference clone pXlrl01. Lines with arrowheads represent fragments obtained from complete digestion of pXlrl01 with Bam HI and Xba I. Fragments 2-7 are numbered according to size.

Lower section : Bam HI/Xba I restriction products of two chromosomal clones pXlcr4 and 5 (lanes 2 and 3 respectively) and pXlrl01 (lanes 1 and 4) resolved on a 1% vertical agarose gel and visualised, after ethidium bromide staining, by U.V. illumination. The top two fragments in lanes 1, 3 and 4 and the top fragment in lane 2 contain some material from the parent vectors linked to rDNA. The fragment numbers for pXlrl01 correspond to those in the upper section. Fragments obtained within the ribosomal transcription unit were the same in pXlcr4, 5 and pXlrl01 (i.e. fragments 4, 5,7). The remaining bands in pXlcr4 and 5 correspond to NTS-containing fragments and are variable because of Bam HI site heterogeneity in the NTS.


NTS
ETS
ITS1
ITS2
Bam HI
Bam HI
Bam HI
Xba I
Xba I
Xba I
Bam HI
Bam HI
2
3
4
5
6
7

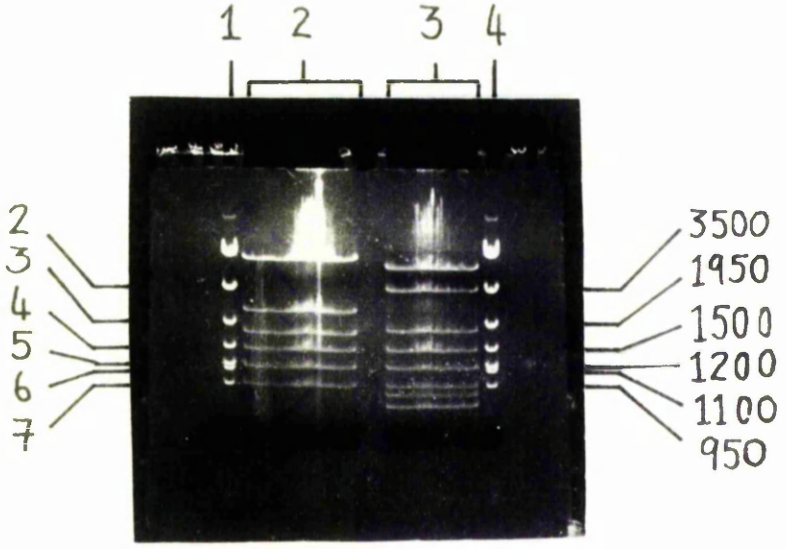

1
2
3
4
2
3
4
5
6
7
3500
1950
1500
1200
1100
950

Figure 6.3

fragments approximately 600bp in size. The ITS1 fragment extends from a Sau 3A site at the 3' end of the 18S gene (Figure 5.3) to a Sau 3A site 20 nucleotides into the start of the 5.8S gene. The ETS fragment extends from position -600 in the ETS to a Sau 3A site at nucleotide +10 at the 5'-start of the 18S gene (Figure 5.3). On the basis of this information I decided to digest pXlcrl-5 with Sau 3A and try and identify these ETS and ITS1 fragments. By a stroke of good luck, a band of approximately 600bp was separated from the many other rDNA and vector bands on a 40cm, 4% polyacrylamide gel. The intensity of the band after ethidium bromide staining confirmed that the ETS and ITS1 fragments had comigrated. The "Sau 600" fragment in each clone was eluted from the gel and the DNA 5'-end labelled. Isolation of specific ETS and ITS1 fragments was achieved by secondary restriction with Ava II which has several recognition sites in both the ETS and ITS1.

A limited amount of sequence analysis was carried out in all three transcribed spacers and short sections of the rRNA coding regions in pXlcrl-5. Fragments were prepared for sequencing as described (Chapter 5, Section 5.3) and then subjected to various complete sequencing runs or to G or C cleavage runs in selected regions of the transcribed spacers and rRNA coding regions.

6.4 The internal transcribed spacers

Figures 6.4 and 6.5 illustrate the sequencing runs carried out and the results obtained for ITS1 and ITS2

Figure 6.4 Summary of ITS1 sequencing runs and variants identified in pX1cr1-5.

pX1r101 is the reference sequence : on this line is indicated every site where a transcribed spacer variant has been found in another rDNA source. Where substitutions were found in other sources, the base that occurs in pX1r101 is shown above the line, together with the nucleotide number below the line. For insertions or deletions in other sources a box is shown above the pX1r101 line and the number of the adjacent nucleotide in pX1r101 below the line. For pX1cr1-5, a solid line represents coverage by complete sequencing in the region indicated; an interrupted line indicates G cleavage only. A gap indicates that no data are available. Substitutions are denoted by showing the base that differs from that in pX1r101. Insertions with respect to pX1r101 are shown thus  , with the number of extra nucleotides, +1 etc. Deletions with respect to pX1r101 are shown thus  , with the number of deficient nucleotides -1 etc. The question mark at position 144 in pX1cr1 means that it was not possible to say with complete certainty that T is present (as in pX1cr4) because of poor T chemistry. Between nucleotides 525 and 551, the inset shows the variant phase-shifted sequence of pX1cr4 (written for the rightwards strand). pX1cr1 showed a similar G cleavage pattern to pX1cr4 but was not fully sequenced.

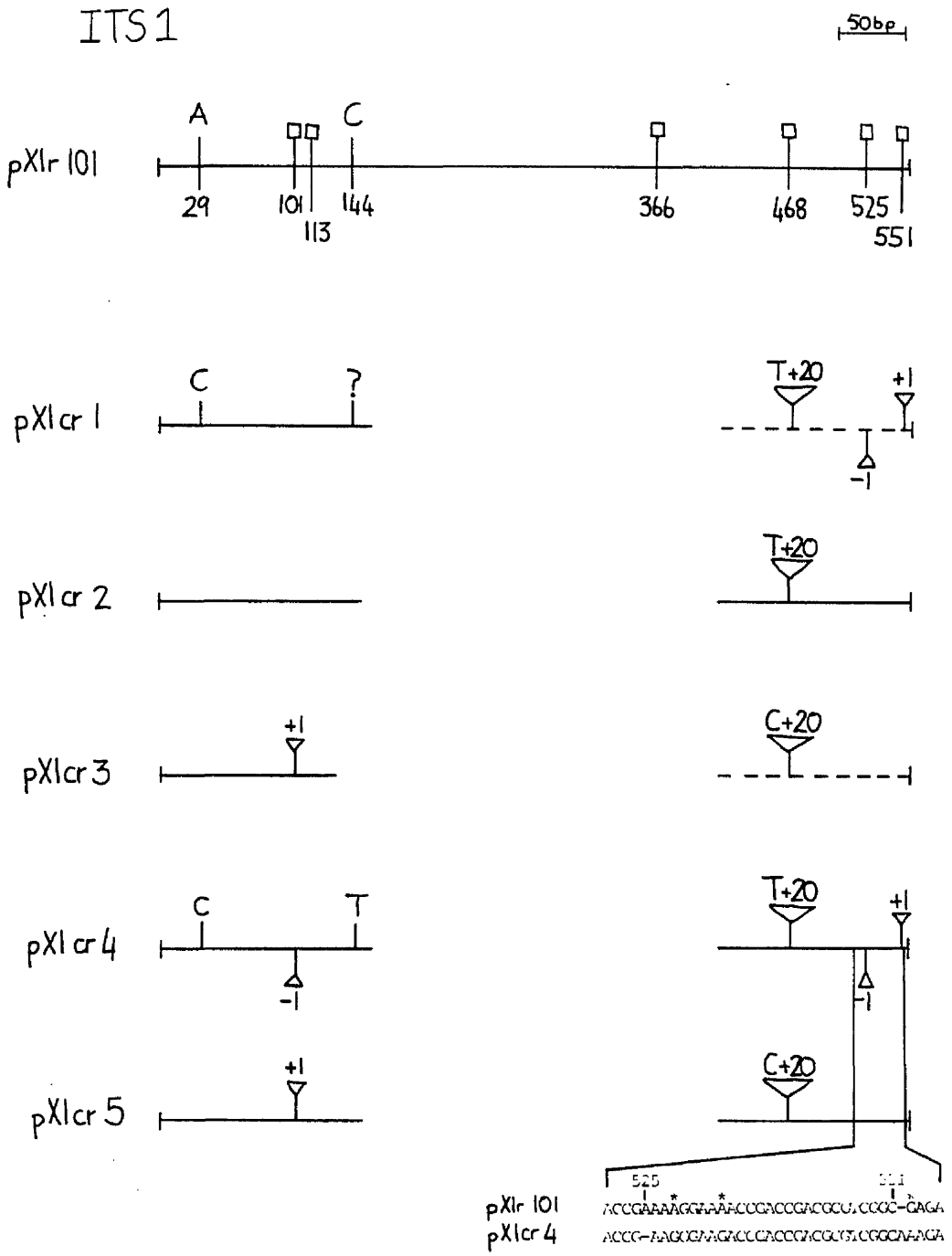

ITS 1
50bp
pXlr 101
A
C
29
101
113
144
366
468
525
551
pXlcr 1
C
?
T+20
+1
-1
pXlcr 2
T+20
pXlcr 3
+1
C+20
pXlcr 4
C
T
T+20
+1
-1
pXlcr 5
+1
C+20
525
551
pXlr 101
pXlcr 4

Figure 6.4

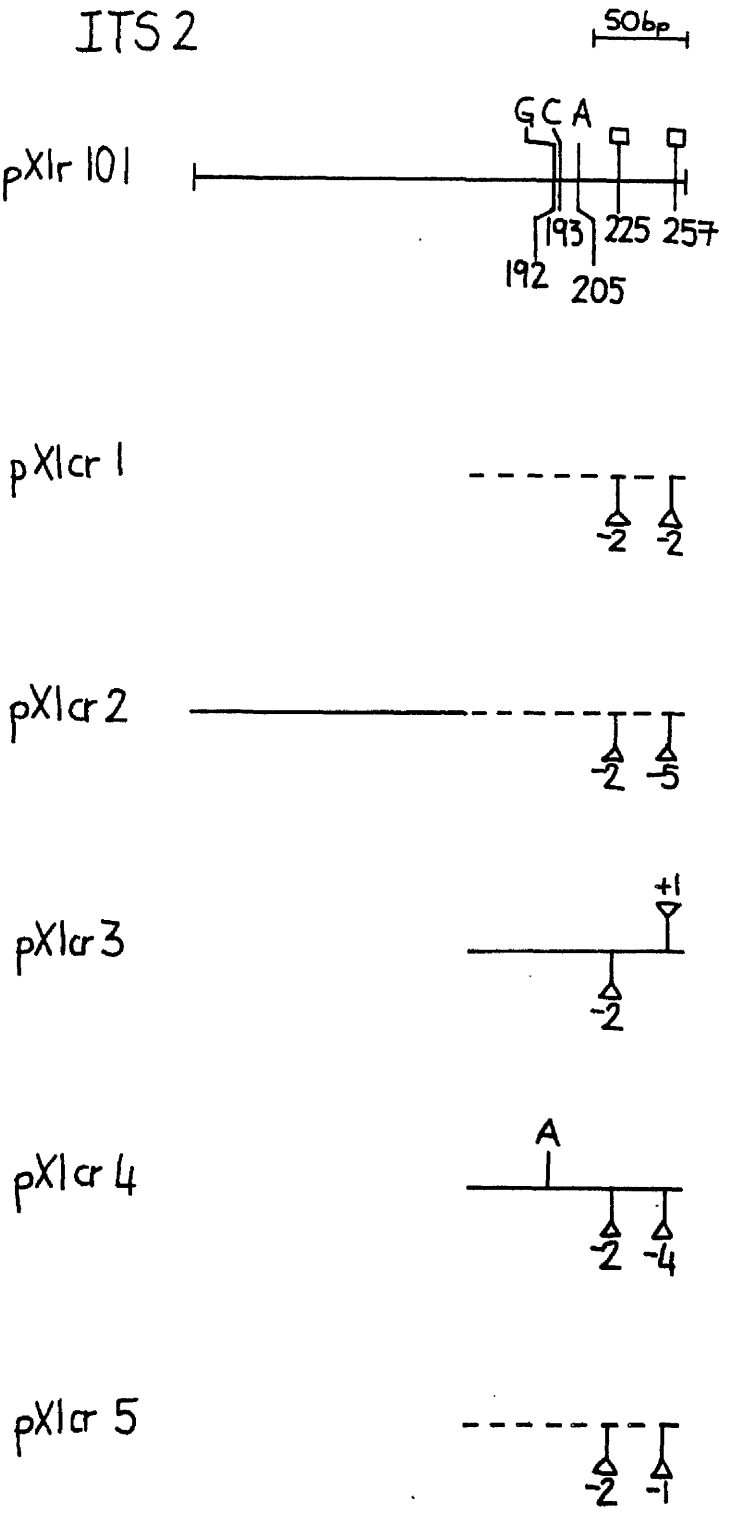


Figure 6.5 Summary of ITS 2 sequencing runs and variants identified in pXlcr1-5.

pXlr101 is the reference sequence. Figure markings are as described in figure 6.4 with one exception. An interrupted line indicates C cleavage only was performed. In pXlcr4, the A residue indicated is at position 192.

respectively. The findings (including those for the ETS, see below) corresponded well to those for amplified rDNA. In addition some new sites of variation were identified.

A full sequencing run extending rightwards from the Sau 3A site at the 3' end of the 18S gene allowed comparative analysis of the first 150 nucleotides of ITS1 in all five clones. In this region it will be recalled, that two sites of variation were observed in amplified rDNA (Figure 5.4). At position 29, amplified rDNA shows A/C heterogeneity. This was also true for the chromosomal clones : pX1cr 2, 3 and 5 contain A whereas pX1cr 1 and 4 contain C (Figure 6.4). Further downstream length variation in the oligo-C tract following nucleotide 100 was observed in the chromosomal clones. I was unable to determine the corresponding number of guanines in the adjoining oligo-G tract because sequencing was performed on the rightwards strand only. However, full sequence determinations in the amplified rDNA clones always revealed equivalent numbers of C and G. By analogy, therefore,length variation in the oligo-G tract is inferred in chromosomal rDNA. A new site of variation, not found in the amplified rDNA clones examined, occurs in pX1cr 4 (and possibly pX1cr 1, see legend to Figure 6.4) at position 144. A C $\longrightarrow$ T change is present.

The 3'-end of ITS1 had revealed sites of major variation in amplified rDNA. Accordingly, this region was studied in great detail in the chromosomal clones (Figure 6.4). Initially a G cleavage experiment was carried out extending leftwards from the Hinf I site at the start of the 5.8S gene.

Full sequencing runs were then performed on pX1cr2, 4 and 5. All of the clones exhibited a G cleavage pattern compatible with an extra block of nucleotide between nucleotides 468 and 469. None of the clones resembled pX1r101 and pX1r103 at this point, all of them contain the extra 20bp segment. Full sequence analysis on three of the clones verified this finding and confirmed the variable nature of nucleotide 468[3] (Figure 6.4).

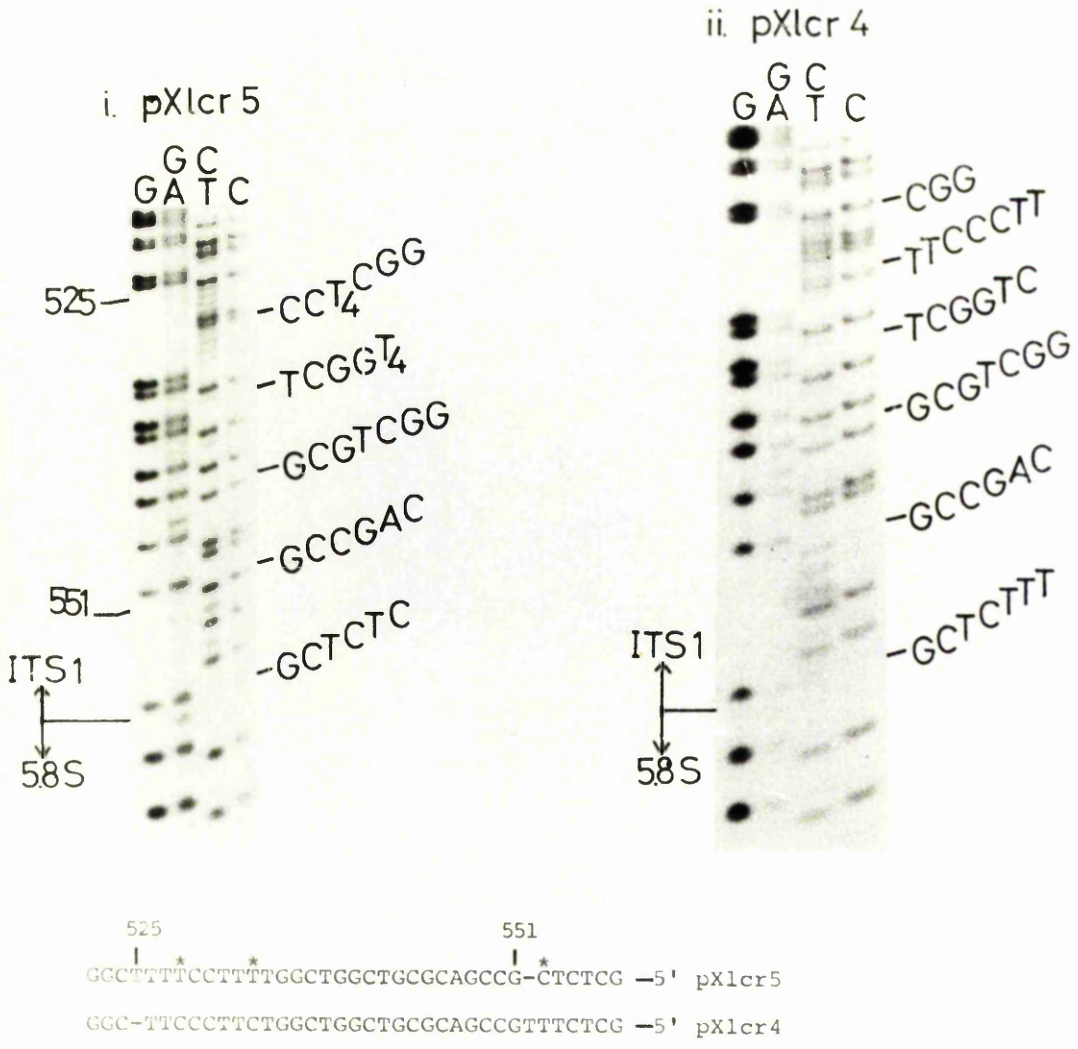
I mentioned in Chapter 5 (Section 5.4.1) that uncloned amplified rDNA revealed a block of ambiguous nucleotides near the start of the 5.8S gene. Analysis of the G cleavage pattern for two clones pX1cr1 and 4 and full sequencing of pX1cr4 showed that both these clones possess variant sequences in this region. Figure 6.6 illustrates the appropriate region of two sequencing gels. pX1cr5 shows the same sequence as pX1r101. pX1cr4 contains the variant sequence. A single base deletion occurs at position 525 in ITS1 followed by a single base insertion following position 551. The region between nucleotides 525 and 551 in the variant sequence is thus "phase-shifted" by one base with respect to the reference sequence. Superposition of the reference sequence and "phase-shifted" sequence would generate a pattern of bands similar to those found in uncloned amplified rDNA. Thus, this chromosomal variant was present in uncloned amplified rDNA but was not found in any of the amplified rDNA clones analysed. In addition, pX1cr4 shows variation within variation : at positions 528/533 and 552 a A $\longrightarrow$ G/A→G and G $\longrightarrow$ A base change has occurred (Figure 6.4, inset).

Figure 6.6 Full sequencing determination of the "phase-shifted" sequence in pX1cr4.

The Figure shows the shortest loading of the full sequencing gels obtained for pX1cr4 and pX1cr5 at the extreme 3' end of ITS1.

Gel(i) shows the sequencing run obtained for pX1cr5 in this region, extending leftwards from the Hinf I site at the start of the 5.8S gene. Gel(ii) shows the equivalent sequencing run for pX1cr4. pX1cr5 exhibits the same sequence type as pX1r101, pX1cr4 exhibits the "phase-shifted" sequence type.

Nucleotides are numbered according to the pX1r101 numbering system. The sequences at the bottom of the Figure are for the leftwards strand. * denotes variant nucleotides between positions 525 and 551.



525 551

GGCTTTTCCTTTTGGCTGGCTGCGCAGCCG-CTCTCG —5' pXlcr5

GGC-TTCCCTTCTGGCTGGCTGCGCAGCCGTTTCTCG —5' pXlcr4

Figure 6.6

ITS2 was analysed in a similar way to ITS1 (Figure 6.5). No sites of variation had been found in the first half of ITS2 in amplified rDNA. This was also true of chromosomal rDNA : full sequence analysis of pX1cr2 in this region revealed 100% homology with the amplified rDNA sequence. Further downstream, length variation in both homopolymeric C tracts was present in pX1cr1-5. The 11 cytosines found in the first tract in pX1r101 appears to be the least common length variant. All the other amplified and chromosomal rDNA sources contain 9 C. The second tract is more variable among the different sources, ranging from 10 to 16 C. Chromosomal rDNA showed one new variant in ITS2, a G $\longrightarrow$ A change in pX1cr4 at position 192 (Figure 6.5) next to the site of C $\longrightarrow$ G variation at 193 in pX1212.

6.5 ETS

Sequence analysis in the ETS covered the 3' region extending leftwards from a Sau 3A site at position +10 in the 18S gene. The sequencing runs performed and the results obtained are summarised in Figure 6.7.

Between nucleotides -20 and -33 amplified rDNA displayed two types of sequence : pX1r101-type sequence common to pX1r101, 102, 103 and fragment 2a from uncloned amplified rDNA, and pX1r14-type sequence found in pX1r14, pX1108 and fragment 2b from uncloned rDNA. Both sequence types are present in chromosomal rDNA : pX1cr2, 3 and 5 show pX1r101-type sequence whereas pX1cr1 and 4 display a pX1r14-type sequence. No new sequence types were observed.

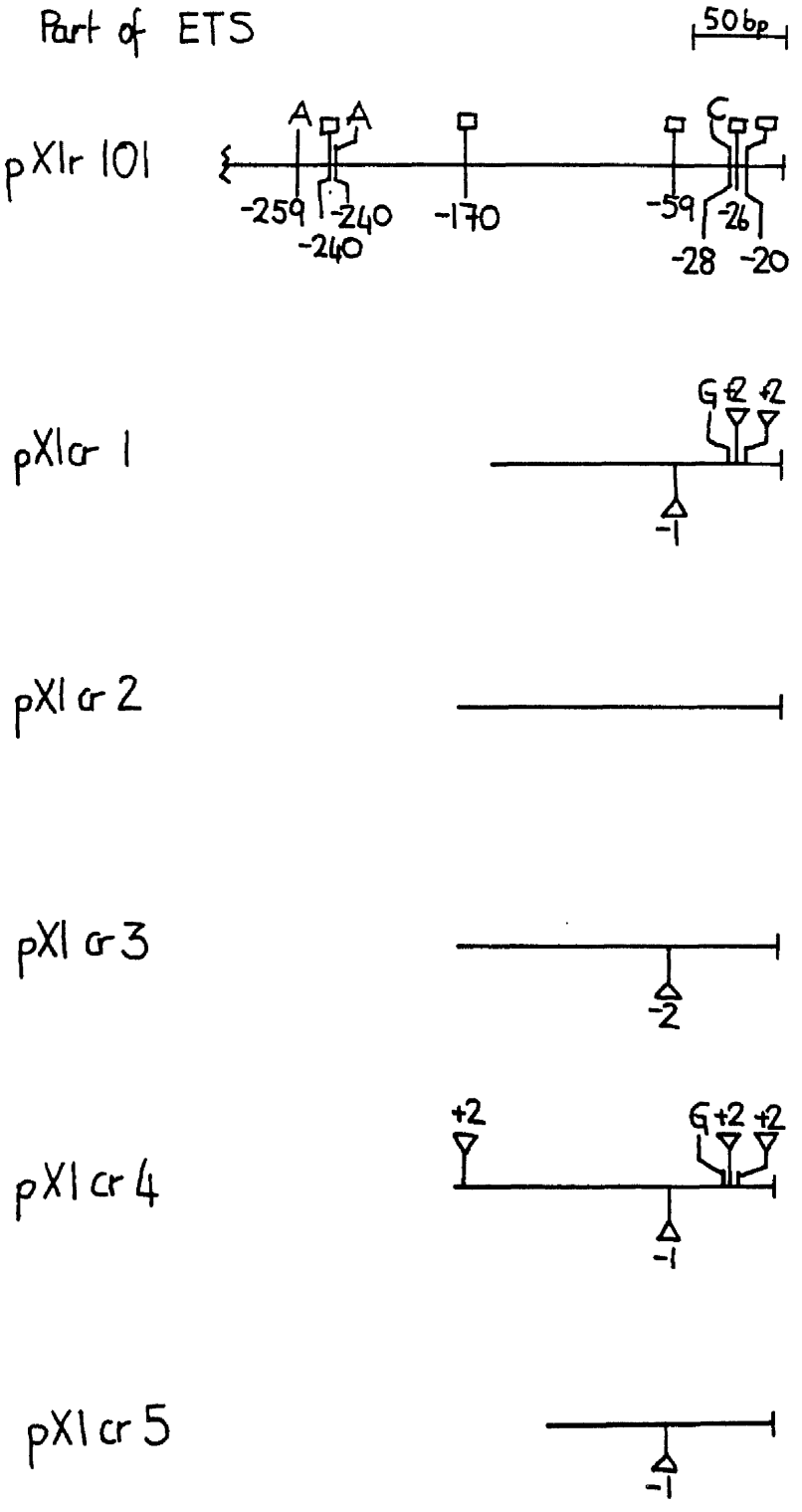


Figure 6.7 Summary of ETS sequencing runs and variants identified in pXlcrl-5.

pXlrl01 is the reference sequence. Figure markings are as described in figure 6.4.

Further upstream, the sites of variation previously identified in amplified rDNA were identified in corresponding positions in the chromosomal clones. pXlcr2 was the only clone to show complete ETS homology in the region examined, with the reference clone, pXlrl01 (Figure 6.7).

6.6 rRNA coding regions

Sequence analysis of the rRNA coding regions was limited to the last 60 nucleotides of the 18S gene in four of the clones and the entire 5.8S gene in one clone pXlcr2. No sites of sequence variation were identified - the sequences show complete homology with the sequences obtained for amplified rDNA. The start of the 28S gene was analysed by a C cleavage experiment extending rightwards from the Bam HI site in ITS2. pXlcr2 showed a slightly different pattern in the 28S region to the remaining clones and full sequence analysis verified the presence of an extra C adjacent to G40 in the sequence. This variant was also identified in the amplified rDNA clone pXlrl02.

6.7 Conclusions

Analysis of the transcribed spacers of *X. laevis* chromosomal rDNA clones has revealed several sites of microheterogeneity previously identified in both cloned and uncloned amplified rDNA. A sequence variant present in uncloned amplified rDNA but not found in any of the amplified rDNA clones was identified in two of the chromosomal clones. In addition, two new sites of variation in the form of single base changes were observed. These results are reassuring:it seems unlikely

that amplification of rDNA during oogenesis introduces sequence variants into the transcribed spacers of amplified rDNA.

In the introduction to this thesis (Section 1.4), I discussed the validity of using amplified rather than chromosomal rDNA for detailed characterisation of X. laevis rDNA. I concluded that when large preparations of amplified rDNA are made from many frogs a representative selection of chromosomal rDNA repeats is present in the amplified rDNA. Both the uncloned and cloned amplified rDNA used in this study originated from groups of frogs which had been bred at the Carnegie Institute of Washington, Department of Embryology 1973-1976, after descent by one or a few generations from the wild. The chromosomal rDNA used in this analysis was obtained from frogs supplied by a commercial company and originated from the wild. The findings reported here indicate that pooled amplified rDNA does give a true representation of chromosomal rDNA by the criteria of direct DNA sequence analysis in the transcribed spacer regions.

7. Analysis of rDNA from a single X. laevis

The preceding three chapters have been concerned with determining the nature and extent of sequence variants in the transcribed spacers of both amplified and chromosomal rDNA purified from many X. laevis frogs. Heterogeneities including single base changes and length variants of one to several nucleotides were identified in all three transcribed spacers. These results however, give no indication of the extent of transcribed spacer variation within single frogs : more specifically within single nucleolar organisers. The most basic level for the introduction of sequence variants within the transcribed spacers of X. laevis rDNA must exist within single nucleolar organisers. I anticipate that the heterogeneities identified in pooled rDNA reflect a low level of sequence variation present within single nucleolar organisers which, given time, have spread through and become established within the rDNA multigene family. My final objective in this project, therefore, was to purify chromosomal and amplified rDNA from individual X. laevis and determine the extent of intranucleolar transcribed spacer variation.

Wild type X. laevis contain their diploid complement of rRNA genes within two nucleolar organiser regions (2-nu). However, mutants of X. laevis exist which contain only one nucleolar organiser. These heterozygotes (1-nu) regulate their genes to produce twice as much 28S and 18S rRNA as the same genes when present in wild-type (2-nu) individuals. rDNA isolated from a 1-nu X. laevis would provide the ideal material for determining the extent of intranucleolar transcribed spacer variation because any sequence variants identified

must originate within the single nucleolar organiser of the animal. However, such frogs must be specially bred and reared and are therefore not easily obtained. I decided instead to study rDNA isolated from a single wild-type (2-nu) X. laevis.

In order to show that transcribed spacer variation exists at the level of a single nucleolar organiser, at least three types of transcribed spacer sequence must be identified in rDNA isolated from a 2-nu animal. Identification of only two types of sequence does not rule out the possibility that each nucleolar organiser may only contain a single sequence type (intranucleolar homogeneity) rather than a mixture of the two sequences (intranucleolar heterogeneity). Identification of a third sequence type eliminates this possibility. It does not matter how the sequence types are arranged within each nucleolar organiser because in all of the possible arrangements one of the nucleolar organisers must contain at least two types of sequence. Thus intranucleolar variation would exist if three types of variant were found in the rDNA of a single frog.

The initial objectives of this particular sequencing project were 1) isolation and purification of chromosomal and amplified rDNA from three individual X. laevis animals and 2) sequence analysis of both ends of ITS1 and the 3' end of ETS purified from the chromosomal and amplified preparations from each frog.

Purification of amplified rDNA was very successful giving yields of approximately 4, 8 and 10µg of rDNA from the three

animals. In contrast, I was unable to obtain very pure preparations of chromosomal rDNA because of contamination of the preparations with total nuclear DNA. I decided therefore, to limit my sequence analysis to amplified rDNA. In addition, detailed analysis of amplified rDNA from each frog generates a substantial amount of work. So in the time left available to me, I constructed ETS and ITS1-containing clones using amplified rDNA from one frog and then concentrated by sequence analysis on the 3' end of ITS1. During the writing up of this thesis, an Honours student in our laboratory, Alan Warren, has extended this sequence analysis to cover the 5' region of ITS1 and the 3' region of the ETS.

7.1 Restriction and cloning strategy

The quickest and most efficient way of achieving my sequencing objective was to utilise the M13 cloning system in conjunction with the dideoxy sequencing method (see Chapter 2). A survey of the available M13 vectors revealed that two related vectors, M13mp10 and M13mp11, contained unique Sst I and Xba I restriction nuclease recognition sites (see Figure 2.4). This combination was of particular interest since a survey of restriction sites in *X. laevis* rDNA revealed the strategic location of Sst I and Xba I sites for cloning the transcribed spacer regions (Figure 7.1). On the basis of this information I decided to clone the ETS Sst I/Xba I fragment into M13mp10 and the ITS1 Xba I/Sst I fragment into M13mp11. Cloning into M13mp10 allows sequence analysis of the 3' end of the ETS whereas cloning into M13mp11 allows sequence analysis of the 3' end of the ITS 1 fragment (Figure 7.2). Sequence analysis

Figure 7.1 Restriction of amplified rDNA purified from a single X. laevis

Upper section : Map of Sst I and Xba I sites in rDNA predicted from amplified rDNA clones. Lines with arrowheads represent fragments obtained from complete digestion of amplified rDNA, from "frog 3", with Sst I and Xba I. Fragments 1-5 are numbered according to size.

Lower section : Restriction digests of amplified rDNA. Restriction products were separated on a 1% horizontal agarose gel and visualised, after ethidium bromide staining, by U.V. illumination. The wide central section represents the preparative Sst I plus Xba I digest of amplified rDNA prepared from "frog 3". The fragment numbers correspond to those in the upper section : the relationship of the fragments in this preparative digest was inferred from the known distances between restriction sites in cloned rDNA and was confirmed by a complete Sst I plus Xba I digest of pX1r101.

N.B. The line marked "X" in the upper section of the figure represents a group of fragments in the amplified rDNA preparation which are variable in length due to variation in the position of the first Sst I site in the NTS. These fragments are present in submolar amounts.

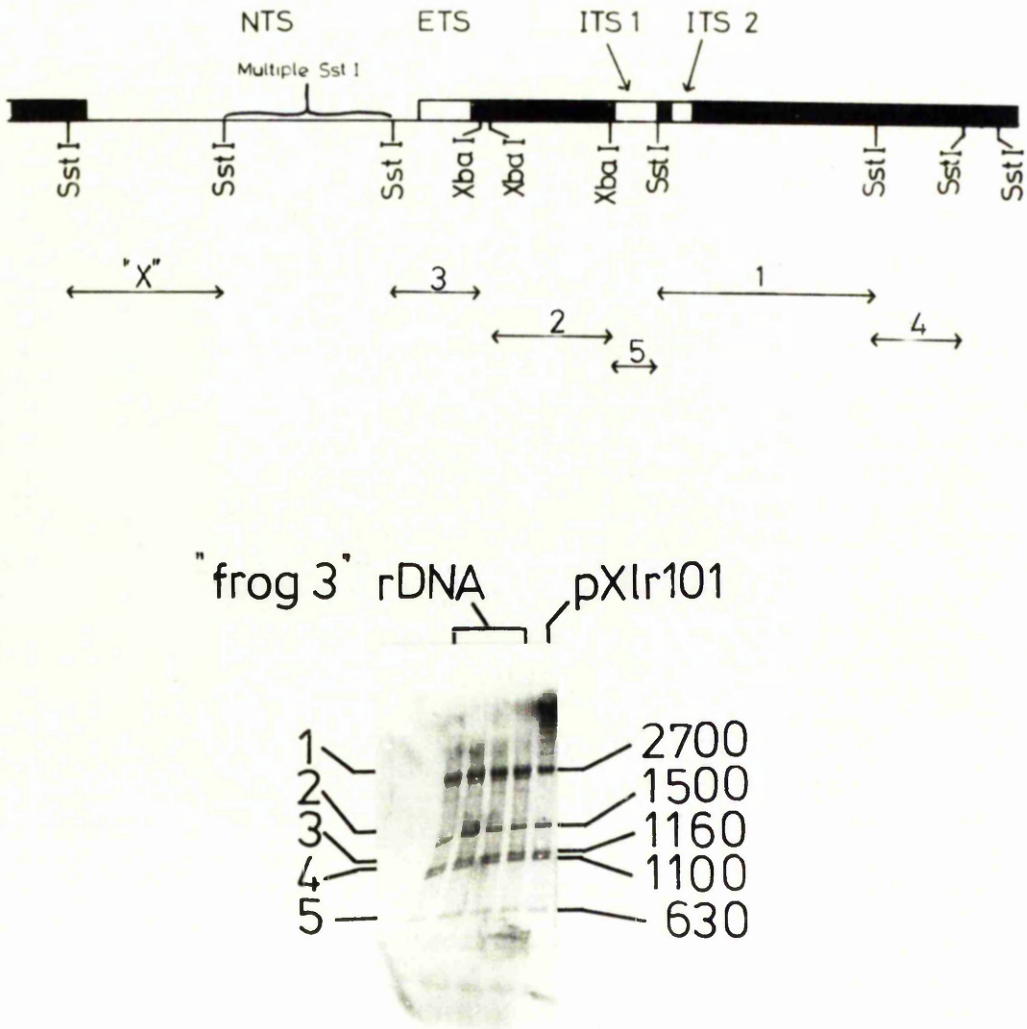
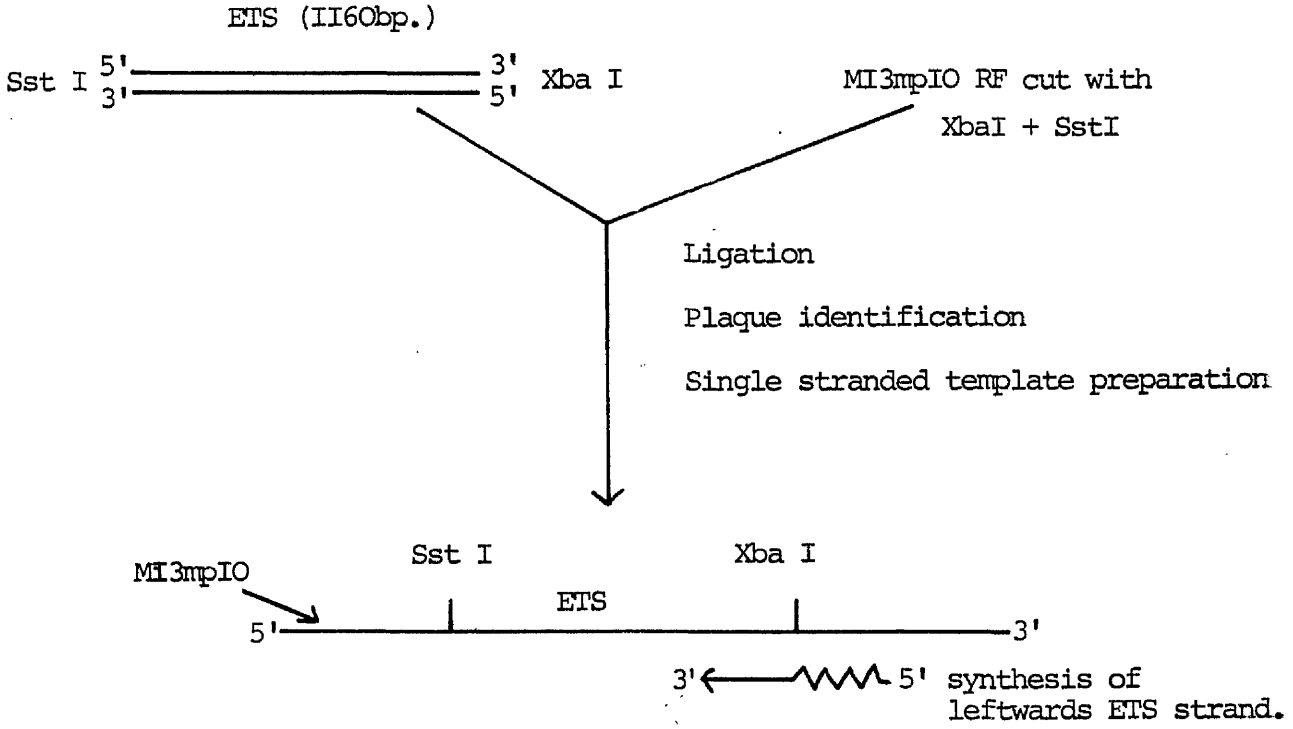

NTS
ETS
ITS 1
ITS 2
Multiple Sst I
Sst I
Sst I
Sst I
Xba I
Xba I
Xba I
Sst I
Sst I
Sst I
Sst I
"X"
3
1
2
4
5
"frog 3" rDNA
pXlr101
1
2
3
4
5
2700
1500
1160
1100
630

Figure 7.1



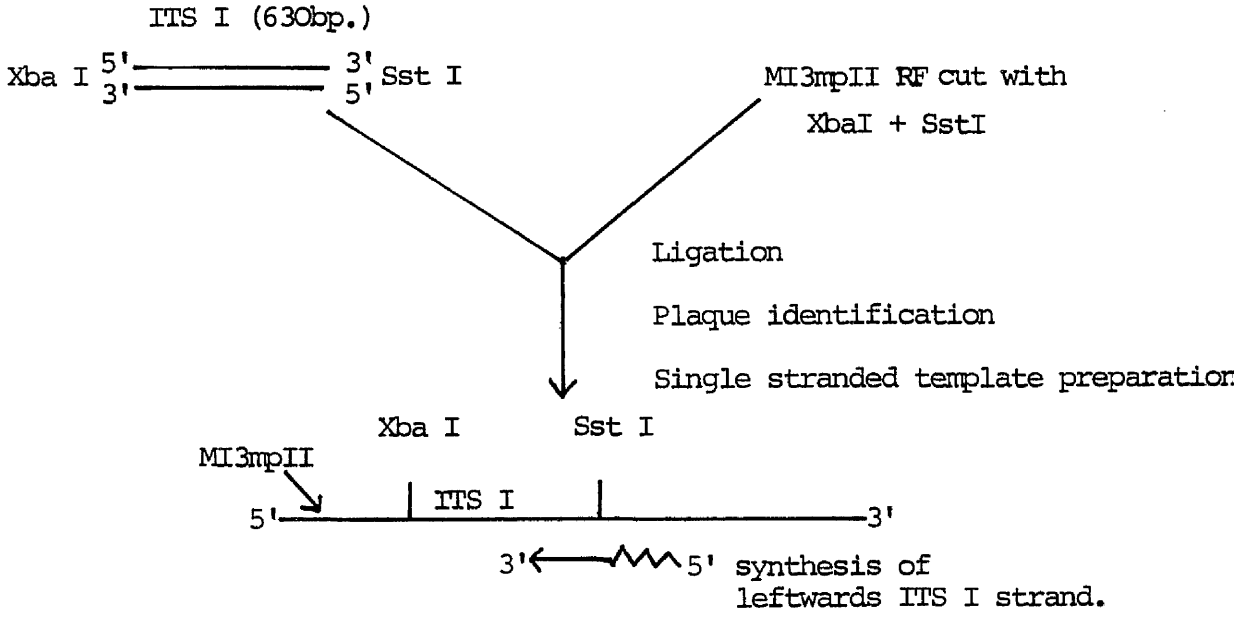


Figure 7.2 M13 cloning strategy for analysis of the 3'region of the ETS and ITS 1 of amplified rDNA.

of the 5' end of ITS1 (by A. Warren) was achieved by cloning an aliquot of the ITS1 Xba I/Sst I fragment into M13mp11.

Figure 7.1 illustrates the restriction fragments obtained when amplified rDNA from "frog 3" was digested to completion with Xba I and Sst I and resolved on a 1% agarose gel. The pattern obtained corresponds well with the restriction pattern of the reference clone pX1r101. The ETS (band 3) and ITS1 (band 5) containing fragments were recovered from the "frog 3" rDNA digest and an aliquot of each fragment used in the subsequent cloning experiment.

Construction of the M13mp10 and mp11/ rDNA clones was performed as described (Chapter 3, Section 5.1) using 0.15pmoles of rDNA and 0.05pmoles of linearised M13 vector in each ligation reaction. Recombinant molecules were identified by their ability to form white plaques on IPTG/BCIG plates (Chapter 3, Section 5.3). In both experiments over 95% of the plaques obtained were white. The ETS cloning experiment yeilded fewer plaques (246 total, 233 white) than the ITS1 cloning experiment (682 total, 670 white). This difference in number may reflect the influence of insert size during the cloning step. As insert size increases from 630bp (ITS1 fragment) to 1160bp (ETS fragment) the cloning efficiency decreases. The plates were stored at $4^{o}C$ prior to dideoxy sequence analysis of a number of white plaques. It has been reported that plaques stored in this way are unstable and have a tendancy to lose insert, especially if insert size is over 1kb. All of my sequencing studies were performed on DNA prepared from plaques within six weeks of the cloning

experiment and I encountered no problems with insert instability. In contrast, sequence analysis of the ETS using DNA prepared four months after the initial cloning experiment (A. Warren) revealed that the recombinant DNA was unstable and deletions had arisen in the insert sequence.

7.2 Analysis of the 3'-end of ITS1 by partial dideoxy sequencing

In my previous search for transcribed spacer variants using the Maxam and Gilbert chemical sequencing method, I chose the G specific reaction to compare the 3' end of ITS1 from cloned and uncloned rDNA. A characteristic G cleavage pattern was obtained for the different sequence types present in this region (see for example Figure 5.6). I decided to adopt a similar screening technique for analysing the newly constructed ITS1 clones using the dideoxy sequencing method. Single stranded DNA was prepared from 81 white plaques and dideoxy sequencing using the G termination reaction only was performed. Of the 81 clones analysed in this way, 69 contained the ITS1 sequence.

The most striking feature revealed by analysis of the G termination pattern was the occurrence of the extra block of 20 nucleotides, following nucleotide 468, in all 69 clones. This 20bp segment had previously been identified in three (of the five) amplified rDNA clones, in uncloned amplified rDNA and in all five chromosomal rDNA clones (Chapters 4-6). Figure 7.3 shows the consensus sequence (leftwards strand) derived previously from full sequencing runs on amplified

Figure 7.3 Screening of the 3' region of ITS1 by partial dideoxy sequencing.

The G termination banding pattern in the region of ITS1 which was shown to vary by the presence or absence of a 20bp block of nucleotides in pooled rDNA, is shown for six M13mp11/ITS1 amplified rDNA clones. Nucleotides are numbered according to the pX1r101 numbering system. The banding pattern illustrated, was identical in all 69 clones examined and comparison with the sequence for pX1212 (sequence of leftwards strand for pX1212 is shown) demonstrated the presence of the 20bp block of nucleotides in all 69 clones. One site of variation was identified - at position 468^3 either a G band (lane 2) or a space (lanes 1, 3 $\longrightarrow$ 6) was obtained, representative of a C or T respectively, on the rightwards strand (see text for explanation).

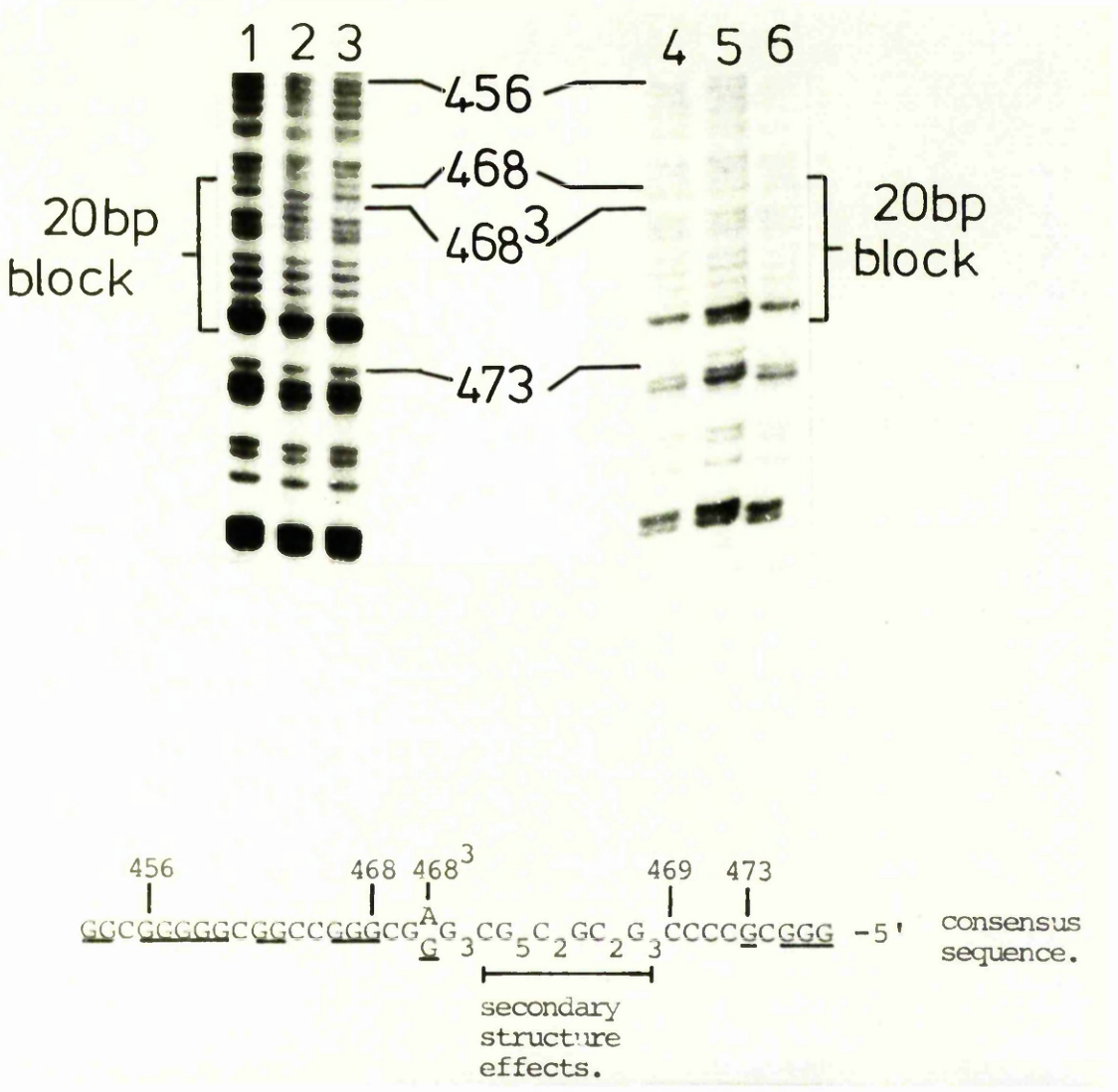

1 2 3
4 5 6
456
468
468 3
473
20bp block
20bp block
456
468
468 3
469
473
GGCGGGGGCGGCCGGGCGA/GG3CG5C2GC2G3CCCCGCGGG -5'
consensus sequence.
secondary structure effects.

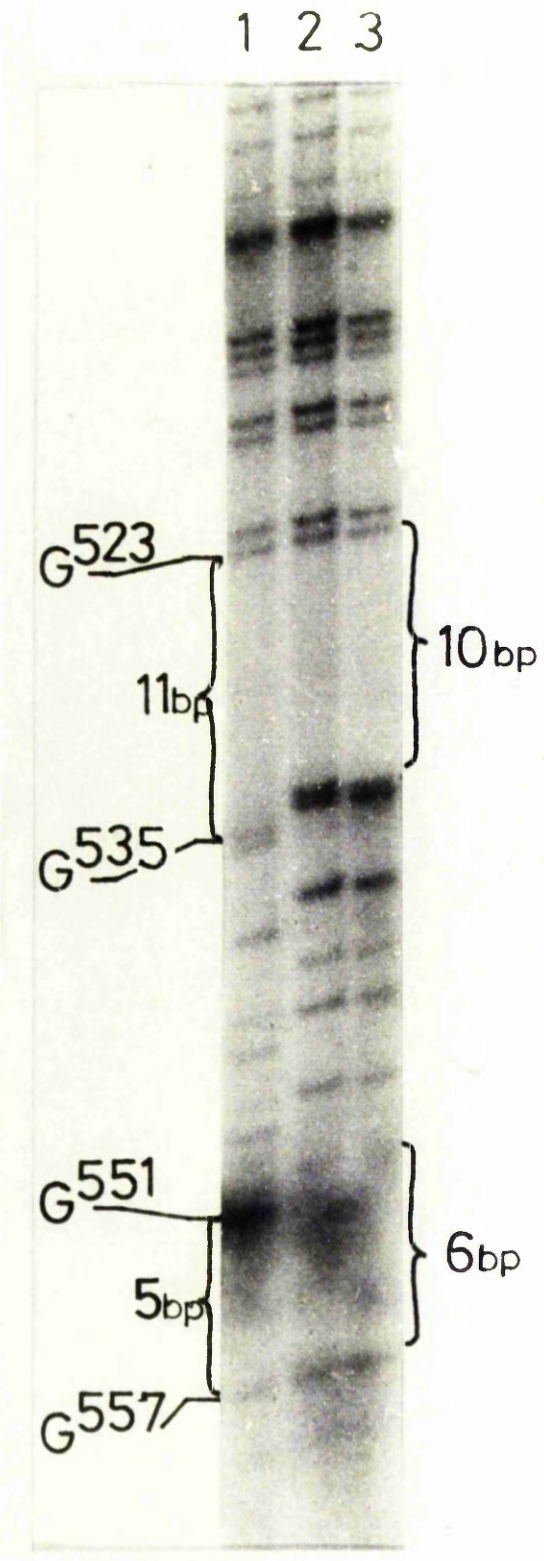
Figure 7.3

and chromosomal rDNA clones, and the G pattern obtained for six newly constructed ITS1 clones in the region of nucleotide 468. The spacing and position of bands in the G pattern is consistent with a DNA sequence containing a 20bp segment between nucleotides 468 and 469. The banding pattern within the 20bp segment gives some spacing anomalies (see Figure 7.3) because of strong local secondary structure. However, it is possible to identify single base variation at nucleotide 468^3. The presence of a band at 468^3 indicates a G, therefore C on the rightwards, RNA-like strand. In contrast, a space at position 468^3 implies, from previous full sequencing runs on amplified and chromosomal rDNA clones, that A is present at this position. Therefore, T is present on the rightwards, RNA-like strand. Of the 69 clones analysed 62 displayed the pattern lacking G at position 468^3 (see Figure 7.3, lanes 1, 3,4,5 and 6) whereas only 7 clones displayed a band signifying G at position 468^3 (see Figure 7.3, lane 2).

At the extrme 3' end of ITS1 analysis of amplified and chromosomal rDNA (Chapters 4-6) demonstrated two types of sequence : the reference type sequence, as in pX1r101, was displayed by all of the clones except two chromosomal clones, whereas a "phase-shifted" sequence was present in pX1cr1 and pX1cr4. Figure 7.4 (lower section) shows the sequence of the leftwards strand of pX1r101 and pX1cr4 in this region. pX1cr4 contains a single base deletion at nucleotide 525 and a single base insertion following nucleotide 551. Each sequence type gives a characteristic G banding pattern : between G557 and G551 (see Figure 7.4) pX1r101 contains five nucleotides whereas pX1cr4 contains six nucleotides. Similarly between

Figure 7.4 Identification of "phase-shifted" sequence at the 3' end of ITS1.

The nucleotide sequence for the leftwards strand of pX1r101 (reference sequence) and pX1cr4 (phase-shifted sequence) between nucleotides 523 and 557 are shown. In addition, the G termination banding pattern in this region is illustrated for three M13mp11/ITS1 amplified rDNA clones. Nucleotides are numbered according to the pX1r101 numbering system and bands in the gel correspond to guanine in the nucleotide sequences shown. Lanes 2 and 3 demonstrate a pX1cr4-type sequence pattern, lane 1 demonstrates a pX1r101-type sequence pattern.


1 2 3
G523
11bp
G535
10bp
G551
5bp
G557
6bp

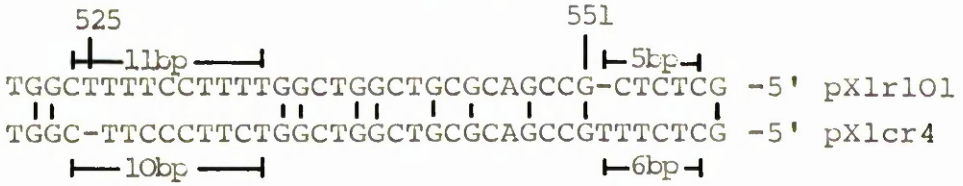

525
551
11bp
5bp
TGGCTTTTCCTTTTGGCTGGCTGCGCAGCCG-CTCTCG -5' pX1r101
TGGC-TTCCCTTCTGGCTGGCTGCGCAGCCGTTTCTCG -5' pX1cr4
10bp
6bp

Figure 7.4

G535 and G523 (see Figure 7.4) pX1r101 contains eleven nucleotides whereas pX1cr4 contains ten nucleotides. Thus the pX1cr4 sequence is "phase-shifted" by a single base with respect to the pX1r101-type sequence. Comparison of the G termination pattern obtained for the 69 ITS1 clones (see main section, Figure 7.4 for an example) revealed that the pX1cr4 sequence type was the more frequent sequence type in this frog. Only 10 clones displayed a banding pattern characteristic of the reference sequence (Figure 7.4, main section, lane 1). This contrasts with the results obtained for the amplified and chromosomal clones for rDNA prepared from many frogs. Here only two, out of the ten clones studied in this region, displayed the "phase-shifted" sequence. The reference sequence was the more common sequence type.

7.3 Conclusions

In order to demonstrate that intranucleolar transcribed spacer variation exists, at least three types of sequence must be identified in rDNA purified from 2-nu X. laevis. In these experiments I have identified four types of sequence at the 3'end of ITS1 (Figure 7.5). All the sequence types contain the extra 20bp segment at position 468, however, some clones contain T at 468^3 others C. In addition, further downstream at the extreme 3' end of ITS1, some clones display a "phase-shifted" pX1cr4-type sequence between position 525 and 551 whereas others display the reference pX1r101-type sequence. The relative abundance of the different sequence in this frog is very different. The most common sequence type, present in 55 of the 69 clones examined, contains T at

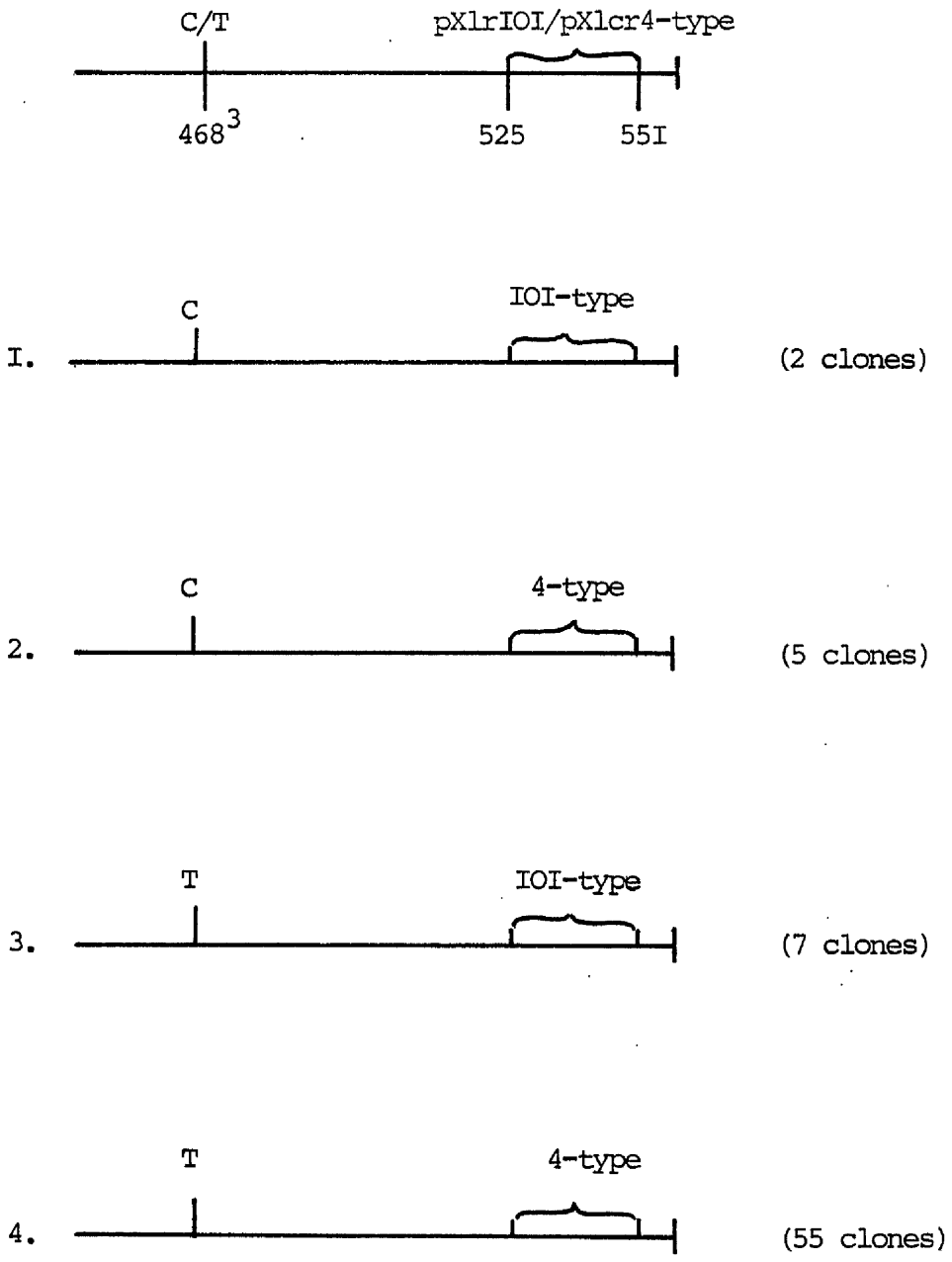


Figure 7.5 Summary of the four different sequence types identified at the 3'end of ITS 1 in amplified oocyte rDNA from a single frog.

position 468[3] and displays the pXlcr4-type sequence further downstream (Figure 7.5, sequence type 4). Thus identification of four sequence types in this sequencing study demonstrates, unequivocably, that transcribed spacer variation within single nucleolar organisers does exist.

8. Discussion

8.1. Nature and extent of variation in the transcribed spacers of X. laevis rDNA

Before this work was commenced it was generally supposed that the transcribed spacers of rDNA within a single species were homogeneous with respect to length. In this work, I have carried out direct DNA sequence analysis on the transcribed spacers of X. laevis from the following sources:-

1. clones from amplified rDNA
2. uncloned amplified rDNA
3. clones from chromosomal rDNA
4. clones from amplified rDNA from a single frog

The regions examined in DNA from these various sources covered the extreme 3' end of the ETS and the whole of both ITS1 and ITS2, a total of approximately 1,100 nucleotides of the rDNA unit structure. Substantial microheterogeneity was observed. Transcribed spacer variants comprising base substitutions and insertions and deletions of one or a few nucleotides were identified at some 20 sites. In addition, a site in ITS1 has been shown to vary by the presence or absence of a block of 20 nucleotides (see below).

Sequence variants identified in amplified rDNA clones were also identified in uncloned amplified rDNA. This finding confirmed that heterogeneities in cloned rDNA are not artefacts introduced during cloning. Comparison of the sequence data obtained for amplified and chromosomal rDNA

LOCATION	DNA SEQUENCE
ETS	
-248 → -221	$GAGAGAGAAAGACGGAAAGAAAGGAGAG^{+}$
-192 → -175	$C_2TC_{10}TC_4^{+}$
-174 → -165	G_{10}^{+}
-59 → -48	C_{12}^{+}
ITS 1	
19 → 32	$GGAGAGAGGG\overset{A}{\underset{C}{}}AGG$
54 → 68	GGAGAGAGAGAGAGA
92 → 113	$C_{10}ACG_{10}^{+}$
378 → 397	GGGGAGGGGAGGGGGGGAGG
525 → 534	$\overset{A}{-}AA\overset{A}{\underset{G}{}}GGAA\overset{A}{\underset{G}{}}A$
ITS 2	
216 → 226	C_{11}^{+}
243 → 257	C_{15}^{+}

Table 8.1 Simple sequence tracts in the transcribed spacers of X.laevis rDNA.

+ means that length variation occurs in this tract.
If a single base change occurs in a tract, this is indicated in the table,

showed that there was very good correspondence between the findings for the two sources. This indicates that the primary reservoir of variant sequences is genomic rDNA and that the variants do not originate during the amplification process. It was important to establish this fact because, in order for variants to spread through a large population of frogs then the variants must occur in the germ line (i.e. chromosomal) rDNA. Analysis of amplified rDNA from a single X. laevis revealed several of the variants which had already been identified in rDNA pooled from many frogs. The significance of this finding is discussed below.

a) Localisation of some heterogeneities in simple sequence tracts.

Specific regions of the transcribed spacers which display a run of 10 or more nucleotides consisting of purines only or pyrimidines only, or which show a strong bias towards a single type of base have been designated as "simple sequence" DNA. All three transcribed spacers contain a number of "simple sequence" tracts (Table 8.1). Several of the transcribed spacer variants occur in these tracts, especially length variants, and eight out of the eleven tracts in the table contain sites of variation. Apparently these simple sequence tracts constitute potential foci for length heterogeneity and mutational "hotspots" within the spacer regions. It is worth noting that although there is a tendency towards simple sequence DNA, there is no evidence for the occurrence of large scale repetitive sequence patterns - a characteristic common to the NTS (1.3.2).

b) 20bp tract in ITS1

The most striking site of variation in the transcribed spacers is located towards the 3' end of ITS1 and is characterised by the presence or absence of a block of 20 nucleotides. Analysis of five amplified and five chromosomal clones (rDNA prepared from many frogs) and 69 amplified rDNA clones (rDNA prepared from a single frog) in this region of ITS1 revealed that only two clones, pX1r101 and pX1r103, lack this 20bp segment. The sample of uncloned amplified rDNA analysed comprised predominantly the version containing the 20bp block of nucleotides with only a small amount of the shorter variant (Figure 5.7).

Two possibilities exist which could explain this finding. It is possible that, in terms of transcribed spacer sequence evolution, the version of ITS1 lacking the 20bp sequence is "older" than the version containing the 20bp sequence. Thus insertion of a 20bp segment has occurred and this new version of ITS1 is now in the process of sweeping through the rDNA family with the simultaneous elimination of the shorter ITS1 version from the family. On the other hand, the predominance of the longer version of ITS1 could mean that this form of ITS1 is in fact "older" than the shorter version. Thus deletion of the 20bp segment could be a recent event and this version of ITS1 is in the primary stages of sweeping through the other members of the rDNA multigene family. In both cases however, the identification of a sequence variant within the 20bp block (position 468^3 T/C heterogeneity) implies that this DNA sequence has been established in the

transcribed spacer long enough to allow the introduction and spread of a new variant within this region through the individual family members.

c) No two transcription units yet examined are identical

No two transcription units examined in this study are identical throughout their transcribed spacer regions. Diversity arises at two levels : different combinations of variants within a single transcribed spacer (ETS, ITS1 or ITS2) and combinations of variants between the three transcribed spacers in a single transcription unit. In order to make this latter observation only those clones containing a complete repeat unit of rDNA can be compared (i.e. pX1r101-103, pX1cr1-5).

d) Identification of linked groups of variants

Examination of the nature of the transcribed spacer variants identified reveals that, despite the extensive diversity observed, some linked groups of variants occur. First, in the ETS between nucleotides -20 and -28 all the clones examined conform to one or other of two alternative patterns, resembling pX1r101 or pX1r14 respectively (Figures 5.11 and 6.7). The pX1r14-type pattern possess two mini-insertions with respect to the pX1r101-type pattern : no intermediate form with only one of the two insertions was found. A second type of linkage pattern is observed in uncloned oocyte rDNA. Fragments 2a and 2b in uncloned amplified rDNA (Figure 5.2) span the entire ETS and part of the NTS. 2a and 2b are length variants due to Bam HI site heterogeneity in the NTS. The longer fragment 2a, contains predominantly the pX1r101-

type sequence whereas the shorter fragment 2b, contains predominantly the pXlrl4-type sequence. Thus in this sample of uncloned rDNA a pattern of sequence variation at the 3' end of the ETS is linked to a pattern of length variation more than 1kb upstream in the NTS. However, not all of the clones examined were found to conform to this linkage pattern. A third example of linkage between variants occurs at the extreme 3' end of ITS1 : the deletion of position 525 is evidently coupled to the insertion following nucleotide 551 (Figure 6.4).

The fact that no intermediate forms of the first (ETS, -20 to -28) and third (ITS1, 525 to 551) type of linkage patterns occur suggests that the pairs of variants within each region have arisen simultaneously or in close succession in the same gene. The gene containing the linked variants would then begin to replace the other members of the family resulting in the pattern observed in this study: in the ETS either pXlrl01 or pXlrl4 type sequence, in ITS1 either a phase-shifted or reference sequence.

8.2 Evolution of the transcribed spacers

All eukaryotic genomes contain families of repeated genes in which individual members of the multigene family show a greater similarity within species than between related species. Variation accumulates between species yet individual members within a species do not evolve independently - a phenomenon known as concerted evolution. Concerted evolution involves three processes : mutation, "homogenising

mechanisms" such as unequal recombination and gene conversion which result in the transmission of variants through the family, and natural selection.

Mutation : The sequence of a single gene can be altered by the chance occurrence of single base changes and insertions and deletions of one or a few nucleotides. It is possible that different regions of a gene may display different mutation rates.

Unequal recombination : The role that unequal recombination plays in the transmission of sequence variants through multigene families and, in particular, the rDNA multigene family has been discussed (1.6). Briefly, when sister chromatids, each containing a tandem array of repeated genes, are paired, the arrays may align out of register with respect to the total array length. Frequent recombination between arrays would thus result in continuous duplication or deletion of individual gene units. In this way, a repeat containing a new mutant will either be lost through deletion or will be duplicated and gradually encompass the entire gene array (Figure 1.3).

Gene conversion : This mechanism is thought to maintain homogeneity within genomically dispersed and tandemly linked families of genes (Dover, 1982, references therein). In gene conversion, two initially slightly different copies end up sharing the DNA sequence of one of them: one copy converts the other. The direction of conversion is often random, but sometimes one variant copy persistently converts the other.

The direction of conversion is often random, but sometimes one variant copy persistently converts the other. Random fluctuations in the direction of conversion, or a bias in the direction of conversion eventually ensures that all copies of a gene family in all individuals of the same species, are of the same type.

Natural selection : Natural selection determines the viability of a particular gene whether the gene is a single copy gene or a member of a multigene family. Consider the rDNA multigene family. If a mutation, introduced into a rRNA gene, is deleterious to ribosome formation, then natural selection will eliminate any gene harbouring the mutant. Hence the overall rate of change within that particular region of the gene is slow. On the other hand, mutants having no deleterious effect i.e. neutral mutants, will be accepted by natural selection and the overall rate of change within that region of the gene will be determined by the mutation rate and the rate at which mutations can sweep through the multigene family via the homogenising mechanisms per unit time.

It is clear from analysis of DNA sequence data that the rates of change within different regions of the rDNA unit are very different. NTS length heterogeneity is of widespread occurrence in eukaryotic rDNA and is correlated with rapid evolutionary change within these regions (Coen et al, 1982). In contrast, the overall rate of change in the rRNA coding regions is slow. Presumably natural selection eliminates gene copies in which mutations, deleterious to rRNA function, have

arisen. The transcribed spacers of eukaryotic rDNA undergo rapid evolutionary change. Clearly natural selection allows these regions to accept and accumulate mutations.

The identification of multiple variants in the transcribed spacers of X. laevis rDNA demonstrates that the overall rate of change within the transcribed spacers is fast enough to allow substantial heterogeneity to be seen at any point in time in these spacer regions - a previously unrecognised fact. Even within single nucleolar organisers, some degree of sequence variation exists. Thus the variants indeed occur on physically linked repeats within individual chromosomes, and are not, for example, a consequence of sampling of different geographical populations of frogs with different but internally homogeneous transcribed spacer arrays.

Comparison of two closely related species of frog - X.laevis and X. borealis has led to the identification of homologous blocks of sequence within all three transcribed spacer regions (Furlong and Maden, 1983; Furlong et al, submitted for publication). This analysis has clearly demonstrated that the transcribed spacers in X. laevis and X. borealis have evolved from common ancestral transcribed spacers sequences. However, although the conserved tracts in X. laevis and X. borealis occur in the same linear order in the transcribed spacers of the two species, the distances and sequence patterns between these tracts differs, implying a history of base changes and insertions and deletions. I would like to suggest that the identification of intraspecies

transcribed spacer variants in X. laevis rDNA in the form of base changes and insertions and deletions of one or a few nucleotides, implies a degree of sequence instability which may underly this phylogenetic variation.

Thus these data suggest that phylogenetic change in the transcribed spacers occurs by the repeated occurrence of small insertions or deletions in addition to point mutations.

8.3 Function of the transcribed spacers

Because of their presence within the ribosomal transcription unit, the transcribed spacers have always been assumed to play some specific role in rRNA maturation. However, although the order of processing steps in eurkaryotic rRNA is well known, little is understood of the precise mechanisms involved in each step and even less is understood of the function of the transcribed spacers in these processing steps. A question which must be asked therefore is "Do the transcribed spacers serve a specific function in rRNA maturation?". In order to answer this question consider the following.

a) How important is primary sequence ?

Comparison of sequence data from a number of eukaryotes (Table 1.1(b) for references) reveals little homology among the transcribed spacers. Such extensive divergence does not seem to suggest any function that is closely related to sequence content. The identification of intra-specific transcribed spacer variation strengthens this proposition

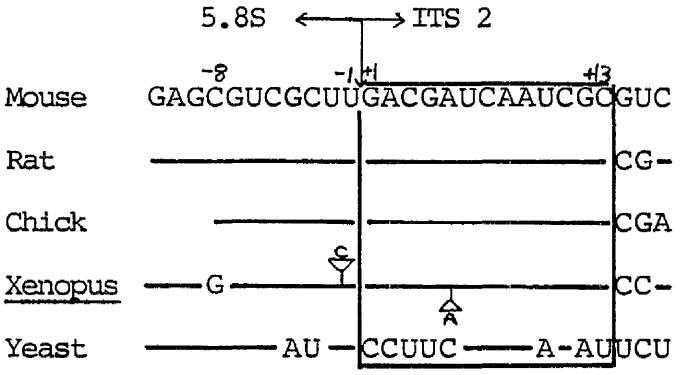


Figure 8.1 Sequence conservation at the 5'start of ITS 2 in a number of eukaryotes.

The data for mouse, rat, Xenopus and yeast are from references citied in Table 1.1b. The chick data are from Crouch et al, 1983. In each sequence, nucleotides identical with the mouse sequence are represented by a straight line.

△ means one nucleotide is deleted in the Xenopus sequence with respect to the mouse sequence.

▽ means that one nucleotide is inserted in the Xenopus sequence with respect to the mouse sequence.

No sequence information upstream of nucleotide -8 in the 5.8S gene is available for chick.

especially when we consider that in this study, I found that no two transcription units were alike within their transcribed spacer regions. In addition, several of the variable sites identified occur quite close to the rRNA coding regions, implying that there is no unique relationship between sequence and function in these rRNA flanking regions. There are variants within 30 nucleotides of the 5' and 3' ends of the 18S sequence and the 5' end of the 5.8S gene (Figures 5.4, 5.8, 5.11). A variable length C tract in ITS2 terminates only a few nucleotides from the 28S boundary (Figure 5.8).

There is however one region in the transcribed spacers, namely the 5' start of ITS2, which shows a significant degree of sequence homology between rat, mouse, chick, yeast and Xenopus (Figure 8.1). This region is also highly homogeneous in X. laevis ITS2 (Figures 5.8, 6.5). The significance of this homology is unclear. A role for ITS2 in base-pairing with a small nucleolar RNA, U3, has been proposed by Bachellerie et al, (1983) and Crouch et al, (1983). However, the ubiquity of this feature in eukaryotic ITS2 is yet to be established.

On the whole, the primary sequence data available suggest that if the transcribed spacers are functionally important, this function must be totally independent of primary sequence. Perhaps transcribed spacer function in the RTU, is simply to maintain distance between the ribosomal coding regions, to allow the molecules responsible for generating mature rRNAs access to their sites of action.

b) Sequence conservation between Xenopus and mouse

Comparison of the primary sequence between closely related species (e.g. X. laevis/X.borealis : Furlong and Maden 1983, Furlong et al, submitted for publication, mouse/rat : Michot et al, 1983), has revealed that embedded in the largely divergent sequences of the transcribed spacers are short tracts of completely conserved or almost completely conserved sequence. None of the X. laevis transcribed spacer variants identified in this study occur within any of the conserved X. laevis/X. borealis tracts. Comparison of these tracts between the two species of frog and two species of rodent reveals that two of the seven ITS conserved tracts in X. laevis/X. borealis are also conserved in mouse. A 20 nucleotide tract in ITS1 of frog (tract 1) contains 14 nucleotides which are exactly conserved in mouse ITS1. Near the start of ITS2, 19 of the 25 nucleotides of tract 1 in frog are conserved in both mouse and rat. Considering the very low extent of overall homology within the entire ITS sequences, on the one hand between both Xenopus species and on the other hand between mouse/Xenopus and rat/Xenopus, this observation is very surprising. It could be interpreted to suggest that at least some of these conserved ITS sequences are functionally significant. Equally, however, these sequences may simply represent mere evolutionary relics conserved by chance within regions undergoing rapid divergence. Whichever of these possibilities is correct remains unknown at present.

c) Secondary structure considerations

In E. coli, the sequences flanking 16S and 23S rRNA form extensive base-paired structures which are recognised and cleaved by RNase III during ribosome maturation (Young and Steitz, 1978; Bram et al, 1980). It is tempting to envisage a similar situation for eukaryotic rRNA formation. However, in X. laevis the sequences flanking 18S rRNA show no significant potential for interaction (Maden et al, 1982a). In S. carlsbergensis an intermediate situation appears to exist : the sequence at the 5' end of 17S rRNA is cleaved simultaneously with a sequence downstream from the 3' end, but the potential for base-pairing between the cleavage sites is limited (Veldman et al, 1981). Thus, it seems unlikely that mature eukaryotic RNAs are generated by molecular mechanisms similar to those in E. coli.

8.4 Conclusions

The degree of heterogeneity in the transcribed spacers of X. laevis rDNA has been critically assessed by direct nucleotide sequence analysis. Identification of sequence variants in these regions has provided novel and important data relevant to the evolution of eukaryotic transcribed spacer rDNA and also queries the function of these sequences in ribosome formation.

I would like to suggest that large parts of the transcribed spacers behave as "neutral" regions because of their ability to accept mutations. Some regions appear to be inherently more mutable than other regions : it is interesting

to observe that nine of the twenty sequence variants identified in this study are located within some of the simple sequence tracts described in Table 8.1, as discussed above. If the transcribed spacers do play a specific role in ribosome formation, then I propose that these "neutral" regions are functionally unimportant, and only those regions which show intraspecies homogeneity or which have a tendency towards phylogenetic conservation (e.g. 5' start ITS2, Figure 8.1) should be considered as functionally significant.

REFERENCES

Abelson, J. (1979) Ann. Rev. Biochem. 48, 1035-1069.

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

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

Attardi, G. & Amaldi, F. (1970) Ann. Rev. Biochem. 39, 183-226.

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

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

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

Bird, A.P. (1980) in Cell Biology (Goldstein, L. & Prescott, D.M., eds.), vol. 3, pp 62-111, Academic Press, London and New York.

Bird, A.P. & Southern, E.M. (1978) J. Mol. Biol. 118, 27-47.

Bird, A.P., Rochaix, J-D. & Bakken, A.H. (1973) in Molecular Cytogenetics (Hamkalo, B.A. & Papaconstantinou, J., eds.), pp 49-58, Plenum, New York.

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

Birnstiel, M.L., Wallace, H., Sirlin, J. & Fischberg, M. (1966) Nat. Cancer Inst. Monogr. 23, 431-447.

Birnstiel, M., Speirs, J., Purdom, I., Jones, K. & Loening, U.E. (1968) Nature 219, 454-463.

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

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

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

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

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

Breathnach, R. & Chambon, P.A. (1981) Ann. Rev. Biochem. 50, 349-383.

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

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

Brown, D.D. & Blackler, A.W. (1972) J. Mol. Biol. 63, 75-83.

Brown, D.D. & Dawid, I.B. (1968) Science 160, 272-280.

Brown, D.D. & Dawid, I.B. (1969) Ann. Rev. Genetics 3, 127-154.

Brown, D.D. & Gurdon, J.B. (1964) Proc. Nat. Acad. Sci. USA 51, 139-146.

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

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

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

Cech, T.R. & Rio, D.C. (1979) Proc. Nat. Acad. Sci. USA 76, 5051-5055.

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

Coen, E., Strachan, T. & Dover, G. (1982) J. Mol. Biol. 158, 17-35.

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

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

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

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

Dover, G. (1982) Nature 299, 111-117.

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

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

Edwards, K. & Kőssel, H. (1981) Nucleic Acids Research 9, 2853-2869.

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

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

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

Gall, J.G. (1968) Proc. Nat. Acad. Sci. USA 60, 553-560.

Gall, J.G. (1969) Genetic Suppl. 61, 1, 121-132.

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

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

Gilbert, W. & Dressler, D. (1968) Cold Spring Harbour Symp. Quant. Biol. 33, 473-484.

Glotz, C. & Brimacombe, R. (1980) Nucleic Acids Research 8, 2377-2395.

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

Goodpasture, C. & Bloom, S.E. (1975) Chromosoma 53, 37-50.

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

Gronenborn, B. & Messing, J. (1978) Nature 272, 375-377.

Grunstein, M. & Hogness, D.S. (1975) Proc. Nat. Acad. Sci. USA 72, 3961-3965.

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

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

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

Hourcade, D., Dressler, D. & Wolfson, J. (1973) Cold Spring Harbour Symp. Quant. Biol. 38, 537-550.

Howell, W.M., Denton, T.E. & Diamond, J.R. (1975) Experientia 31, 260-262.

Jeanteur, P., Amaldi, F. & Attardi, G. (1968) J. Mol. Biol. 33, 757-775.

Lewin, B. (1976) Nature 260, 574-576.

Lima-de Faria, A., Birnstiel, M. & Jaworska, H. (1969) Genetic Suppl. 61, 1, 145-159.

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

Macgregor, H.C. (1968) J. Cell Science 3, 437-444.

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

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

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

Marmur, J. (1961) J. Mol. Biol. 3, 208-218.

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

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

Messing, J., Gronenborn, B., Müller-Hill, B. & Hofschneider, P.H. (1977) Proc. Nat. Acad. Sci. USA 74, 3642-3646.

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

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

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

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

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

Pene, J.J., Knight, E. Jnr. & Darnell, J.E. Jnr. (1968) J. Mol. Biol. 33, 609-623.

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

Perry, R.P., Cheng, T.Y., Freed, J.J., Greenberg, J.R., Kelley, D.E. & Tartof, K.D. (1970) Proc. Nat. Acad. Sci. USA 65, 609-616.

Rochaix, J-D., Bird, A. & Bakken, A. (1974) J. Mol. Biol. 87, 473-487.

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

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

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

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

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

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

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

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

Schwarz, Zs. & Kössel, H. (1980) Nature 283, 739-742.

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

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

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

Smith, G.P. (1973) Cold Spring Harbour Symp. Quant. Biol. 38, 507-513.

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

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

Southern, E.M. (1979) Methods in Enzymology 68, 152-176.

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

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

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

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

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

Tartoff, K.D. (1975) Ann. Rev. Genetics 9, 355-385.

Tocchini-Valentini, G.P. & Crippa, M. (1971) Lepetite Colloq. Biol. Med. 2, 237-243.

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

Twigg, A.J. & Sherratt, D. (1980) Nature 283, 216-218.

Van Etten, R.A., Walberg, M.W. & Clayton, D.A. (1981) Cell 22, 157-170.

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

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

Vincent, W.S., Halvorson, H.O., Chen, H-R. & Shin, D. (1969) Exp. Cell Res. 57, 240-250.

Wallace, H. & Birnstiel, M.L. (1966) Biochim. Biophys. Acta. 114, 296-310.

Wallace, H., Morray, J. & Langridge, W.H.R. (1971) Nature 230, 201-203.

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

Wellauer, P.K. & Dawid, I.B. (1974) J. Mol. Biol. 89, 377-395.

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

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

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

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

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

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

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

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. & Noller, H.F. (1980) Nucleic Acids Research 8, 2275-2293.

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

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